INTERACTIVE CONTROL ENGINEERING

Control of Biological Systems and Human Performance

Interactive Control Engineering: Principles and Applications

Edited by John D. Hedengren

Control technology

1 Process Dynamics, Modeling, and Identification
2 Practical PID Control
3 Feedforward, Cascade, and Ratio Control
4 Override, Split-Range, and Constraint Control
5 Delays, Slow Measurements, and State Estimation
6 Multivariable and Plantwide Control
7 Nonlinear Dynamics, Oscillations, and Stabilization
8 Batch, Run-to-Run, and Hybrid Control
9 Model Predictive Control and Economic Optimization

Industry applications

10 Control in Upstream and Midstream Oil and Gas
11 Control in Petroleum Refining
12 Control in Petrochemicals and Polymers
13 Control in Chemical Plants
14 Control in Semiconductor Manufacturing
15 Control in Mining and Mineral Processing
16 Control in Iron, Steel, and Metals Production
17 Control of Energy Conversion and Storage
18 Control of Nuclear and Fusion Systems
19 Control of Biological Systems and Human Performance
20 Motion Control, Robotics, and Autonomous Systems

INTERACTIVE CONTROL ENGINEERING
Principles and Applications

Volume 19

Control of Biological Systems and Human Performance

Variable subjects, sparse samples and limits that cannot be crossed

John D. Hedengren
Editor

Advanced Process Solutions, LLC

Interactive Control Engineering: Principles and Applications
Volume 19: Control of Biological Systems and Human Performance
Edited by John D. Hedengren

First edition, 2026

Copyright © 2026 Advanced Process Solutions, LLC. All rights reserved.

Produced by Advanced Process Solutions, LLC.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means without the prior written permission of the publisher, except for brief quotations in a review or scholarly work.

The simulators described in this book are teaching models. Their equations, parameters and results are simplified for instruction and are not validated for the design, operation or safety assessment of any real plant, vehicle or medical device. The editor and the publisher accept no liability for use of this material outside that purpose.

Names of products and companies are used for identification only and may be trademarks of their owners.

Every chapter has a companion simulator that runs in a web browser at https://apmonitor.com/apps/. This volume online, with chapter files and citation records: https://apmonitor.com/apps/books/v19-biological-human/

How to cite this volume. John D. Hedengren, Ed., *Control of Biological Systems and Human Performance*, vol. 19 of *Interactive Control Engineering: Principles and Applications*. Advanced Process Solutions, LLC, 2026.

Typeset in Times and Helvetica-family type with LaTeX and the memoir class.

Preface

This book is one of 20 volumes of *Interactive Control Engineering: Principles and Applications*. Each chapter is about one plant, machine or system and one control lesson, and each has a simulator that runs in a web browser with nothing to install. The simulators are collected at https://apmonitor.com/apps/. Nine volumes are organized by control technology and eleven by industry or application area; the last pages of this volume list them all.

How to read this book with the apps

Chapter 1 teaches the theory this volume relies on and explains, once, the simulator framework that the apps share. Each chapter after it opens with the control question it answers, its learning objectives, what to read first, and the address of its app, also given as a code to scan. The chapter then develops the model, the controllers and the measured results. Every number reported was measured on the app, so a reader can open the app, load the named scenario and reproduce it.

Every chapter ends with exercises. Each is labeled with its kind and the time it should take, and none takes more than ten minutes. Concept checks and calculations can be done on paper. App challenges name a scenario, the settings and the run length. One exercise in each chapter, labeled *Review the agent*, presents a tuning, a model or a claim proposed by an AI assistant and asks what is wrong with it. Short answers to the odd-numbered exercises are at the back of the book, and worked solutions are available to instructors.

How AI agents were used

AI agents did much of the development work behind this series. Under the editor's direction they wrote and tested the simulators, drafted the papers from which the chapters are drawn, measured the reported results on the running apps, converted the papers to book form, and drafted the introductory and closing chapters and the exercises. Other agents then checked that work: they re-measured reported numbers, audited what each app displays against the state of its model, reviewed the papers as referees, and solved the exercises without seeing the answers. The editor set the scope, chose the plants and the lessons, reviewed the results and is responsible for the content.

That way of working is also a subject of the book. The last chapter of each volume describes what the agents got wrong while this volume was made and how the errors were caught, and it sets out what the editor believes this means for engineers: more of the routine work can be delegated, and more of the engineer's value lies in checking, supervising, choosing the problem and knowing the process.

About this volume

This volume is about control of living systems. Chapters 2 to 5 cover medical loops: insulin delivery, anesthesia, ventilation and kidney replacement. Chapters 6 and 7 pace an athlete whose capacity is used up as it is spent. Chapters 8 to 13 cover cell culture, fermentation, digestion and the farm. Chapter 14 steers an epidemic from counts that are partial and a week old.

The subjects differ from one individual to the next, samples are sparse, and the actuators act in one direction against limits that cannot be crossed. The simulators are teaching models and are not medical devices. Each chapter names the technology volume that teaches its main technique.

John D. Hedengren
2026

Contents

Preface v

Notation x

1 Introduction: Controlling Living Systems 1
- 1.1 The Question This Volume Answers . . . 1
- 1.2 A Short History . . . 2
- 1.3 Notation and Conventions . . . 4
- 1.4 Models, Variability and Limits of Living Plants . . . 5
- 1.5 Instruments and Final Elements in This Volume . . . 11
- 1.6 The Simulator Framework . . . 14
- 1.7 Working With AI Agents While Studying . . . 17
- 1.8 Chapter Roadmap . . . 18
- 1.9 Exercises . . . 21

2 Artificial Pancreas Simulator for Direct Action and One-Sided Insulin Delivery 23
- 2.1 Introduction . . . 24
- 2.2 Background . . . 25
- 2.3 Process Model . . . 26
- 2.4 Steady-State and Dynamic Analysis . . . 30
- 2.5 Controller Design . . . 32
- 2.6 Asymmetric Performance Metrics . . . 36
- 2.7 Results . . . 37
- 2.8 Predictive Control and the Announced Meal . . . 41
- 2.9 Estimator Choice and Relay-Written Constants . . . 44
- 2.10 Technology in Practice . . . 45
- 2.11 Limitations . . . 47
- 2.12 Discussion . . . 48
- 2.13 Conclusion . . . 48
- 2.14 Exercises . . . 51

3 Closed-Loop Anesthesia Simulator for Model Inversion with Feedback Correction 54
- 3.1 Introduction . . . 55
- 3.2 Automation of Intravenous Anesthesia . . . 56
- 3.3 Patient Model . . . 57
- 3.4 Dynamic Analysis . . . 63
- 3.5 Control Structure as Implemented . . . 65
- 3.6 Scenarios and Measured Results . . . 68
- 3.7 Technology in Practice . . . 77
- 3.8 Limitations . . . 79
- 3.9 Conclusion . . . 80
- 3.10 Exercises . . . 82

4 Mechanical Ventilator Simulator for Patient-Triggered Cycles and Breath-by-Breath Identification 86
- 4.1 Introduction . . . 87
- 4.2 Breathing Machines and the Patient Behind Them . . . 88
- 4.3 A Lung Behind a Wye . . . 89

4.4 What the Equations Promise . . . 93
4.5 The Ventilator's Loops . . . 95
4.6 Nine Cases and an Hour . . . 98
4.7 Technology in Practice . . . 103
4.8 Limitations . . . 105
4.9 Conclusion . . . 105
4.10 Exercises . . . 107

5 Continuous Kidney Replacement Simulator for Cumulative Balance and Measurement Redundancy 111
5.1 Introduction . . . 112
5.2 Continuous Therapy and Automatic Balancing . . . 113
5.3 Circuit Model . . . 114
5.4 A Small Difference of Large Flows . . . 119
5.5 Controller as Implemented . . . 121
5.6 Cases and Measured Outcomes . . . 124
5.7 Technology in Practice . . . 128
5.8 Limitations . . . 131
5.9 Conclusion . . . 131
5.10 Exercises . . . 133

6 Treadmill Pacing Simulator for Constraint Selection and Fixed-Horizon Optimization 137
6.1 Introduction . . . 138
6.2 Background: Controlling a Person on a Belt . . . 139
6.3 Plant Model . . . 140
6.4 Dynamic Analysis . . . 148
6.5 Controller Design . . . 149
6.6 Performance Measures . . . 152
6.7 Results . . . 153
6.8 Technology in Practice . . . 159
6.9 Limitations . . . 162
6.10 Discussion . . . 162
6.11 Conclusion . . . 163
6.12 Exercises . . . 165

7 Cycling Exertion Simulator to Teach Control of a Depleting Actuator 169
7.1 Introduction . . . 170
7.2 Background: Ergometry, Critical Power and Pacing . . . 171
7.3 Rider, Bicycle and Road Model . . . 172
7.4 Dynamic Analysis . . . 181
7.5 Controller Design . . . 184
7.6 Performance Measures . . . 187
7.7 Results . . . 188
7.8 Technology in Practice . . . 194
7.9 Limitations . . . 195
7.10 Discussion . . . 196
7.11 Conclusion . . . 196
7.12 Exercises . . . 198

8 Feeding an Antibody Culture Whose Appetite Outruns Its Daily Sample 202
8.1 Introduction . . . 203
8.2 From Hybridomas to Cultures Fed on Their Own Measurements . . . 204

8.3 The Culture as Modeled . . . 205
8.4 What a Day, a Bubble and a Liter of Base Allow . . . 211
8.5 The Controller and What It Reads . . . 212
8.6 Ten Cases and a Fourteen-Day Challenge . . . 217
8.7 Technology in Practice . . . 226
8.8 Limitations . . . 228
8.9 Conclusion . . . 228
8.10 Exercises . . . 230

9 Refolding Insulin Precursor When a Better Fold Can Cost the Campaign 234
9.1 Introduction . . . 235
9.2 From Pancreatic Extract to a Folded Precursor . . . 236
9.3 The Route as Modeled . . . 237
9.4 Code Verification and Agreement With the Pilot Study . . . 241
9.5 The Controllers and What They Read . . . 243
9.6 Ten Cases, Measured . . . 245
9.7 Technology in Practice . . . 248
9.8 Limitations . . . 250
9.9 Conclusion . . . 251
9.10 Exercises . . . 253

10 Lager Fermenter Simulator With On/Off Glycol Valves and a Day-Old Diacetyl Result 256
10.1 Introduction . . . 257
10.2 Refrigeration, Fermenting Vessels and Diacetyl . . . 258
10.3 The Tank as Modeled . . . 259
10.4 Implemented Control System . . . 264
10.5 Cases and What They Measure . . . 267
10.6 Technology in Practice . . . 273
10.7 Limitations . . . 275
10.8 Conclusion . . . 275
10.9 Exercises . . . 277

11 Feeding an Anaerobic Digester Toward Its Fold on a Daily Titration 280
11.1 Introduction . . . 281
11.2 Failing Digesters, Inhibition and the Titration . . . 282
11.3 The Digester as Modeled . . . 283
11.4 The Fold . . . 288
11.5 The Control Layers as Implemented . . . 289
11.6 Nine Cases and a Season, Measured . . . 291
11.7 Three Kinds of Evidence . . . 296
11.8 Technology in Practice . . . 296
11.9 Limitations . . . 298
11.10 Conclusion . . . 299
11.11 Exercises . . . 301

12 Recirculating Fish Farm Simulator Where Nitrite Answers the Feed Days Later 304
12.1 Introduction . . . 305
12.2 Reused Water and the Farms Built Around It . . . 306
12.3 The Farm as Modeled . . . 307
12.4 Two Clocks in One Tank . . . 312
12.5 The Control Layers as Implemented . . . 313
12.6 Ten Cases, Measured . . . 315

12.7 Technology in Practice . . . 318
12.8 Limitations . . . 321
12.9 Conclusion . . . 321
12.10 Exercises . . . 323

13 Greenhouse Climate Simulator for Temperature Integration Under a Humidity Limit 326
13.1 Introduction . . . 327
13.2 From Hand-Set Vents to a Planned Daily Mean . . . 328
13.3 One Hectare Under Glass . . . 329
13.4 The Climate Computer as Implemented . . . 335
13.5 The Planner: A Receding-Horizon Linear Program . . . 337
13.6 Nine Cases and a Spring Week . . . 338
13.7 Technology in Practice . . . 343
13.8 Limitations . . . 345
13.9 Conclusion . . . 346
13.10 Exercises . . . 348

14 Estimating and Steering an Epidemic From Week-Old, Partial Counts 351
14.1 Introduction . . . 352
14.2 Compartments, Undercounts and Planned Measures . . . 353
14.3 A Declared City and What It Publishes . . . 354
14.4 Growth, Final Size and Three Delays . . . 357
14.5 Three Estimators on One Augmented State . . . 357
14.6 Dial Policies as Implemented . . . 359
14.7 Measurements on Ten Cases . . . 360
14.8 The City on the Estimation Ladder . . . 366
14.9 Technology in Practice . . . 366
14.10 Limitations . . . 368
14.11 Conclusion . . . 368
14.12 Exercises . . . 370

15 Summary and Outlook 373
15.1 What the Chapters Showed . . . 373
15.2 Choosing a Method . . . 377
15.3 Where Practice Is Heading . . . 377
15.4 Engineering With AI Agents . . . 379
15.5 Gaps in the Control of Living Systems . . . 382
15.6 Exercises . . . 386

Answers to Selected Exercises 388

Glossary 393

Index of Control Topics 396

Index of Applications 399

About the Editor 402

Volumes in This Series 403

Notation

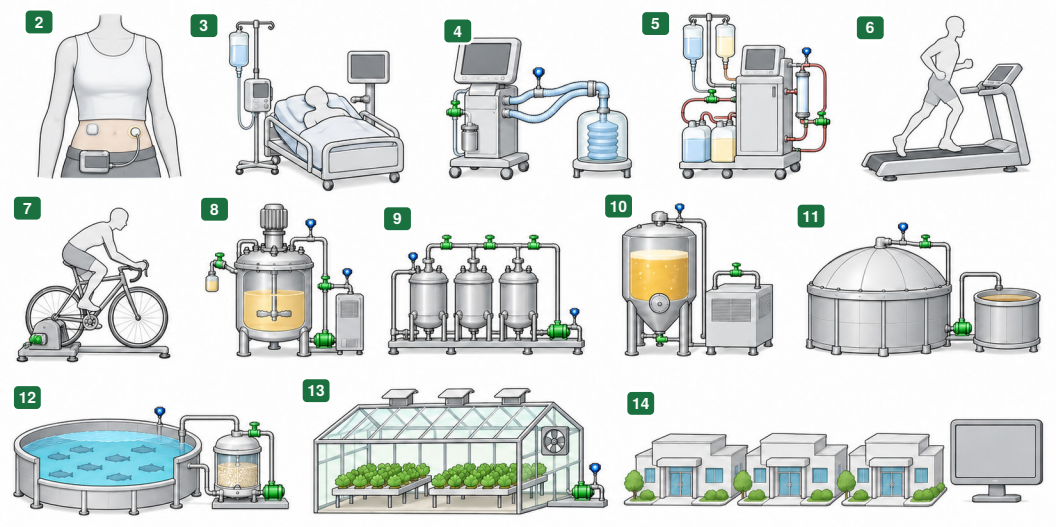
The chapters share the symbols below. A symbol used by one chapter only is defined where it first appears, and each chapter states its own units.

Symbol	Meaning
SP	setpoint, the value a controller is asked to hold
PV	process variable, the measurement a controller acts on
CV	controlled variable
MV	manipulated variable; also OP, the controller output
DV	disturbance variable
e	control error, SP $-$ PV
K_c	controller gain
τ_I, τ_D	integral (reset) time and derivative time
K_p	process gain, the steady-state change in PV per unit change in MV
τ_p	process time constant
θ_p	process dead time
τ_c	closed-loop time constant chosen in a tuning rule
Δt	controller sample time
s	Laplace variable
x, u, y, d	state, input, output and disturbance of a state-space model
A, B, C	matrices of the linear model $\dot{x} = Ax + Bu$, $y = Cx$
Q, R	weights on outputs (or states) and on input moves in an optimal controller
$\hat{x}$	an estimate of x

Abbreviations used throughout: PID, proportional-integral-derivative; PI, proportional-integral; FOPDT, first-order plus dead time; MPC, model predictive control; MHE, moving horizon estimation; IAE, integral of the absolute error; RGA, relative gain array. Times are in seconds unless a chapter says otherwise. Controller output and valve position are in percent of span. A scenario or case number refers to the numbered list in the chapter's app.

CHAPTER 1

Introduction: Controlling Living Systems



The chapters 2 artificial pancreas 3 closed-loop anesthesia 4 mechanical ventilator 5 kidney replacement 6 treadmill pacing 7 cycling exertion 8 antibody culture 9 insulin refolding 10 lager fermenter 11 anaerobic digester 12 recirculating fish farm 13 greenhouse climate 14 epidemic response

Notice that the top row is the bedside and a runner, the middle row a rider and the vessels in which cells make medicines, beer and gas, and the bottom row a fish farm, a glasshouse and a city. Illustration, not to scale.

IN PLAIN TERMS **Keeping Living Things Within Safe Limits**

The main idea. A pump gives a patient a drug, a feed keeps cells growing, and a cold valve holds a tank of beer at its temperature. In each case the thing being controlled is alive, and no two of them are quite the same.

The challenge. The value that matters most is often measured once a day or not at all. A pump can add a drug but cannot take it back, and some limits must never be crossed, such as a blood sugar that falls too low.

Why it matters. Good control keeps patients safe, makes medicines and food reliably, and helps a city keep its hospital from filling up.

THREE WAYS TO READ THIS CHAPTER

Explore	7 min	This page, Section 1.2 with its timeline and two History boxes, Figure 1.2, the Did you know? boxes in Sections 1.4.5, 1.4.6 and 1.4.7, and Key points in Section 1.8. No equations.
Study	45 min	The whole chapter in order, working each number of the examples P-1 and C-1, then Exercises 1 to 10.
Practice	5 min	This page, Section 1.4.7 with Table 1.1, Section 1.5 with Figures 1.7 and 1.9, and Section 1.8.

1.1 The Question This Volume Answers

A patient, a runner, a culture of cells, a tank of fish and a city have properties that a heat exchanger does not. No two of them are alike, and the model the controller carries describes an average one. The variable that matters is often measured once a day, a week late or not at all. The plant grows, so the demand it puts on the actuator doubles while the controller waits for its next sample. Many of the

actuators push one way only, and the limits that the plant must not cross are hard on one side and soft on the other. This volume answers one question: how does a loop hold a living plant inside limits it may not exceed when the individual differs from the model, the measurement is sparse and the actuator cannot take back what it has given?

Seven control problems define the field, and every chapter meets several of them.

- *The variable that matters is not measured.* No probe reads the drug at its site of action in the brain, the breath that reaches the lung behind a leak, the nitrifiers on a biofilter or the infections in a city.
- *Each individual differs from the model.* The propofol an adult needs varies by a factor of three, the plant gain of a runner on a belt differs by a factor of 2.1 between four runners, and a yeast strain's diacetyl runs to 1.96 times what its estimator predicts.
- *Measurements are sparse and late.* A culture is counted once a day, a digester is titrated each morning and reported six hours later, and an epidemic is reported a week after the onsets it counts.
- *The plant grows.* Cells double, their appetite doubles with them, a biofilter must be grown before it can carry its load, and an epidemic doubles in days.
- *Actuators push one way.* A pump can add insulin or propofol and cannot withdraw it, a glycol valve can only cool, and a feed pump can only feed.
- *Limits are hard and asymmetric.* A deep low glucose is an immediate danger and a high one a slow one, and a plateau pressure, a nitrite level or a count of hospital beds is a line the loop must not cross.
- *Some actuators are stores.* The power a rider buys with effort above a sustainable level is drawn from a work capacity that runs out, and a digester's buffer spends itself on the acids it neutralizes.

Thirteen chapters answer the question on thirteen plants, each with an app that runs in a browser. Chapters 2 to 5 are at the bedside: an insulin pump, an anesthetic infusion, a ventilator and a kidney replacement circuit. Chapters 6 and 7 pace a runner and a rider. Chapters 8 to 10 make an antibody, refold insulin and ferment a lager. Chapters 11 to 13 feed a digester, a fish farm and a glasshouse crop, and Chapter 14 steers an epidemic.

A reader who has worked through the volume can do the following on a living plant that is new to the reader.

- Write a compartment model, find its time constants and the infusion that holds a concentration, and say why that infusion keeps falling.
- Write a growth law, predict how fast demand grows, and find the fold beyond which an inhibited culture washes out.
- State what a population model gets wrong on an individual and which structure removes the error.
- Correct a model from a sample at the time the sample was drawn, and say what interval a growing plant allows between samples.
- Size the margin a one-way actuator needs in front of a hard limit, and place an override where a limit binds.

This chapter develops each of these on two small examples and names the chapter that uses it. PID forms and tuning are taken as known from Volume 2, feedforward and cascade from Volume 3, selectors and overrides from Volume 4, delays and estimation from Volume 5, multiplicity from Volume 7, batch operation from Volume 8 and predictive control from Volume 9.

1.2 A Short History

The milestones of Figure 1.1 follow the plants of this volume, from a pandemic and a purified extract to the first automatic loops at the bedside, in the glasshouse and in the digester.

In 1918 seventeen cities of the United States brought in measures against pandemic influenza, among them the closure of schools, churches and theaters. Cities that put several measures in place early in their epidemic had peak death rates about half those of cities that did not [1], the steering problem of Chapter 14. An apparatus for the prolonged administration of artificial respiration,

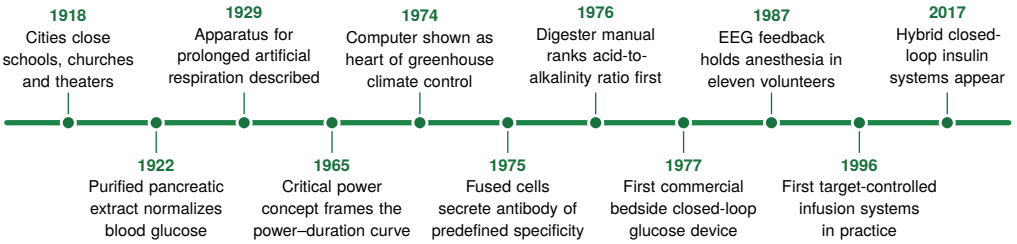


Figure 1.1 Milestones of biomedical and biological control. The dates are from [1] (1918), [2] (1922 and 2017), [3] (1929), [4] (1965), [5] (1974), [6] (1975), [7] (1976), [8] (1977), [9] (1987) and [10] (1996). Notice that the first loops closed by machine on living plants all fall within the 1970s and 1980s. Events are evenly spaced, not to a time scale.

described in 1929, enclosed the body below the neck in a tank and breathed for a child paralyzed by poliomyelitis almost continuously for 122 hours [3], a forerunner of the ventilator of Chapter 4. In 1965 the concept of critical power gave the power a person can hold over a given time a simple frame of two parameters, a fixed energy reserve and the critical power itself, and in 1981 the concept was applied to cycling [4]. That reserve is the store of Chapter 6 and Chapter 7.

> ⧖ HISTORY **1974 to 1987: First Automatic Loops on Living Plants**
>
> At the Dutch horticultural fair of 1974 a computer was shown for the first time as the heart of a climate control system for greenhouses, and within two years systems from five producers were on sale [5]. In 1977 the first continuous arteriovenous hemofiltration was given in Göttingen. Its removal was first set by hand, by hanging the filtrate bag higher or lower, and later systems balanced the fluid automatically with scales and electronic clamps [11]. A report of 1978 found such balancing accurate enough to replace the bed scales that bedridden patients had needed [12]. In 1987 a loop held anesthesia in 13 volunteers by feedback on the median frequency of the electroencephalogram, with a setpoint of 2 to 3 Hz, and succeeded in 11 of them [9]. The climate loop of Chapter 13, the balance of Chapter 5 and the depth loop of Chapter 3 are the modern forms of these three problems.

A report of 1975 described continuous cultures of fused cells that secreted antibody of a predefined specificity [6], the first monoclonal antibodies, a class of product that Chapter 8 makes in cultured cells. In February 1976 a federal operations manual for sludge digesters ranked the ratio of volatile acids to alkalinity first among the indicators that warn of an upset [7], the measurement on which Chapter 11 feeds. The first target-controlled infusion systems came into practice in 1996, with a model for propofol [10], the model inversion of Section 1.4.2.

> ⧖ HISTORY **1922 to 2017: Insulin, From an Extract to a Closed Loop**
>
> On 23 January 1922 a purified pancreatic extract was given under the skin to a patient with diabetes. The sugar and ketones in his urine disappeared, his blood glucose became normal, and he lived 13 more years on daily injections [2]. Half a century later the dose was put under feedback at the bedside. Five teams reported closed loops between 1974 and 1978 that measured glucose in a vein and computed the insulin from the glucose and its rate of change, a proportional-derivative law, and in 1977 one of them became the first commercial device [8]. The first commercially used insulin pump followed in 1978, glucose sensors under the skin were joined to pumps from the early 2000s, and in 2017 hybrid closed-loop systems combined a continuous monitor, a control algorithm and a pump [2]. Chapter 9 makes the insulin, and Chapter 2 closes the loop on it.

1.3 Notation and Conventions

Concentrations are in the units of each field: glucose in mg/dL at the bedside and in g/L in a culture, drugs in mg/L, which equals µg/mL, and nitrogen in mg N/L. A volume V is in liters unless a chapter gives cubic meters, a clearance Cl is a volume cleared per unit time, and a growth rate μ is per hour or per day as the chapter's clock runs. The clocks differ. Chapter 2 runs a model clock in hours on rate constants per minute. Chapter 3 counts minutes, Chapter 4 seconds and milliseconds, and Chapter 6 and Chapter 7 seconds. Chapter 5 closes its balance in hours over drive loops in seconds. Chapter 8, Chapter 9 and Chapter 10 count hours and days, Chapter 12 days, hours and minutes, Chapter 13 minutes and hours, and Chapter 11 and Chapter 14 days.

The framework's h is its integration step and its a the controller action. Loop action follows Section 1.6.2: $a = +1$ is reverse acting, for a plant whose measurement rises with the manipulated variable, and $a = -1$ is direct acting. An insulin pump on glucose is direct acting, because insulin lowers glucose, and so is the depth loop of an anesthetic, because more propofol lowers the depth index. So are a ventilator's expiratory valve on airway pressure, a glycol valve on a fermenting zone, a glasshouse vent on air temperature, a blower on dissolved carbon dioxide and a dial of distancing measures on a hospital census. Base on pH, a jacket on a culture's temperature, gas on dissolved oxygen and a belt on a runner's heart rate are reverse acting. Several chapters write the action with their own symbol: $\mathscr{A}$ in Chapter 3, Chapter 4 and Chapter 5, γ in Chapter 10, γ_a in Chapter 13 and α in Chapter 12.

The chapters keep their own symbols, and many collide. ρ is the slack price of the framework's predictive controller, a patient's sensitivity in Chapter 3, the share of infections reported in Chapter 14 and a ration fraction in Chapter 12. γ is also the Hill slope in Chapter 3, the inverse infectious period in Chapter 14 and the psychrometric constant in Chapter 13. λ is a move suppression in Chapter 2, a bias weight in Chapter 10 and Chapter 8 (λ_X) and the latent heat in Chapter 13. σ is the monitor noise in Chapter 2, the lactate switch in Chapter 8 and the inverse latent period in Chapter 14. D is a dilution rate in Chapter 8 and Chapter 11, the diacetyl concentration in Chapter 10 and the vaccine doses a day in Chapter 14. φ is a central-volume multiplier in Chapter 3, a throttle parameter in Chapter 5 (which writes the filtration fraction FF), the relative humidity in Chapter 6 and Chapter 7, and the air exchange in Chapter 13, while ϕ is the accumulated fatigue in Chapter 6 and Chapter 7. C is a compliance in Chapter 4 and a concentration nearly everywhere else. k_e is the plasma insulin clearance in Chapter 2 and an expiratory valve coefficient in Chapter 4, and E is the exertion in Chapter 7, the exposed class in Chapter 14 and the ethanol in Chapter 10. a itself is a pressure coefficient in Chapter 3, the airflow in Chapter 6 and Chapter 7, the vent opening in Chapter 13 and a correlation coefficient in Chapter 14, none of them the action. μ is a specific growth rate except as the peak-power factor of Chapter 6 and the mean daily report of Chapter 14. This chapter writes the clearance ratio of an individual ν, a letter that no chapter uses.

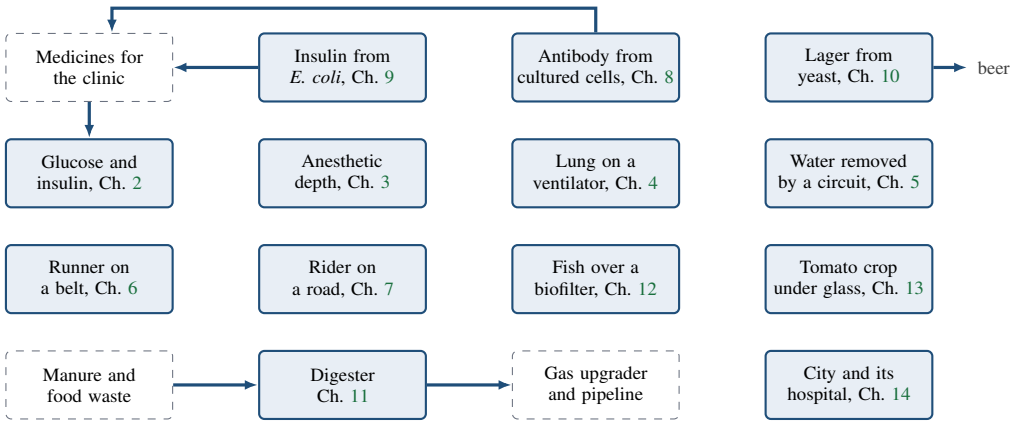


Figure 1.2 The Plants of the Volume. Filled Blocks Are the Plants of Chapters 2 to 14, Dashed Blocks Lie Outside the Volume. Top Row: Cells Making Medicines and Food. Second Row: the Patient at the Bedside, Who Receives the Insulin. Third Row: People Who Exercise and a Farm's Living Stock. Bottom Row: Microbes That Turn Waste Into Gas, and a City in an Epidemic.

1.4 Models, Variability and Limits of Living Plants

1.4.1 The Flow Scheme

Figure 1.2 places each chapter's plant. The top row is the manufacture of medicines and food by cells: an antibody secreted by cultured mammalian cells, insulin refolded from a bacterial precursor, and beer fermented by yeast. The insulin goes to the second row, the patient at the bedside, where a pump delivers it under a glucose monitor and other pumps deliver an anesthetic, a ventilator breathes and a circuit removes water. The third row holds people who exercise and a farm's living stock, fish over a biofilter and a tomato crop under glass. The bottom row is microbes that turn waste into gas, and a city whose epidemic fills its hospital.

Two running examples carry the rest of this section, because no single plant has both a bloodstream and a growing culture. *P-1* is a patient receiving a drug by intravenous infusion. Its central compartment has $V_1 = 10\,\mathrm{L}$ and an elimination clearance $\mathrm{Cl}_1 = 0.5\,\mathrm{L/min}$, its peripheral compartment has $V_2 = 40\,\mathrm{L}$ and exchanges with the central one at $\mathrm{Cl}_2 = 0.4\,\mathrm{L/min}$, and its effect site follows the plasma with $k_{e0} = 0.25\,\mathrm{min}^{-1}$. The target is 2 mg/L, inside a teaching window of 1.5 to 4 mg/L. *C-1* is a fed-batch culture that starts at 10^6 cells per milliliter and 4 g/L of glucose, doubles in 24 h and takes 0.5 g/L of glucose for each 10^6 cells per milliliter it grows. It is counted at 08:00 each day, and the count is reported 4 h later. Every number of both is declared here or computed from these declarations by the script kept with this chapter's record. The numbers are chosen for clear arithmetic and are not taken from any one chapter.

1.4.2 Compartments and Pharmacokinetics

A compartment is a volume that is treated as well mixed. Its balance in concentration is Equation (1.1) for an infusion u in mg/min into one compartment that is cleared at Cl_1.

$$V_1 \frac{\mathrm{d}C_1}{\mathrm{d}t} = u - \mathrm{Cl}_1 C_1, \qquad \tau = \frac{V_1}{\mathrm{Cl}_1}, \qquad C_{1,ss} = \frac{u}{\mathrm{Cl}_1} \tag{1.1}$$

On P-1 alone, the central compartment would have $\tau = 20\,\mathrm{min}$, a maintenance rate of 1 mg/min for the target and a loading dose of $V_1 C^* = 20\,\mathrm{mg}$ to reach it at once. Without the loading dose the plasma would take three time constants, 60 min, to reach 95 % of the target.

Drug also leaves the blood for tissue and returns from it, and the effect is produced at a site that the blood reaches with a lag. The model then adds a peripheral compartment, and an effect site of

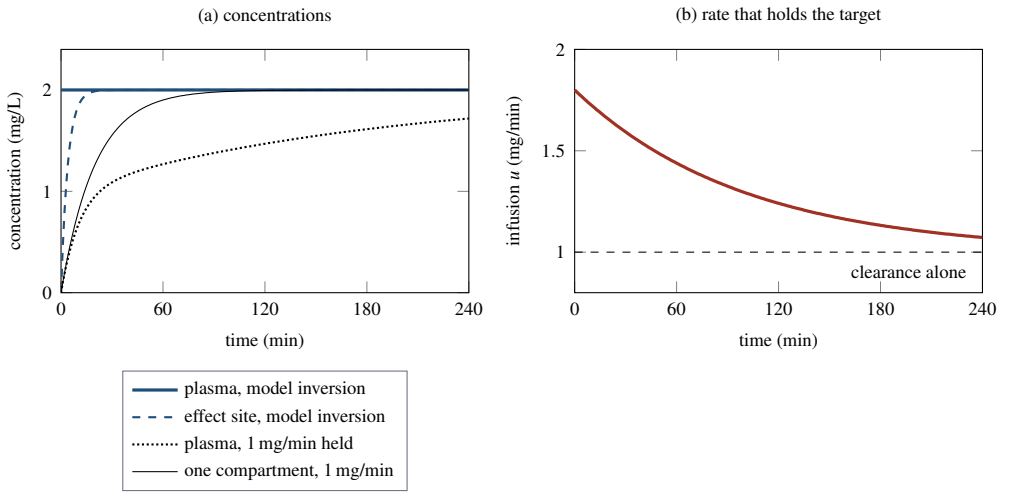


Figure 1.3 P-1 Started by Model Inversion and by a Constant Infusion. (a) Under Model Inversion the Plasma Is at 2 mg/L From the Start (Thick Solid) and the Effect Site Follows It (Dashed). A Constant 1 mg/min Without a Bolus Reaches the Target in About an Hour in One Compartment (Thin Solid) but Is Still Short After Four Hours in Two (Dotted). (b) The Rate That Holds the Target Falls for Hours Toward the 1 mg/min That Clearance Alone Needs.

negligible volume linked to the plasma [13], as in Equation (1.2).

$$\begin{aligned} V_1 \frac{\mathrm{d}C_1}{\mathrm{d}t} &= u - \mathrm{Cl}_1 C_1 - \mathrm{Cl}_2 (C_1 - C_2) \\ V_2 \frac{\mathrm{d}C_2}{\mathrm{d}t} &= \mathrm{Cl}_2 (C_1 - C_2), \qquad \frac{\mathrm{d}C_e}{\mathrm{d}t} = k_{e0} (C_1 - C_e) \end{aligned} \tag{1.2}$$

The two compartments of P-1 have time constants of 10.6 min and 189 min. A constant 1 mg/min without a loading dose gives 1.27 mg/L after an hour and 1.72 mg/L after four, because the peripheral compartment is still filling.

A target-controlled pump inverts the model instead [14]. It gives the bolus $V_1 C^*$ and then the rate of Equation (1.3), which holds $C_1 = C^*$ exactly while C_2 fills with the time constant V_2/Cl_2.

$$u(t) = \mathrm{Cl}_1 C^* + \mathrm{Cl}_2 C^* \, \mathrm{e}^{-\mathrm{Cl}_2 t / V_2} \tag{1.3}$$

On P-1 the rate starts at 1.80 mg/min, falls to 1.44 mg/min after an hour and to 1.07 mg/min after four, and only then nears the 1 mg/min that clearance alone needs, as Figure 1.3 shows. The plasma is at the target from the first second, while the effect site reaches 95 % of it at $\ln 20 / k_{e0} = 12.0$ min.

Chapter 3 is this example at full scale, with three compartments and an effect site for each of two drugs. Its slow compartment fills with a time constant of 285 min, so the rate that holds a maintenance concentration falls from 3.43 to 2.84 mg/kg/h over four hours toward 2.39 mg/kg/h, and the pump inverts a copy of the model that never sees the patient. In Chapter 2 the remote insulin action follows plasma insulin exactly as an effect site follows plasma, with a time constant $1/p_2 = 81.3$ min behind a plasma clearance of $1/k_e = 11.1$ min, on the glucose and insulin compartments of the minimal model [15]. In Chapter 4 the lung is one compartment of compliance C behind a resistance R, with $RC = 0.5$ s, and Chapter 5 closes its balance on a weighed running total that integrates the net removal. The physiological states of Chapter 6 and Chapter 7 follow the belt or the effort through lags in series, reaching 63 % of a belt step in 40, 96, 257 and 475 s in the first. The compartments of Chapter 14 hold people rather than drug, from susceptible to removed [16], and Chapter 10 splits one tank into three exchanging zones.

1.4.3 Growth Kinetics

Cells grow at a specific rate μ that depends on their substrate S, in the saturating form of Equation (1.4) [17], and they consume substrate in proportion to their growth through a yield Y.

$$\mu = \mu_{max}\frac{S}{K_S + S}, \qquad \frac{\mathrm{d}X}{\mathrm{d}t} = (\mu - k_d)X, \qquad q = \frac{\mu}{Y} \tag{1.4}$$

where X is the cell density, k_d a death rate and q the uptake per cell. A specific rate is the property that makes a living plant unlike a reactor. The demand for substrate is proportional to the cells, which multiply, so a constant feed that suits today falls short tomorrow. C-1 grows at $\mu = \ln 2/24 = 0.0289\,\mathrm{h}^{-1}$ while glucose is plentiful, and with $K_S = 0.05\,\mathrm{g/L}$ it still grows at 95 % of that rate at 1 g/L. The glucose it takes in day k is $X_0(2^{k+1} - 2^k)/Y$, so the daily uptake runs 0.5, 1, 2, 4 and 8 g/L over its first five days. Chapter 8 measures the same appetite on a real growth law, 0.53 to 4.27 g/L a day over five days with each day 44 to 94 % above the last, and Chapter 10 shows the heat of fermentation growing with the yeast: with the valves shut its beer warms 0.56 °C in the first day and runs away later.

Some substrates inhibit the organism that consumes them. The Haldane form of Equation (1.5) rises, peaks at $S_M = \sqrt{K_S K_I}$ and falls [18]. In a continuous vessel with dilution rate D, a steady population needs $\mu(S) = D + k_d$.

$$\mu = \frac{\mu_{max}\,S}{K_S + S + S^2/K_I}, \qquad \mu(S_M) = \frac{\mu_{max}}{1 + 2\sqrt{K_S/K_I}}, \qquad D_c = \mu(S_M) - k_d \tag{1.5}$$

Below D_c there are two roots. The lower one is a stable steady state and the upper one bounds the basin from which the population washes out. Above D_c there are none, and the steady states on either side of D_c form a fold. With the methanogen constants of Chapter 11, $\mu_{max} = 0.164\,\mathrm{d}^{-1}$, $K_S = 20$, $K_I = 130\,\mathrm{mmol/L}$ and $k_d = 0.02\,\mathrm{d}^{-1}$, the equation gives $S_M = 51.0\,\mathrm{mmol/L}$, $\mu(S_M) = 0.0919\,\mathrm{d}^{-1}$ and $D_c = 0.0719\,\mathrm{d}^{-1}$, a feed of 287.6 m^3/d to its 4000 m^3 tank. That is the fold the chapter reports, 1.80 times its design feed. At the design feed of 160 m^3/d the two roots are 12.2 and 213 mmol/L of acids, and they meet at S_M as the feed approaches the fold.

Two other chapters use rates of a different shape. In Chapter 12 the biofilter is two nitrifier populations that grow to the load they carry, so a larger ration waits for the slower population to grow and the nitrite returns days after the feed changes. In Chapter 14 the infections grow at a rate r that satisfies $(1 + r/\sigma)(1 + r/\gamma) = R_0$ for exponential latent and infectious stages [19], which doubles them every 4.69 d at $R_0 = 3$. Chapter 9 sets a first-order fold against a second-order aggregation of the same protein, so a charge of unfolded protein U_0 added at once folds the share $\ln(1 + \alpha)/\alpha$ with $\alpha = k_a U_0/k_f$, 75.3 % at $\alpha = 0.72$, and a charge fed slowly folds more.

1.4.4 Variability Between Individuals

A population model describes the typical individual and the spread around it [20], and a loop that inverts it is exact only on the typical individual. Let an individual's clearance be ν times the population's, with the volumes the same. A pump that holds the model at C_T then delivers, in the steady state, a concentration C_T/ν, as in Equation (1.6). The window $[C_{lo}, C_{hi}]$ holds every individual with ν between ν_{lo} and ν_{hi} only if one target satisfies both ends, which needs the population's spread to be no wider than the window.

$$C_{1,\infty} = \frac{C_T}{\nu}, \qquad \nu_{hi}C_{lo} \le C_T \le \nu_{lo}C_{hi} \;\Rightarrow\; \frac{\nu_{hi}}{\nu_{lo}} \le \frac{C_{hi}}{C_{lo}} \tag{1.6}$$

Take three patients of P-1 with $\nu = 0.6$, 1.0 and 1.6. Held at the 2 mg/L target by the population model, they head for 3.33, 2.00 and 1.25 mg/L, and the third leaves the window within the first hour. The population's spread, $1.6/0.6 = 2.67$, equals the window's, $4/1.5 = 2.67$, so open-loop dosing could hold all three only at a target of exactly 2.4 mg/L, with no margin at either end.

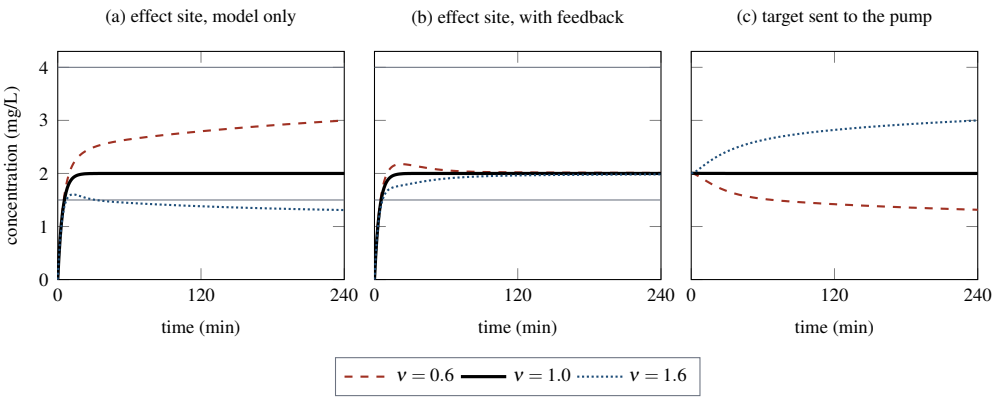


Figure 1.4 Three Patients of P-1 Whose Clearance Is 0.6 (Dashed), 1.0 (Thick Solid) and 1.6 (Dotted) Times the Population's. (a) The Pump Inverts the Population Model and Holds It at 2 mg/L: Only the Population Patient Reaches the Target, and the Fast Clearer Leaves the Window (Thin Lines at 1.5 and 4 mg/L). (b) An Outer PI Loop on the Measured Effect Site Brings All Three to the Target. (c) The Target That Loop Sends to the Pump Moves Toward ν Times 2 mg/L.

Feedback from a measurement of the effect removes the error. In Figure 1.4 an effect monitor reads the effect site once a minute, and an outer PI loop ($K_c = 0.5$, $\tau_I = 8$ min) moves the target sent to the model inversion. It compares the reading with the effect site the model predicts for the population patient, so it acts only on what the model did not predict and leaves the population patient alone. All three patients are within 5 % of the target for good from 45 min, 12 min and 58 min, and after four hours the targets the loop sends are 1.32, 2.00 and 3.00 mg/L, on their way to the νC^* of 1.2, 2.0 and 3.2 mg/L. The commanded target is a parameter estimate, made by an integral that nobody told which parameter was wrong. On the slow clearer the loop's corrections stop the pump for 4.8 min in all, a reminder that a downward correction is made by the patient's own elimination.

Chapter 3 measures exactly this. Under model inversion alone its settled depth index spans 52.1 to 70.4 as the patient's sensitivity moves from 0.70 to 2.0 times the population's, and with the depth loop closed it spans 50.6 to 51.7 while the commanded concentration moves from 1.90 to 5.03 µg/mL. Identifying a patient takes about half an hour, because the outer loop must be slower than the effect-site lag, a pace published closed loops also show [21]. Other chapters use other remedies. Chapter 6 carries its gains as a fraction of the runner's own threshold pace, a schedule on the one number known for each runner, because its plant gain moves by a factor of 2.1 across four runners. Chapter 7 scales the power loop by the rider's critical power so that its loop gain below the sustainable exertion is 0.667 for every rider. Chapter 4 identifies the lung on every breath, and Chapter 10, Chapter 8 and Chapter 9 correct a declared model from laboratory samples, the subject of the next subsection. Variability that no measurement reveals is the hardest kind. Exercise raises insulin sensitivity in Chapter 2 with no change in any measured quantity, and under a hidden inhibition in Chapter 11 a predictive feed controller loses to the sampled PI loop.

1.4.5 Sparse and Late Measurements

A culture that doubles between samples takes more in each interval than in all the earlier ones together. A bolus sized from a sample restores the substrate that the sample had, not what the broth has when the report arrives, so the broth ends each interval below the target by what the cells took since the draw. The interval's uptake must fit inside the band, as in Equation (1.7).

$$\frac{1}{Y}\int_{t_k}^{t_{k+1}} \mu X \, \mathrm{d}t \;\le\; G_{hi} - G_{lo} \tag{1.7}$$

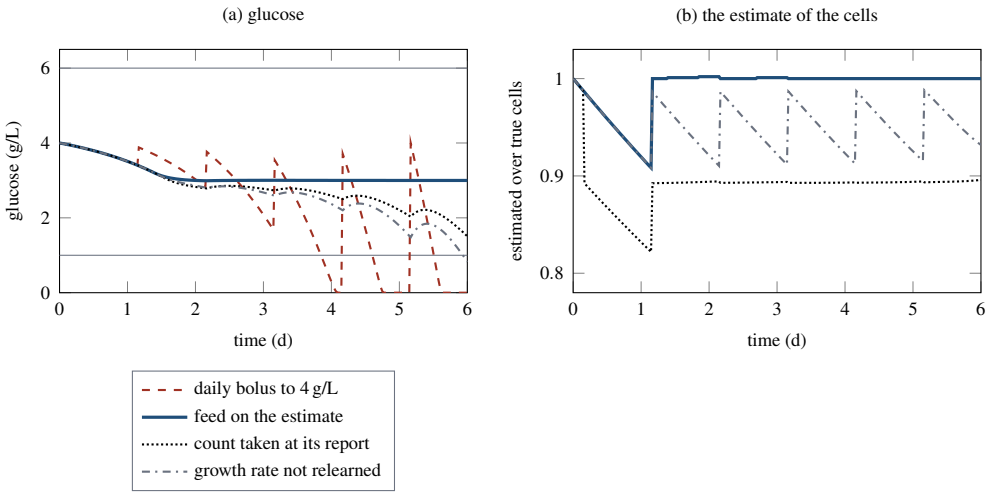


Figure 1.5 C-1 Over Six Days. (a) A Daily Bolus Sized From the Morning Sample (Dashed) Leaves the Band of 1 to 6 g/L (Thin Lines) on Day 3.85 and Runs Out on Day 4.06. A Feed on the Estimated Uptake, Corrected at the Time Each Count Was Drawn, Holds 3 to 3.5 g/L (Solid). The Same Feed With Each Count Applied as if Taken at Its Report Runs Low (Dotted), and Without Relearning the Growth Rate It Leaves the Band on Day 5.93 (Dash-Dotted). (b) The Estimated Cells Over the True Cells for the Three Feeds: Exact After Day 1, About 11 % Low Throughout, and a Sawtooth Corrected at Each Report.

C-1 fed a daily bolus to 4 g/L fails that test on its fifth day, when the day's uptake of 8 g/L exceeds the 5 g/L band of 1 to 6 g/L. The test is necessary and not sufficient. With its target below the top of the band and a four-hour report, the bolus fails a day earlier. Its morning samples read 4.00, 3.51, 3.03, 2.08 and 0.29 g/L, and in Figure 1.5 it leaves the band on day 3.85 and runs out of glucose on day 4.06. Starved for 21.9 h in six days, it ends with 31.2 million cells per milliliter against 59.9 million for the feed described next.

The remedy is to feed what the cells are estimated to take, from a model that runs between samples and is corrected by each one at the instant the sample was drawn [22, 23]. Equation (1.8) is the correction in its logarithmic form, with the sample drawn at t_s and reported later.

$$b \leftarrow b + \lambda \left[\ln X_{lab} - \ln \hat{X}(t_s) - b \right], \qquad \hat{X}_c(t) = \hat{X}(t)\, \mathrm{e}^{b} \tag{1.8}$$

where b is the bias and $0 < \lambda \leq 1$ its weight. With a constant model error the bias closes the fraction $1-(1-\lambda)^k$ of it after k samples. C-1's feed uses $\lambda = 1$, starts from a model growth rate of $0.025\,\mathrm{h}^{-1}$, which is 13 % slow, and relearns the rate from the last two counts. It holds glucose between 2.99 and 3.51 g/L from day 1 to day 6. The time stamp matters. Applying each count as if it had been drawn at the report leaves the estimate low by $\mathrm{e}^{-\mu\tau}$, here $\mathrm{e}^{-0.0289\times 4} = 0.891$, and the glucose then falls to 1.51 g/L. Keeping the slow model growth rate instead of relearning it leaves the estimate up to 9 % low before each report, a share of an uptake that doubles daily, and the glucose leaves the band on day 5.93.

Several chapters read a key variable this way. Chapter 8 counts its cells once a day with a four-hour report, and a bolus to 4 g/L leaves its culture outside the band for 34.8 h, while a culture held warm until day 8 eats up to 5.93 g/L in a day, more than its band is wide. Chapter 10 corrects a diacetyl model by a result that arrives a day after its sample, and Chapter 9 corrects its unfolded protein by each assay at the hour of the draw. In Chapter 11 the titration drawn at 08:00 is reported six hours later, the gas of an overloaded tank peaks on day 2.63 and the titration warns on day 3.58, and a loading gain of 30 swings the feed between its limits where a gain of 8 settles. The reports of

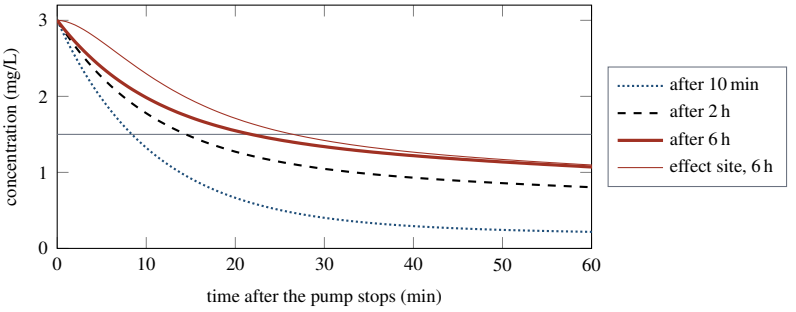


Figure 1.6 P-1 Held at 3 mg/L for Ten Minutes (Dotted), Two Hours (Dashed) and Six Hours (Thick Solid), Then Left With the Pump Stopped. The Plasma Reaches 1.5 mg/L (Thin Gray Line) After 8.4, 14.5 and 21.8 min, and After Six Hours the Effect Site (Thin Solid) Takes 26.5 min.

Chapter 14 count infections about eleven days old, scaled by an unmeasured share, and the hospital census, counted precisely, is what lets all three of its estimators tell a testing cut from a turn in the epidemic [24]. Chapter 5 shows the limit of any one measurement: a balance loop that holds a drifting scale moves the patient to keep the scale right, and only the pumps' own counts, a second measurement of the same flows, reveal the drift.

> **DID YOU KNOW? Five to Ten Minutes**
>
> A federal manual of 1976 told the operators of sludge digesters to run the pH and temperature tests within 5 to 10 min of drawing a sample. A sample left standing too long releases carbon dioxide, and its pH rises [7]. A late result can also be a wrong one.

1.4.6 One-Sided Actuators, Stores and Safety Limits

When a pump can only add, the way down is the plant's own removal. In one compartment a stopped pump lets the concentration fall as in Equation (1.9), and with a peripheral compartment the fall slows the longer the infusion has run, because the tissue returns what it took [25].

$$t_{lim} = \tau \ln \frac{C_0}{C_{lim}} \tag{1.9}$$

For P-1 held at 3 mg/L, one compartment would need $20\,\text{min} \times \ln 2 = 13.9\,\text{min}$ to fall to 1.5 mg/L. With the peripheral compartment, Figure 1.6 gives 8.4 min after a ten-minute infusion, 14.5 min after two hours and 21.8 min after six, and the effect site takes 4.7 min longer again. Raising the target from 1.5 to 3 mg/L takes a 15 mg bolus, under two seconds at a pump capacity of 500 mg/min. The actuator moves the plant up in seconds and leaves it to come down in tens of minutes.

The chapters meet the same asymmetry. In Chapter 2 exercise drops glucose from 74.9 to 47.8 mg/dL within 0.04 h, before the controller's next execution, and the pump then sits at zero for 0.84 h while the insulin already given goes on acting. Deployed systems therefore suspend delivery on a prediction [26] or bound the insulin on board [27]. In Chapter 3 the depth index is lowest a quarter of a minute after the stimulation stops, sooner than any infusion change can reach the effect site. The glycol valves of Chapter 10 can only cool a tank whose yeast make the heat, and the inspiratory valve of Chapter 4 can only add gas, so no controller holds the airway pressure against a leak larger than the open valve passes.

> **DID YOU KNOW? A Twentieth of a Unit an Hour**
>
> Insulin pumps let the programmed delivery be changed in increments as small as 0.05 to 0.25 units of insulin an hour [2]. Like the pump of P-1, it still cannot take back a unit it has delivered.

Because the two sides of a limit carry different costs, a symmetric error integral misjudges such a loop. Glucose is reported as time in range with separate thresholds below [28], and an objective of the form of Equation (1.10) weights the low side by w.

$$J = \int \max(0,\, y - y_{hi})\,\mathrm{d}t + w \int \max(0,\, y_{lo} - y)\,\mathrm{d}t \tag{1.10}$$

Chapter 2 finds that a run that falls well below its low flag beats a run that never does on the symmetric score, and that the ranking reverses at $w = 7.9$, the ratio of the two runs' excursion areas. The margin a loop keeps from a hard limit has to cover both the spread of individuals of Section 1.4.4 and the slow way back of Equation (1.9). Where a limit is approached through another variable, an override takes the actuator, as Volume 4 develops. Chapter 6 and Chapter 7 share one actuator among limits on a low selector, Chapter 5 caps its blood pump on access pressure and its ultrafiltration on transmembrane pressure, taking the replacement fluid before the removal, Chapter 12 cuts the ration on a low select of four limits, and Chapter 13 opens a screen gap, a vent and then heat as the humidity nears its limit.

Some actuators are stores. A rider's effort above a sustainable level draws on a work capacity W' above critical power [29], and when the power available falls in proportion to the capacity left [30], a held effort E above the sustainable E_c fades onto critical power with the time constant of Equation (1.11).

$$\tau_W = \frac{W'_0\,(1 - E_c)}{(P_{pk} - CP)\,(E - E_c)} \tag{1.11}$$

For the recreational rider of Chapter 7, with $W'_0 = 16\,\mathrm{kJ}$, $P_{pk} = 950\,\mathrm{W}$, $CP = 220\,\mathrm{W}$ and $E_c = 0.75$, it gives 21.9 s at full effort and 54.8 s at $E = 0.85$, the chapter's values. The actuator loses its upper range without moving. Chapter 6 drains the same reserve at exactly the excess power, and Chapter 11 spends its alkalinity on the acids it neutralizes.

1.4.7 How Each Loop Handles Variability and Limits

Each chapter's plant can be read in three parts: what varies between individuals or during the run, the limit that may not be exceeded, and the structure that keeps the loop inside it. Table 1.1 lists the three for each chapter. Chapter 15 adds the measured results and compares how each loop handled variability and limits it could not exceed.

> **DID YOU KNOW? Half a Liter an Hour**
>
> A clinical report of 1978 on automatic hemofiltration found that low blood pressure, nausea and muscle cramps during treatment could be reduced, provided that the effective fluid withdrawal did not exceed 0.5 L/h [12]. The reference prescription of Chapter 5 removes 100 mL/h.

1.5 Instruments and Final Elements in This Volume

Figure 1.7 lists what the controllers of each chapter read and what they move. Eight chapters close a loop on a model's estimate rather than on a reading, and six wait for a laboratory sample, an assay or a report, more than read any single kind of sensor. A pump or feeder is the commonest handle, in seven chapters, and two chapters move a person: a rider who follows a cue and a city that adopts a measure.

A monitor on the body reads a signal that stands for the variable that matters. A continuous glucose monitor reads the glucose of the fluid under the skin every 1 to 5 min, and its reading lags blood

Table 1.1 What Varies, the Limit and the Structure That Holds It, by Chapter

Ch.	What varies	Limit	Structure
2	Insulin sensitivity with exercise, unannounced meals, basal drift	Low glucose flag, pump floor at zero	Direct-acting PI with back-calculation, MPC told of meals, MHE with a meal state
3	Drug requirement by a factor of three, surgical stimulation	Surgical depth band, arterial pressure	Depth loop correcting a model-inverting pump, pressure loop, MPC with learned potency
4	Compliance, resistance, effort, leak	Plateau pressure, volume per kilogram	Per-breath least squares, protective cap from the identified compliance
5	Pump tubing, wear, scale drift, filter fouling	Transmembrane and access pressure, filtration fraction	Weighed balance loop, overrides, each scale checked against its pump
6	Runner, heat, fuel	Lactate, heart rate, core temperature	Gains scheduled on threshold pace, low selector with external reset, MPC with terminal value
7	Rider, a work capacity that drains	Heart rate, lactate, exertion range	Power loop, two overrides with margin-scaled reset, MPC
8	Appetite that doubles, cooled cells, probe drift	Glucose band never at zero, osmolality	Feed on estimated uptake, daily correction, hourly Raman trim, MPC
9	Folding constants, harvest load	Unfolded ceiling, hold times, lot limits	Scheduler on an assay-corrected estimate, transfer and endpoint rules
10	Yeast strain, mixing that stops	Recipe temperatures, diacetyl, shared glycol	Time-proportioned PI per zone, corrected estimator, glycol coordinator
11	Feed, temperature, hidden inhibition	Titration ratio at the fold	Sampled PI on a daily titration, economic supervisor, MPC
12	Biofilter capacity, fish biomass	Nitrite, ammonia, oxygen	Oxygen cascade with meal feedforward, feed governor, soft sensor, film observer
13	Weather and its forecast	Hourly humidity limit	Pipe cascade with external reset, three-stage humidity override, planner over the forecast
14	Reported share, testing, a variant	Hospital beds	Estimators anchored by the census, PI on the census, MPC with a margin, scenario tree

glucose by 5 to 10 min, a delay that matters most when the reading is used for decisions in real time [31].

Weight is the measurement a kidney replacement machine trusts most. Weighing with one or more scales is the commonest way to balance its fluids, because it is the most reliable over long treatments, and the weight stands in for the flow [32]. A scale rests on a load cell, strain gauges bonded to a metal element that stretches and compresses with the applied force [33] (Figure 1.8), the gauge that also sits in the power meter of Chapter 7.

In United States regulation an infusion pump moves fluid into a patient in a controlled manner by a piston, roller or peristaltic mechanism, and it may detect air in the line or a blocked line and sound an alarm [34]. Each of those mechanisms only pushes, which is why the pumps of Chapters 2, 3 and 5 are the one-way actuators of Section 1.4.6. The glycol valves of Chapter 10 are on/off solenoid valves, opened and closed by the current through a coil [33], so the loop on each zone works through the share of time its valve is open.

Figure 1.9 places the volume's loops on their platforms. Supervisory and lower loops in industry run with cycle times from milliseconds to minutes [35]. The apps span a wider range, from the 2 ms servo of Chapter 4 to the daily policy of Chapter 14, and several loops have a person above them. While a target-controlled pump runs, the clinical response and the depth of anesthesia are monitored, and closed loops on such an endpoint are still experimental [10].



Figure 1.7 Sensors and final elements by chapter: a round mark where the chapter's controllers read the device and a square mark where they move it, as its variables table and control structure name them. A flow set through a flow loop counts as a valve, and the last column of reads marks a controlled variable that is a model's estimate rather than a reading. Notice that the estimate and the sample outnumber every sensor.

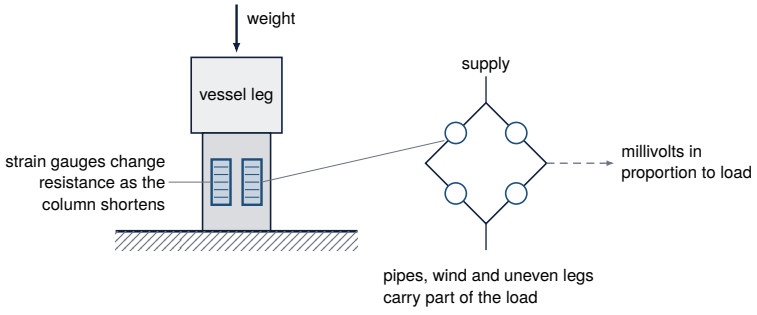


Figure 1.8 Principle of a strain-gauge load cell. Notice that a load that reaches the support by another path is not weighed, one way a scale can drift from the fluid it holds, the fault that Chapter 5 finds by comparing each scale with its pump.

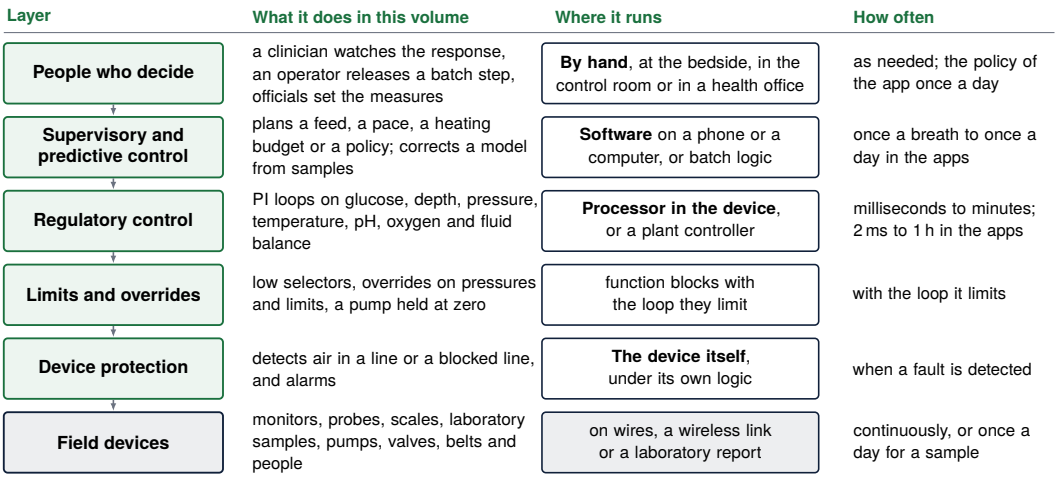


Figure 1.9 The control layers of this volume, the platform that hosts each and how often each acts. Loop periods are from [35] and the app periods from the chapters' technology sections. Notice that the top layer is a person, not a program.

Table 1.2 What Every App Provides. Standalone Apps Are Described in Their Own Chapters.

Element	What it does
Run, Pause, Reset	Start, hold and restart the run. Reset keeps the reader's settings.
Speed	Simulated time per second of real time. It changes the pace and never the result.
Manual and Auto	Hand each actuator to the reader or to the controller, without a bump.
Scenario list	Numbered cases, each with a briefing, a completion rule and a reload button.
Control panel	Setpoints, manual outputs, tuning, options and disturbances, at a Beginner or an Advanced level.
Trends	SP and PV above the MV for each loop, with limits, event marks and a cursor.
Scorecard	IAE, overshoot, settling time, MV travel and time at a limit for each loop.
Learn card	Predict, Try It, two hints and Explain, opened by the reader in that order.
Show Model	Equations, live variable values, editable parameters and assumptions.
Export CSV	Every logged sample of the run as a text file with a header row.
Instructions	A separate page with background, an operating guide and the model's limitations.

1.6 The Simulator Framework

Every chapter of this book has an app that runs in a web browser, collected with the others at https://apmonitor.com/apps/. About 240 apps exist across the series. All but a few are built on one shared framework, so the engine, the PID algorithm, the trends and the data export behave the same way in every chapter. The exceptions are standalone studios, such as the cascade studio and the two Temperature Control Lab apps, which follow the same ideas in their own code. This section describes the framework once and states where the standalone apps differ. Table 1.2 lists what a reader meets.

1.6.1 The Simulation Engine

Each app carries a plant model written as ordinary differential equations $\frac{dx}{dt} = f(x,u,d,p)$ in the states x, the inputs u, the disturbances d and the parameters p. The engine integrates the model with the classical fourth-order Runge–Kutta method on a fixed step h set for the fastest dynamics of the app. Some apps supply their own step, such as explicit Euler for a discontinuous right-hand side or an exact update for a model driven by events.

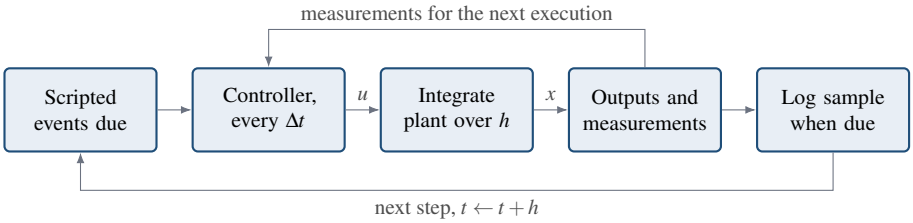


Figure 1.10 Work of the Engine in One Integration Step

Three clocks run at separate rates. The plant advances every step h. The controller executes every control period Δt, equal to h unless the app declares a slower controller, and holds its outputs between executions. The log takes one sample at a fixed interval of simulated time for the trends, the scorecard and the exported file. The order of work in one step is shown in Figure 1.10. The controller also executes once when a scenario is loaded, so every run starts from an initialized controller.

Drawing the screen is kept apart from these clocks. A speed of 5 means five units of the app's time base per second of real time. Speed, frame rate and pauses change how fast a run is shown and never what it computes, so a phone and a workstation produce the same trajectory and the same file.

Noise comes from a deterministic pseudorandom generator. Its seed is declared by the app, or by a scenario that needs its own noise record, and it is restored whenever a scenario is loaded or Reset is pressed. The framework gives the reader no seed setting, so *default seed* in this book means the seed the app ships with. A run repeats bit for bit when the same scenario is loaded and the same settings are made before Run. A change made by hand during a run lands on a different step each time, which is one reason numeric answers carry a tolerance. The standalone simulators draw noise from the browser's own generator, so a noisy run there repeats in its statistics and not sample for sample.

Run and Pause start and hold the clock. Reset restarts the run at $t = 0$, paused, and keeps the reader's settings: every setpoint, tuning value, option, model parameter, disturbance slider and mode the reader changed is put back before the plant and the controller are initialized. Scripted events replay, and a one-shot disturbance button pressed during the run is cleared. The reload button beside the scenario list restores the scenario's own settings instead.

1.6.2 The Velocity-Form PID Algorithm

One PID implementation serves the PID loops of every framework app. It is the velocity (incremental) form: each execution adds the change in the proportional and derivative terms to a running sum, and the integral is a separate term, as shown in Equations (1.12) to (1.15) for execution k and control period Δt.

$$e_k = a\,(\mathrm{SP}_k - \mathrm{PV}_k), \qquad P_k = K_c\,e_k, \qquad D_k = -a\,K_c\,\tau_D\,\frac{\mathrm{PV}^f_k - \mathrm{PV}^f_{k-1}}{\Delta t} \tag{1.12}$$

$$S_k = S_{k-1} + (P_k - P_{k-1}) + (D_k - D_{k-1}) \tag{1.13}$$

$$v_k = u_b + S_k + I_{k-1}, \qquad u_k = \mathrm{sat}(v_k) \tag{1.14}$$

$$I_k = I_{k-1} + \left[\frac{K_c}{\tau_I}\,e_k + \frac{u_k - v_k}{T_t}\right]\Delta t \tag{1.15}$$

where $a = \pm 1$ is the controller action, PV^f is the measurement after an optional first-order filter, u_b is the output bias, S is the running sum of proportional and derivative increments, v is the output before limiting, u is the output sent to the actuator, I is the integral term and T_t is the tracking time constant, equal to τ_I unless a loop sets another value. The function sat clamps v_k to the actuator range and then applies a rate limit if the loop declares one.

The limit acts on the emitted output u_k only. The sum S_k is never clamped, so it equals $P_k + D_k$ and the output before limiting is the position form $u_b + K_c e_k + D_k + I_{k-1}$. Anti-windup touches the integral alone. While the output is limited, the last term of Equation (1.15) drives I toward the value that reproduces the output applied, which is back-calculation [36]. With no integral action, Equation (1.15) is skipped. Clamping the running sum would instead discard each increment that crosses a limit and keep each one that returns, so symmetric noise would push the output off the limit even in a proportional-only loop.

The derivative acts on the measurement, so a setpoint step gives a proportional kick and no derivative kick. A change of K_c acts on the whole error at the next execution, as in the position form. The action $a = +1$ suits a process whose PV rises with the MV, a reverse-acting controller in the ISA convention, and $a = -1$ gives the direct-acting one.

In Manual the controller keeps executing and sets $I_k = u_{\text{man}} - u_b - S_k$, so the output it would send equals the manual value u_{man}. The first automatic output differs from it by one period's change in P and D. In Auto the engine copies each controller output to the manual station, so a switch to Manual is also bumpless. The standalone studios with PID loops use velocity-form code of their own, with the limit on the output and anti-windup on the integral alone, in which a gain change applies only to later increments.

1.6.3 The Predictive Control and Estimator Options

More than 100 apps offer model predictive control (MPC) as an option beside the structure they open with, which stays the default. Volume 9 develops the theory. In most of them the controller plans with a small linear model in discrete time as shown in Equation (1.16).

$$x_{k+1} = A\,x_k + B\,u_k + E\,w_k + c, \qquad y_k = C\,x_k \tag{1.16}$$

where w is a measured or scheduled disturbance and c is a constant left by linearization. The model comes from a step test at the design point, from the app's balances linearized there or again at every decision, or from the model the reader fits. It never reads the true states or parameters of the run, so a parameter the reader changes in Show Model is model mismatch, as on a real plant.

At each decision the controller minimizes the objective of Equation (1.17) over a horizon of N steps.

$$J = \sum_{k=1}^{N} (y_k - r_k)^{\top} Q\,(y_k - r_k) + \sum_{j} \Delta u_j^{\top} R\,\Delta u_j + \sum_{i} \rho_i\, s_i^2 \tag{1.17}$$

where r is the setpoint, Q weights the tracking error, R is the move suppression on each input change Δu_j and ρ_i is the price of a slack $s_i \geq 0$ that lets an output band be broken when it cannot be held. Actuator limits are hard constraints. The quadratic program is solved in the browser by a dual active-set method [37]. The first move is applied, and the plan is made again at the next decision.

The plan starts from an estimate of the model state. About half of these apps use a steady-state Kalman filter on the model with an added disturbance state, which absorbs what the model did not predict and removes steady offset without an integral term [38]. Others correct the model output with a filtered bias to the same end. About a dozen apps also offer moving horizon estimation (MHE), which solves the estimation problem over a window of recent samples and can hold bounds that a Kalman filter cannot, such as a concentration that is never negative [39].

With the MPC selected, the tuning fields of the loops it replaces leave the Control panel. In most of these apps the Advanced level shows up to three settings, each a multiple of the shipped value: the prediction horizon, the move suppression and the band price. The trend draws the accepted plan ahead of the present time, and a list of constraint prices shows which limits bind. A plan that is not optimal, not finite or outside its limits is rejected, and a fallback, in most apps the shipped controller, holds the actuator until a plan succeeds. The Model Fitting and Tuning Studio builds the same kind of controller from the reader's own data in its tab, MPC from the fit.

1.6.4 Scenarios, Progress and Data

An app teaches through scenarios, the numbered cases at the top of the Control panel, nine to thirteen in most apps. Loading one sets the model parameters, the controller settings, the disturbances, the mode, the initial state and any scripted events. The information button beside the list opens a briefing in four parts: what the case is, why it matters, what to do and one detail for the practicing engineer. The Learn card holds the lesson back, with Predict and Try It before the run, hints on request and Explain afterward.

Each scenario has a completion rule that the app judges from the run on the simulated clock: a run length, the scripted events passed, a setting made by the reader, a state held for a stated time or a target met. Choosing a case earns nothing. A completed case is marked in the list, and a badge row shows the count and a level, Explorer after one case, Practitioner at half and Completed at all. Progress is kept in the reader's browser with no account. A certificate of progress or of completion can be issued from it, as a record of self-paced work and not an assessed credential.

One case, usually the last, is the graded challenge. It states a target, such as an error band to hold through a sequence of upsets. Changing the model parameters or the disturbances voids the attempt in most apps, and a record made with Reveal suggested tuning says so.

Show Model opens the equations, live values of the variables, the editable parameters and the assumptions. The trends plot SP and PV above the MV for each loop, with actuator limits and marks at setpoint moves, disturbances, mode changes and scripted events. The scorecard reports for each loop the integral of the absolute error (IAE), overshoot, settling time, MV travel and share of time at a limit, with the app's own measures, and it does not combine them into one score. IAE and travel accumulate from Reset, so two runs are compared at the same clock reading.

Export CSV saves the log as a text file with a header row of column names that carry their units and one row per logged sample. In most apps the time column is in seconds, minutes or hours according to the run length. The log holds the most recent 60 000 samples unless the app sets another size. Apps with the MPC option begin the file with comment lines that start with #, which a script must skip.

An app exercise in this book names the scenario by its number and name as the app shows them, every setting that differs from the scenario's defaults, the run length in simulated time, the seed (the default seed unless stated), what to read and where (a trend, a scorecard tile or a column of the exported file) and a tolerance on a numeric answer. Settings are made before Run unless the exercise gives a time.

1.7 Working With AI Agents While Studying

A reader of this book is likely to have an AI assistant within reach, and the apps make its work easy to test. This section gives the editor's view of how to use one while working through a chapter. In short, the assistant may do the algebra and the scripting, and the reader keeps the checking.

What to delegate. Tasks with a clear input and a result that can be tested are good candidates. An assistant can derive the linearization of a chapter's balance at an operating point and the transfer function that follows from it. It can write a short script that reads the exported CSV file and fits a first-order-plus-dead-time model to a step test. It can propose a first tuning from that fit with a named rule, convert units, explain a term in a model or outline a report. The assistant knows nothing about the app until it is told. Give it the equations and parameter values from Show Model, the column names of the exported file and the scenario that produced the data, as described in Section 1.6.4. Ask it to state its assumptions and the units of every result.

What to check, and how. An assistant returns a fluent answer whether or not the answer is right, so each result needs a test that it could fail. Four tests cover most of the errors a chapter will produce.

Table 1.3 Roadmap of the Volume

Ch.	Plant	The lesson	Level	Time
2	Artificial pancreas	A pump that only adds, and a cost that is not symmetric.	introductory	3 h
3	Intravenous anesthesia	An exact loop on a model is not a correct loop on a patient.	intermediate	3 h
4	Mechanical ventilator	The measured breath and the delivered breath part company.	advanced	2.5 h
5	Kidney replacement	A loop that holds its measurement need not hold the patient.	intermediate	2 h
6	Runner on a treadmill	One actuator, four limits and a deadline.	intermediate	3.5 h
7	Road cyclist	An actuator whose authority is a store.	advanced	4 h
8	Antibody culture	A doubling appetite fed on a sample a day old.	advanced	3 h
9	Insulin refolding	A better fold can cost the campaign.	intermediate	2 h
10	Lager fermenter	Triggers and probes decide the batch more than tuning.	introductory	2 h
11	Anaerobic digester	More gas can mean a tank near its fold.	introductory	2.5 h
12	Recirculating fish farm	A biofilter is an actuator that has to be grown.	advanced	2.5 h
13	Glasshouse crop	A daily temperature budget under an hourly humidity limit.	advanced	2.5 h
14	City in an epidemic	Estimate before steering on counts a week old.	advanced	2 h

- *Units.* Every term of an equation carries the same units, and a process gain is in PV units per MV unit. A gain computed with the valve as a fraction is 100 times the gain of a loop whose valve is in percent.
- *Signs.* Decide from the physics which way the PV moves when the MV rises. Compare that with the sign of the fitted gain and with the controller action of Section 1.6.2.
- *Steady state.* Set the derivatives to zero. The model must return the operating point that Show Model reports, and a linearization must give zero deviation there.
- *A run on the app.* Load the named scenario at the default seed, enter the proposed numbers and run. The trend and the scorecard decide. A tuning that reads well on paper and holds the valve at a limit for half the run has failed.

A fitted model gets one more test: compare it with data it was not fitted to, such as a second step in the opposite direction. When a check fails, report the failure to the assistant with the evidence and ask for the cause, then check the correction the same way.

How the exercises train the habit. One exercise in each chapter is marked *Review the agent*. It presents a tuning, a model or a claim as an AI assistant might propose it, with a flaw, and asks the reader to find the flaw in about five minutes. The flaws are ordinary ones: a unit, a sign, a steady state that does not close, a result measured on the wrong scenario or at a different run length, or a limit that was ignored. Each is found with one of the tests above. The purpose is not distrust. An engineer who can say why an answer is right can hand an assistant much larger tasks and still stand behind the result. That larger picture, in which agents do most of the routine work and the engineer supervises and checks, is the subject of Chapter 15.

1.8 Chapter Roadmap

Table 1.3 lists the chapters with the lesson each adds to this one. The level says how much of Section 1.4 a chapter assumes. The time is an estimate for reading the chapter once and running each case of its app once, at about eight minutes a case.

Chapters 2, 10 and 11 form a first course: a one-way actuator, an on-off valve with a day-old result, and a fold watched through a daily titration. They rest on Sections 1.4.3, 1.4.5 and 1.4.6. Chapters 3, 5, 6 and 9 add a model inside the loop, redundancy, a selector and a schedule. The six advanced chapters combine estimation or prediction with constraints and are best left until Section 1.4.7 is

familiar. A reader at the bedside can read Chapters 2 to 5 in order, a reader of sport Chapters 6 and 7, and a reader of bioprocesses Chapters 8 to 12, ending with the glasshouse and the city.

KEY POINTS

- A living plant differs from its model, is measured sparsely and late, and is often moved by an actuator that cannot take back what it has given.
- A loop that inverts a population model is exact only on the typical individual, and feedback from a measurement of the effect removes the error.
- A bolus sized from a sample restores what the sample had, so a model between samples is corrected at the time each sample was drawn.
- When a pump can only add, the way down is the plant's own removal, and the margin from a hard limit must cover the spread of individuals and that slow way back.

Where else this applies. Selectors and overrides are the subject of Volume 4, estimation from late samples of Volume 5 and batch operation of Volume 8.

References

[1] R. J. Hatchett, C. E. Mecher, and M. Lipsitch, "Public health interventions and epidemic intensity during the 1918 influenza pandemic," *Proceedings of the National Academy of Sciences*, vol. 104, no. 18, pp. 7582–7587, 2007.

[2] C. Lambert and E. Delgado, "100 years since the discovery of insulin, from its discovery to the insulins of the future," *Biomedicines*, vol. 12, no. 3, p. 533, 2024.

[3] P. Drinker and L. A. Shaw, "An apparatus for the prolonged administration of artificial respiration: I. A design for adults and children," *Journal of Clinical Investigation*, vol. 7, no. 2, pp. 229–247, 1929.

[4] A. Chorley and K. L. Lamb, "The application of critical power, the work capacity above critical power (W'), and its reconstitution: A narrative review of current evidence and implications for cycling training prescription," *Sports*, vol. 8, no. 9, p. 123, 2020.

[5] I. Th. H. Gieling, "Commercial greenhouse computersystems: Inventarisation of implemented research, carried out by commercial firms," *Acta Horticulturae*, no. 106, pp. 59–66, 1980.

[6] G. Köhler and C. Milstein, "Continuous cultures of fused cells secreting antibody of predefined specificity," *Nature*, vol. 256, no. 5517, pp. 495–497, 1975.

[7] C. Zickefoose and R. B. J. Hayes, "Operations manual: Anaerobic sludge digestion," United States Environmental Protection Agency, Office of Water Program Operations, Washington, DC, Tech. Rep. EPA 430/9-76-001, 1976. [Online]. Available: https://permanent.fdlp.gov/gpo81586/00000IKT.pdf

[8] C. Cobelli, E. Renard, and B. Kovatchev, "Artificial pancreas: Past, present, future," *Diabetes*, vol. 60, no. 11, pp. 2672–2682, 2011.

[9] H. Schwilden, J. Schüttler, and H. Stoeckel, "Closed-loop feedback control of methohexital anesthesia by quantitative EEG analysis in humans," *Anesthesiology*, vol. 67, no. 3, pp. 341–347, 1987.

[10] P. U. Bidkar, A. Dey, P. Chatterjee, R. Ramadurai, and J. J. Joy, "Target-controlled infusion – past, present, and future," *Journal of Anaesthesiology Clinical Pharmacology*, vol. 40, no. 3, pp. 371–380, 2024.

[11] C. Ronco, "Continuous renal replacement therapy: Forty-year anniversary," *The International Journal of Artificial Organs*, vol. 40, no. 6, pp. 257–264, 2017.

[12] P. Kramer, W. Wigger, D. Matthaei, C. Langescheid, J. Rieger, C. Fuchs, K.-W. Rumpf, and F. Scheler, "Clinical experience with continuously monitored fluid balance in automatic hemofiltration," *Artificial Organs*, vol. 2, no. 2, pp. 147–149, 1978.

[13] L. B. Sheiner, D. R. Stanski, S. Vozeh, R. D. Miller, and J. Ham, "Simultaneous modeling of pharmacokinetics and pharmacodynamics: Application to d-tubocurarine," *Clinical Pharmacology & Therapeutics*, vol. 25, no. 3, pp. 358–371, 1979.

[14] S. L. Shafer and K. M. Gregg, "Algorithms to rapidly achieve and maintain stable drug concentrations at the site of drug effect with a computer-controlled infusion pump," *Journal of Pharmacokinetics and Biopharmaceutics*, vol. 20, no. 2, pp. 147–169, 1992.

[15] R. N. Bergman, Y. Z. Ider, C. R. Bowden, and C. Cobelli, "Quantitative estimation of insulin sensitivity," *American Journal of Physiology-Endocrinology and Metabolism*, vol. 236, no. 6, p. E667, 1979.

[16] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London. Series A*, vol. 115, no. 772, pp. 700–721, 1927.

[17] J. Monod, "The growth of bacterial cultures," *Annual Review of Microbiology*, vol. 3, no. 1, pp. 371–394, 1949.

[18] J. F. Andrews, "A mathematical model for the continuous culture of microorganisms utilizing inhibitory substrates," *Biotechnology and Bioengineering*, vol. 10, no. 6, pp. 707–723, 1968.

[19] J. Wallinga and M. Lipsitch, "How generation intervals shape the relationship between growth rates and reproductive numbers," *Proceedings of the Royal Society B: Biological Sciences*, vol. 274, no. 1609, pp. 599–604, 2007.

[20] L. B. Sheiner, B. Rosenberg, and V. V. Marathe, "Estimation of population characteristics of pharmacokinetic parameters from routine clinical data," *Journal of Pharmacokinetics and Biopharmaceutics*, vol. 5, no. 5, pp. 445–479, 1977.

[21] A. R. Absalom, N. Sutcliffe, and G. N. Kenny, "Closed-loop control of anesthesia using bispectral index," *Anesthesiology*, vol. 96, no. 1, pp. 67–73, 2002.

[22] R. Luttmann, D. G. Bracewell, G. Cornelissen, K. V. Gernaey, J. Glassey, V. C. Hass, C. Kaiser, C. Preusse, G. Striedner, and C.-F. Mandenius, "Soft sensors in bioprocessing: A status report and recommendations," *Biotechnology Journal*, vol. 7, no. 8, pp. 1040–1048, 2012.

[23] D. Dochain, "State and parameter estimation in chemical and biochemical processes: A tutorial," *Journal of Process Control*, vol. 13, no. 8, pp. 801–818, 2003.

[24] E. L. Haseltine and J. B. Rawlings, "Critical evaluation of extended Kalman filtering and moving-horizon estimation," *Industrial & Engineering Chemistry Research*, vol. 44, no. 8, pp. 2451–2460, 2005.

[25] M. A. Hughes, P. S. A. Glass, and J. R. Jacobs, "Context-sensitive half-time in multicompartment pharmacokinetic models for intravenous anesthetic drugs," *Anesthesiology*, vol. 76, no. 3, pp. 334–341, 1992.

[26] B. W. Bequette, "A critical assessment of algorithms and challenges in the development of a closed-loop artificial pancreas," *Diabetes Technology & Therapeutics*, vol. 7, no. 1, pp. 28–47, 2005.

[27] C. Ellingsen, E. Dassau, H. Zisser, B. Grosman, M. W. Percival, L. Jovanovič, and F. J. Doyle, "Safety constraints in an artificial pancreatic β cell: An implementation of model predictive control with insulin on board," *Journal of Diabetes Science and Technology*, vol. 3, no. 3, pp. 536–544, 2009.

[28] T. Battelino, T. Danne, R. M. Bergenstal, S. A. Amiel, R. Beck *et al.*, "Clinical targets for continuous glucose monitoring data interpretation: Recommendations from the international consensus on time in range," *Diabetes Care*, vol. 42, no. 8, pp. 1593–1603, 2019.

[29] A. M. Jones, A. Vanhatalo, M. Burnley, R. H. Morton, and D. C. Poole, "Critical power: Implications for determination of $\dot{V}O_2$max and exercise tolerance," *Medicine & Science in Sports & Exercise*, vol. 42, no. 10, pp. 1876–1890, 2010.

[30] R. H. Morton, "A 3-parameter critical power model," *Ergonomics*, vol. 39, no. 4, pp. 611–619, 1996.

[31] G. Cappon, M. Vettoretti, G. Sparacino, and A. Facchinetti, "Continuous glucose monitoring sensors for diabetes management: A review of technologies and applications," *Diabetes & Metabolism Journal*, vol. 43, no. 4, pp. 383–397, 2019.

[32] G. Villa, M. Neri, R. Bellomo, J. Cerda, A. R. De Gaudio, S. De Rosa, F. Garzotto, P. M. Honore, J. Kellum, A. Lorenzin, D. Payen, Z. Ricci, S. Samoni, J.-L. Vincent, J. Wendon, M. Zaccaria, and C. Ronco, "Nomenclature for renal replacement therapy and blood purification techniques in critically ill patients: practical applications," *Critical Care*, vol. 20, p. 283, 2016.

[33] T. R. Kuphaldt, *Lessons in Industrial Instrumentation*, version 2.33 ed. Open textbook, Creative Commons Attribution 4.0 license, ibiblio.org/kuphaldt/socratic/sinst, 2024.

[34] U.S. Food and Drug Administration, "Infusion pump," Code of Federal Regulations, Title 21, Section 880.5725, revised as of April 1, 2024, U.S. Government Publishing Office, govinfo.gov, 2024.

[35] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule, and M. Thompson, "Guide to operational technology (OT) security," National Institute of Standards and Technology, Tech. Rep. NIST Special Publication 800-82r3, 2023.

[36] K. J. Åström and L. Rundqwist, "Integrator windup and how to avoid it," in *Proc. 1989 American Control Conference*, 1989, pp. 1693–1698.

[37] D. Goldfarb and A. Idnani, "A numerically stable dual method for solving strictly convex quadratic programs," *Mathematical Programming*, vol. 27, no. 1, pp. 1–33, 1983.

[38] G. Pannocchia and J. B. Rawlings, “Disturbance models for offset-free model-predictive control,” *AIChE Journal*, vol. 49, no. 2, pp. 426–437, 2003.

[39] C. V. Rao, J. B. Rawlings, and D. Q. Mayne, “Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations,” *IEEE Transactions on Automatic Control*, vol. 48, no. 2, pp. 246–258, 2003.

1.9 Exercises

1.1 Theory and calculation · 4 min

A drug is infused into one well-mixed compartment of 20 L that is cleared at 1 L/min. The target concentration is 3 mg/L.

(a) Find the time constant, the maintenance rate and the loading dose. *Checkpoint:* 20 min, 3 mg/min, 60 mg.

(b) Without the loading dose, how long does the concentration take to reach 95 % of the target? *Checkpoint:* about 60 min.

1.2 Theory and calculation · 4 min

Patient P-1 of this chapter has $V_1 = 10\,\mathrm{L}$, $V_2 = 40\,\mathrm{L}$, $\mathrm{Cl}_1 = 0.5\,\mathrm{L/min}$ and $\mathrm{Cl}_2 = 0.4\,\mathrm{L/min}$. A pump gives the bolus that brings the plasma to 3 mg/L and then holds it there exactly.

(a) Find the infusion rate just after the bolus and 100 min later. *Checkpoint:* 2.70 mg/min and 1.94 mg/min.

(b) Find the rate the pump approaches after many hours, and say where the difference goes in the meantime. *Checkpoint:* 1.50 mg/min, the rest fills the peripheral compartment.

1.3 Concept check · 2 min

A target-controlled pump inverts a population model and holds the model’s plasma at 2 mg/L exactly. The patient’s volumes match the model, but the patient clears the drug 1.6 times as fast as the population. After four hours, what is true of the patient’s effect-site concentration?

(a) It equals 2 mg/L, because the pump’s loop is exact.

(b) It lies above 2 mg/L, because the pump gives extra drug.

(c) It lies below 2 mg/L and heads for about 1.25 mg/L.

(d) It oscillates about 2 mg/L at the effect-site rate.

1.4 Theory and calculation · 4 min

A drug’s window is 2 to 4 mg/L. A population model is used open loop, and patients’ clearances range from 0.8 to 1.25 times the population’s.

(a) A pump holds the model at 2.5 mg/L. Find the steady concentrations of the two extreme patients. *Checkpoint:* 3.125 mg/L and 2.0 mg/L.

(b) Find the range of model targets that keeps every patient in the window, and say whether open-loop dosing is possible. *Checkpoint:* 2.5 to 3.2 mg/L, yes.

1.5 Theory and calculation · 5 min

A culture doubles every 20 h and takes glucose in proportion to its growth. Over its first 20 h it takes 0.4 g/L. It is sampled once a day and fed a daily bolus, and its glucose must stay inside a band 5 g/L wide.

(a) By what factor does the culture grow in one day, and how much glucose does it take on day 0 and day 1 (the first and second days)? *Checkpoint:* 2.30, 0.52 g/L and 1.19 g/L.

(b) On which day does the daily uptake first exceed the band, so that no daily bolus can hold it? *Checkpoint:* day 3, with 6.29 g/L.

1.6 Theory and calculation · 5 min

Methanogens grow by a Haldane law with $\mu_{max} = 0.5\,\mathrm{d}^{-1}$, $K_S = 4\,\mathrm{mmol/L}$, $K_I = 100\,\mathrm{mmol/L}$ and a death rate of $0.05\,\mathrm{d}^{-1}$, in a stirred tank that keeps no biomass back.

(a) Find the acid level of the peak, the peak growth rate and the critical dilution rate. *Checkpoint:* 20 mmol/L, $0.357\,\mathrm{d}^{-1}$, $0.307\,\mathrm{d}^{-1}$.

(b) At a dilution rate of $0.2\,\mathrm{d}^{-1}$, find the two steady acid levels and say which one the tank should run at. *Checkpoint:* 4.17 mmol/L and 95.8 mmol/L, the lower.

1.7 Theory and calculation · 4 min

A cell count is drawn at 08:00 and reported at 14:00. The culture grows at $0.03\,\mathrm{h}^{-1}$.

(a) An estimator resets its cells to the count when the report arrives, as if the count had been drawn then. How far low is its estimate? *Checkpoint:* 16.5 %.

(b) An estimator instead updates a bias with weight $\lambda = 0.4$, compared at the hour of the draw. Against a constant model error, what fraction of the error has it removed after three counts? *Checkpoint:* 0.784.

1.8 Concept check · 2 min

A loop holds a weighed running total of fluid removed on its target. One scale begins to drift slowly. The loop's own error stays small. What reveals that the patient's true removal is falling behind?

(a) A second, independent measurement of the same flows, such as the pumps' own counts.

(b) A faster integral action on the balance loop.

(c) A derivative term on the weighed running total.

(d) A larger proportional gain so that the error stays smaller still.

1.9 Theory and calculation · 5 min

Two actuators push one way.

(a) A drug in one compartment with a time constant of 30 min stands at 5 mg/L, above a limit of 4 mg/L. The pump stops. How long until the concentration is back at the limit, and how long until it reaches 2 mg/L? *Checkpoint:* 6.7 min and 27.5 min.

(b) A rider with $W_0' = 20\,\mathrm{kJ}$, $P_{pk} = 1000\,\mathrm{W}$, $CP = 250\,\mathrm{W}$ and $E_c = 0.75$ holds an exertion of 0.9 from a full reserve. Find the fade time constant, the power at the start and the power a minute later. *Checkpoint:* 44.4 s, 700 W, 367 W.

1.10 Concept check · 2 min

Run A spends an area of 40 mg/dL h above its glucose band and none below it. Run B spends 5 mg/dL h below the band and none above. An objective weights the area below the band by w and the area above by one. For which weights does the objective rank A as the better run?

(a) For every w, because A never goes low.

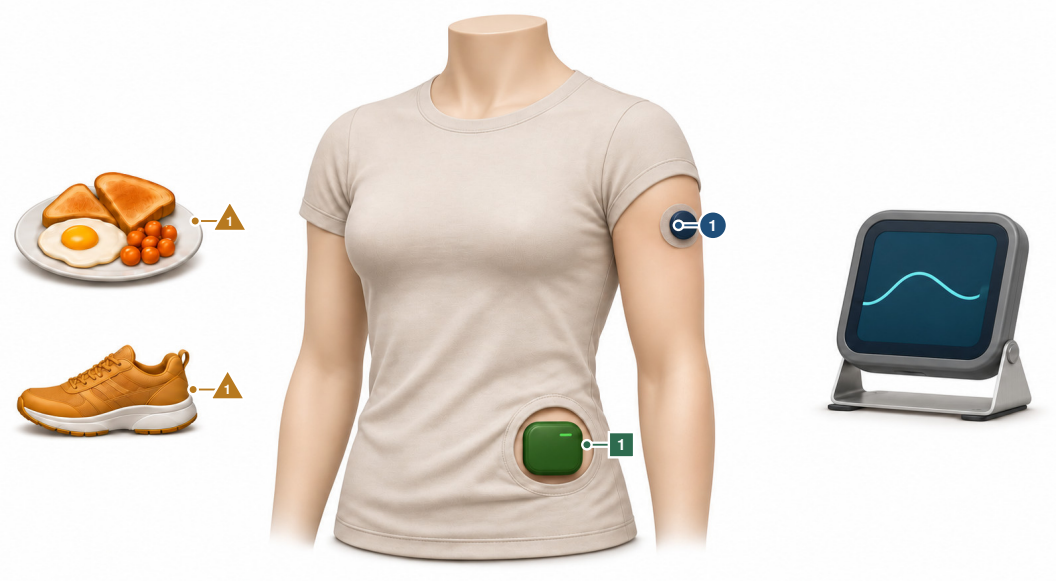
(b) For w below 8 only.

(c) For no w, because B has the smaller total area.

(d) For w above 8.

CHAPTER 2

Artificial Pancreas Simulator for Direct Action and One-Sided Insulin Delivery



The controller moves 1 insulin pump worn on the abdomen

The controller reads 1 continuous glucose monitor on the arm

What disturbs it 1 a meal, or exercise

Notice that the monitor on the arm and the pump on the abdomen are separate devices, and that the pump can add insulin or stop but can never take it back. Illustration, not to scale.

IN PLAIN TERMS **Insulin That Can Only Push One Way**

The main idea. People with type 1 diabetes no longer make their own insulin, so a small pump worn on the body adds it. An automatic system reads the blood sugar from a sensor and sets the speed of the pump.

The challenge. Insulin only lowers blood sugar, and the pump cannot take back what it has given, so stopping it does not stop a fall for more than an hour. A low sugar is also more dangerous than a high one of the same size.

Why it matters. A score that treats highs and lows alike can pick a setting that is worse on the side that matters most.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	The first page, the history boxes of Section 2.2, the fact boxes of Sections 2.6 and 2.7, the try-it boxes of Section 2.8, and the key points
Study	65 min	The whole chapter with its pause points and the twelve exercises
Practice	15 min	The first page, Section 2.7, Section 2.10, Section 2.11 and the open challenge

The control question An insulin pump can only lower glucose, it stops at zero, and a low glucose costs more than a high one of the same size. How should a direct-acting PI loop be tuned when its actuator pushes one way and its error is not symmetric, and what does a predictive controller gain when it is told of a meal?

Learning objectives

- Derive the fasting steady-state map of the minimal model and the pump rate that holds a chosen glucose.
- Identify a first-order plus dead-time model from a pump step and explain why its dead time is the plasma insulin lag.
- Explain why the loop is direct acting and why no retuning shortens a low once the pump is at zero.
- Compare tunings on one-sided excursion areas, and show where the symmetric integrated error misranks them.
- Separate what a meal announcement, a soft glucose band and a bounded meal estimate each change in a predictive controller.

Prerequisites Sections 1.3 (loop action), 1.4.2 (compartments), 1.4.6 (one-sided actuators and asymmetric cost), 1.6.2 (velocity form) and 1.6.4 (scenarios and data). Section 1.6.3 is needed only for the predictive controller and the estimators.

The app https://apmonitor.com/apps/diabetic/

See also Volume 2, Chapter 2 for velocity-form PID with back-calculation and a relay test placed by its switching band, and Volume 2, Chapter 3 for anti-windup at a bound the supply sets. Volume 9, Chapter 9 for control against a band rather than a setpoint, and Volume 9, Chapter 5 for a predictive controller weighed against releases ahead of published demand.

2.1 Introduction

Most regulatory loops taught in a first course have an actuator that works in both directions. A heater can be turned up or down, a valve can be opened or closed, and the error signal that drives the actuator is scored with a metric that treats an excursion above setpoint and an excursion below setpoint as equally bad. Subcutaneous insulin delivery satisfies neither condition. An infusion pump adds insulin, and insulin removes glucose, so the process gain from the manipulated variable to the controlled variable is negative and the controller must be direct acting in the ISA sense, with an output that rises as the measurement rises [1]. In the convention of Section 1.3 the loop has the action $a = -1$. The pump has no negative setting, because there is no way to withdraw insulin that has already reached the tissue, so the manipulated variable is bounded below at zero and the loop has authority in one direction only. The cost of the two error signs is also different. A sustained high glucose is a long-term risk, while a sufficiently deep low is an immediate one.

The dynamics compound the actuator constraint. A meal delivers glucose to the blood within an hour, while the insulin path from the pump runs through a plasma compartment and then through a remote action compartment whose time constant in this model is $1/p_2 = 81.3\,\text{min}$. The disturbance is therefore faster than the actuator, and the same lag that delays the effect of switching the pump on also delays the effect of switching it off. Once glucose is falling, shutting the pump down does not stop the fall, because the insulin already in the action compartment continues to clear glucose for more than an hour.

Teaching simulators for glucose control frequently present the plant as a first-order lag with a fixed gain and an unbounded actuator. However, a symmetric linear plant with a two-sided actuator removes exactly the three properties that make this loop worth studying, because it hides the sign reversal, it hides the shutdown floor, and it makes a symmetric error integral a legitimate score. This chapter studies a simulation of Type 1 diabetic glucose regulation that retains the sign reversal, the zero floor on the pump and the asymmetric cost of error, and it quantifies what each of them does to the tuning of a proportional-integral controller. Neither the model nor the control algorithms are new.

Section 2.2 traces automated insulin delivery from bedside instruments to hybrid closed loop. Section 2.3 sets out the six-state model, its disturbances, instrument, actuator and timing, with

every relation and constant a reimplementation needs. Section 2.4 derives the closed-form fasting steady-state map of the model, shows that its slope falls by a factor of 1.75 between infusion rates of 3.0 and 4.0 µU/min, and shows that the apparent dead time of the identified first-order plus dead-time model is the plasma insulin clearance time constant $1/k_e$. Section 2.5 describes the direct-acting PID, a linear model predictive controller that can replace it, two moving horizon estimators and a relay autotune. Section 2.6 defines the one-sided scores, and Section 2.7 measures the loop over a full simulated day. In a sweep of integral gain over that day a fivefold change in integral gain moves the integrated absolute error by 8.7 % and the glucose nadir by 6.4 mg/dL, which quantifies the claim that a symmetric metric is not a suitable objective for this loop. Section 2.8 compares the predictive controller with the PID to separate what a meal announcement and a soft glucose band each change, and Section 2.9 compares the estimators and the constants a relay test writes. Section 2.10 describes the sensor, pump and controller platform of real systems, and Sections 2.11 to 2.13 state the limitations of the model and what the results mean beyond this plant.

The chapter rests on standard process control fundamentals. The model is a set of compartment mass balances with first-order exchange between pools, and the loop behavior follows from the time-scale separation between a fast glucose balance and two slower insulin lags [1, 2]. The principles it applies are controller action against a negative process gain, actuator saturation and anti-windup, identification of a first-order plus dead-time model from a step test, steady-state offset and its removal by bias or integral action, and performance evaluation under sampled and noisy measurements [3]. Loops with one-sided actuators appear industrially wherever the actuator works only in the safe direction, such as pressure relief and emergency cooling, so the tuning question posed here is not specific to glucose. Section 1.4.6 sets out the general case, and Chapter 3 meets the same one-way pump with an anesthetic.

2.2 Background

Automated insulin delivery began as bedside instrumentation rather than as a wearable device [4, 5], as the first history box below tells.

> ⧖ HISTORY **Bedside Machines Close the Loop (1964 to 1979)**
>
> Automatic control of blood glucose began at the hospital bedside. After pioneering work in 1964, five teams reported closed-loop results between 1974 and 1978. Their machines measured glucose in blood from a vein and infused insulin, and glucose as well, into the vein. Their control laws were of the proportional-derivative kind, which computed the insulin dose from the glucose value and its rate of change. One of these systems became the first commercial device in 1977. The intravenous route could not leave the hospital, but the machines proved that external control of glucose was feasible. In 1979 two studies showed that insulin could be delivered continuously under the skin, the route a wearable pump uses today [6].

The move from bedside to wearable removed the glucose infusion channel and replaced intravenous measurement with a subcutaneous sensor, which added transport lag on both sides of the loop. Feasibility of a fully automatic subcutaneous system was established with a proportional integral derivative algorithm operating on a continuous glucose monitor, and the reported result was glucose inside 70 mg/dL to 180 mg/dL for 75 % of the time under closed loop against 63 % open loop [7]. Model predictive formulations followed, first as simulation studies [8] and then in clinical protocols [9]. A review of the algorithms available at that stage set out why the problem resists a textbook solution, naming the sensor lag, the actuator lag, the unmeasured meal and the impossibility of negative insulin as the four difficulties any candidate must address [10].

Safety layers rather than better regulators account for much of the progress since. Constraining the controller by the insulin already delivered, an approach usually called insulin on board, prevents the algorithm from stacking correction doses whose effect has not yet appeared, and it does so as an explicit constraint inside the optimization rather than as a detuning [11]. Testing moved to virtual

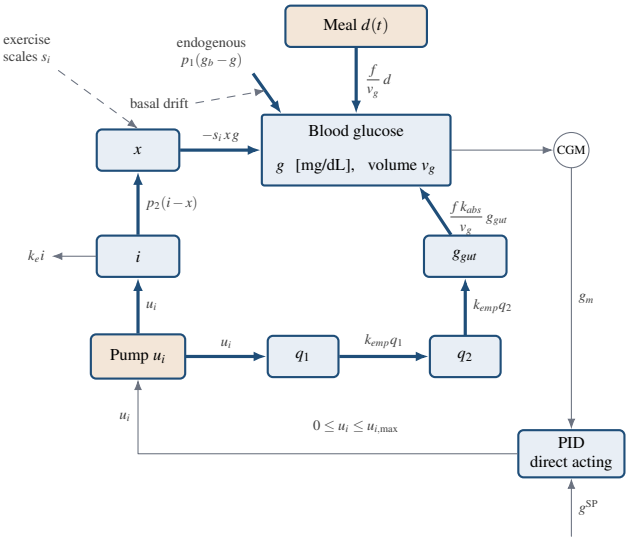


Figure 2.1 Compartments and flows of the expanded Bergman minimal model as implemented. The infusion rate u_i enters both the plasma insulin pool and the gastric transit chain, and the meal enters the blood directly through $f d / v_g$, which is how the published assignment equations are written.

patients, as the second history box tells. Systems that automate basal delivery while leaving the meal bolus to the user, usually called hybrid closed loop, reached clinical practice on that foundation [12], and a six-month randomized multicenter trial reported time in the 70 mg/dL to 180 mg/dL range of 71 % under closed-loop control against 59 % in the control group [13].

> ⧖ HISTORY **Virtual Patients in Place of Animal Trials (2009)**
>
> A report of 2009 described a simulator for testing closed-loop algorithms before any clinical study. It held 300 virtual subjects, 100 adults, 100 adolescents and 100 children, built from the data of real people to span the variability seen across the population with type 1 diabetes, together with models of sensor error and of insulin delivery by pump [14]. The United States Food and Drug Administration accepted it as a substitute for animal trials in the preclinical testing of closed-loop control strategies. Because an experiment in the computer costs a fraction of the time and money of an animal trial, it became a primary test bed for new control algorithms [6]. The simulator of this chapter holds one virtual subject, as Section 2.3 states.

2.3 Process Model

The simulator implements the expanded Bergman minimal model exactly as it is written in the APMonitor diabetic blood glucose assignment. The original minimal model was built to estimate insulin sensitivity from an intravenous glucose tolerance test using three states, glucose, remote insulin action and plasma insulin [15]. The expansion adds a three-compartment chain that delays material on its way into the blood, which is the same structural device used by larger meal models to reproduce the shape of a postprandial glucose curve [16]. Figure 2.1 sets out the compartments, the flows between them, the pump and meal inputs, and the point of measurement.

2.3.1 Roles of the Signals

The controlled variable is blood glucose as reported by the continuous glucose monitor, the manipulated variable is the pump infusion rate, and the disturbances are the meal appearance rate, the exercise term and the basal drift. The complete classification is given in Table 2.1. The distinction

Table 2.1 Classification of Signals in the Glucose Loop

Symbol	Role	Unit	Range
g	state, CV source	mg/dL	≥ 0
g_m	measurement (CV)	mg/dL	≥ 0
x	state, remote action	µU/mL	≥ 0
i	state, plasma insulin	µU/mL	≥ 0
q_1	state, transit pool 1	model units	≥ 0
q_2	state, transit pool 2	model units	≥ 0
g_{gut}	state, gut glucose	model units	≥ 0
u_i	MV, infusion rate	µU/min	0 to $u_{i,\max}$
d	DV, glucose appearance	model units	≥ 0
e_x	DV, exercise	%	0 to 100
b_d	DV, basal drift	%	-30 to $+30$
g^{SP}	setpoint	mg/dL	40 to 250

between the true state g and the reported measurement g_m matters throughout, because the controller only ever sees g_m while every performance number quoted in Section 2.7 is computed from g.

2.3.2 Governing Equations

The six state equations are written with every parameter in reciprocal minutes and every right-hand side multiplied by 60, so the independent variable of the simulation is time in hours while the parameter values stay in the units in which they are published. The set is given in Equation (2.1).

$$\frac{\mathrm{d}g}{\mathrm{d}t} = 60\left[-p_1(g-g_b) - s_i x g + \frac{f k_{abs}}{v_g} g_{gut} + \frac{f}{v_g} d\right] \tag{2.1a}$$

$$\frac{\mathrm{d}x}{\mathrm{d}t} = 60\, p_2\,(i-x) \tag{2.1b}$$

$$\frac{\mathrm{d}i}{\mathrm{d}t} = 60\,(-k_e i + u_i) \tag{2.1c}$$

$$\frac{\mathrm{d}q_1}{\mathrm{d}t} = 60\,(u_i - k_{emp} q_1) \tag{2.1d}$$

$$\frac{\mathrm{d}q_2}{\mathrm{d}t} = -60\,k_{emp}(q_2 - q_1) \tag{2.1e}$$

$$\frac{\mathrm{d}g_{gut}}{\mathrm{d}t} = 60\,(k_{emp} q_2 - k_{abs} g_{gut}) \tag{2.1f}$$

where g is blood glucose, x is remote insulin action, i is plasma insulin, q_1 and q_2 are the two transit pools, g_{gut} is the gut glucose pool, u_i is the infusion rate, d is the glucose appearance rate, g_b is the basal glucose parameter, p_1 is the glucose clearance rate, p_2 is the insulin action rate, s_i is insulin sensitivity, k_e is plasma insulin clearance, k_{abs} is gut absorption, k_{emp} is gastric emptying, f is the fraction of appearing glucose that reaches the blood and v_g is the glucose distribution volume. Every state is clamped at zero after each integration step, because g, i and the three pools are concentrations and masses that cannot go negative.

Two features of Equation (2.1) are consequences of following the published assignment rather than of a modeling choice made here. The transit chain (2.1d) to (2.1f) is driven by the infusion rate u_i and not by the ingested carbohydrate, so q_1, q_2 and g_{gut} track the pump. The meal reaches the blood only through the direct term $f d / v_g$ in (2.1a). A reimplementation that routes the meal into q_1 instead is a different model and will not reproduce the numbers reported here.

The nonlinearity of the plant lives in the single product $s_i x g$ in (2.1a). Glucose clearance by insulin is proportional to both the insulin action and the glucose concentration, so the incremental effect of

insulin falls as glucose falls, and a loop that recovers from a postprandial peak can still overshoot below the setpoint during the descent.

2.3.3 Meals, Exercise and Basal Drift

The glucose appearance rate is the sum of a fasting baseline and the contribution of every meal that has started, as shown in Equation (2.2).

$$d(t) = d_0 \left(1 + \frac{b_d}{100}\right) + \sum_k m(t - t_k, A) \tag{2.2a}$$

$$m(\tau, A) = \begin{cases} 0, & \tau < 0 \\ A\,\tau, & 0 \le \tau < 1\,\mathrm{h} \\ A\,\mathrm{e}^{-(\tau-1)}, & \tau \ge 1\,\mathrm{h} \end{cases} \tag{2.2b}$$

where $d_0 = 1000$ is the fasting baseline in model units, b_d is the basal drift in percent, t_k is the start time of meal k, and A is the meal amplitude. Each meal ramps linearly to its full amplitude over one hour and then decays with a one-hour time constant, so the appearance rate peaks exactly one hour after the meal begins. The scheduled plan places meals at 08:00, 12:00 and 18:00 on the simulated clock. The schedule is off in every case as the application loads it and is switched on as a disturbance setting, and an additional unscheduled meal can be injected at any simulated time with the current amplitude. Three of the shipped cases script one such meal of amplitude 1800 at $t = 2\,\mathrm{h}$.

The baseline term d_0 is not part of the published assignment code. It is required here because the published initial condition is a steady state only when the appearance rate carries that value. With $d = 1000$ and $u_i = 3.0\,\mu\mathrm{U/min}$ the steady state of Equation (2.1) lies within 0.005 of every published value, and the reference case starts at that steady state, where no state moves, which is what makes a fasting step test meaningful.

Exercise scales insulin sensitivity by $s_i(1 + e_x/100)$, so the insulin already present clears glucose faster without any change in the pump rate. Basal drift scales only the baseline term of d and stands in for a change in endogenous glucose production. Neither disturbance is available to the controller, which sees only g_m.

2.3.4 Instrument and Actuator

The continuous glucose monitor adds a zero-mean Gaussian deviate of standard deviation σ to the true glucose and clamps the result at zero, as shown in Equation (2.3).

$$g_m = \max(0,\ g + \sigma\,\mathcal{N}(0,1)) \tag{2.3}$$

The deviate is drawn once per control update by the Box-Muller transform from a generator that is reseeded with a fixed seed whenever a case starts, so a run with noise repeats exactly. There is no sensor lag and no calibration error in the instrument model, so the subcutaneous transport delay of a real monitor is absent. The pump applies the commanded rate without lag, without a minimum delivery increment and without a delivery quantum, and it is clamped to $0 \le u_i \le u_{i,\max}$ with $u_{i,\max} = 10\,\mu\mathrm{U/min}$ by default. The manual pump setting follows the same maximum.

2.3.5 Parameters, Limits and Initial Conditions

The constants of Equation (2.1) and the settings that the application exposes are listed in Table 2.2. Five model parameters, the pump maximum and the sensor noise are adjustable inputs, and the remainder are fixed. The insulin distribution volume $v_i = 12\,\mathrm{L}$ appears in the published parameter list but in no equation of the implemented model, so it has no effect on any trajectory.

Equation (2.1b) equates x with i, so the remote action state carries the units of plasma insulin and s_i carries mL per µU per minute in order that $s_i x g$ have the units of g divided by time. The three

Table 2.2 Model Parameters, Defaults and Adjustable Ranges

Symbol	Meaning	Value	Range in settings
g_b	basal glucose parameter	291 mg/dL	fixed
p_1	glucose clearance rate	0.0317 min^{-1}	fixed
p_2	insulin action rate	0.0123 min^{-1}	fixed
f	fraction reaching blood	0.8	fixed
v_i	insulin volume (unused)	12 L	fixed
s_i	insulin sensitivity	0.029	0.01 to 0.06
k_e	plasma insulin clearance	0.09 min^{-1}	0.03 to 0.2
k_{abs}	gut absorption rate	0.012 min^{-1}	0.004 to 0.03
k_{emp}	gastric emptying rate	0.18 min^{-1}	0.05 to 0.35
v_g	glucose volume	12 L	5 to 25
$u_{i,\max}$	pump upper limit	10 µU/min	5 to 20
σ	CGM noise standard dev.	1.5 mg/dL	0 to 5
d_0	fasting appearance baseline	1000	fixed
A	meal amplitude	1200	0 to 2400
e_x	exercise	0 %	0 to 100
b_d	basal drift	0 %	−30 to +30
g^{SP}	glucose setpoint	80 mg/dL	40 to 250
K_c	proportional gain	0.3	0 to 1.5
τ_I	integral time	3 h	0.002 to 48[a]
τ_D	derivative time	0 h	0 to 16.8
T_p	MPC prediction horizon	6 h	1 to 12
λ	MPC move suppression	3	0 to 30
u_{bias}	controller bias	u_i at start	not set directly

[a] The top of the range, 48 h, switches integral action off.

transit states carry the infusion rate integrated over time, and the coefficient $f k_{abs}/v_g$ that returns g_{gut} to the blood is applied as a numerical factor rather than as a dimensionally consistent conversion, which is how the published assignment writes it.

The published initial condition is $[g, x, i, q_1, q_2, g_{gut}] = [76.22, 33.33, 33.33, 16.67, 16.67, 250.0]$ with the pump at the basal rate 3.0 µU/min. Setting the derivatives of Equation (2.1) to zero at that infusion rate gives $i = x = u_i/k_e = 33.33$, $q_1 = q_2 = u_i/k_{emp} = 16.67$, $g_{gut} = u_i/k_{abs} = 250$ and $g = 76.216\,\text{mg/dL}$ from Equation (2.4), so the published vector is the steady state to the digits it is written with. The application starts every case at a steady state computed this way. In manual mode it is the steady state of the pump rate in force, and in automatic mode it is the steady state of the setpoint in force, with the pump at the rate of Equation (2.5) that holds it. The parameters and disturbances in force enter the calculation, so a setting chosen before a run starts changes the starting state and is not a step. The reference case therefore starts at $[76.216, 33.333, 33.333, 16.667, 16.667, 250.0]$, and every automatic case starts at 80 mg/dL with the pump at 2.853 µU/min.

2.3.6 Integration and Timing

The engine of Section 1.6.1 advances the state by explicit Euler on a fixed step of $1/500$ hours, or 7.2 sec. The controller and the continuous glucose monitor execute on every 84th step, so the control period is $\Delta t = 0.168\,\text{h}$ (10.08 min), which is the period the application declares. A nominal ten-minute period would not be a whole number of steps. When a control period has elapsed, the measurement is read and the controller executes first, the state is then advanced by one Euler step, and each state is clamped at zero. A change the operator makes to a disturbance or a model parameter reaches the plant at the next controller execution, while an unscheduled meal enters at once. The speed setting spans 0.5 to 10 simulated hours per real second and leaves the trajectory unchanged. At the default $1\times$ a twenty-four hour day takes 24 sec.

The run is logged once per control period, and the exported time column is the model clock that the meal schedule uses, on which the 08:00 meal begins at $t = 8\,\text{h}$. A row stamped t holds the state at t, the glucose reading on which the controller acts at t and the infusion rate held over the control period that ends at t. Every time quoted in this chapter is model time.

2.3.7 Assumptions

The model is a lumped compartment description with a small number of pools, and each of the following simplifications has a regime in which it fails.

One well-mixed glucose pool. Blood glucose is a single concentration in a fixed volume v_g, so there is no distinction between plasma and interstitial glucose. A real continuous glucose monitor reads interstitial fluid and lags plasma by five to fifteen minutes, and that lag is absent here.

Insulin action through one remote compartment. The path from plasma insulin to glucose clearance is a single first-order lag with rate p_2. Subcutaneous depot kinetics, which dominate the delay in a real wearable pump, are not represented at all, so the simulated actuator is faster than a real one.

Linear endogenous balance. The term $-p_1(g - g_b)$ is linear and restores glucose toward $g_b = 291\,\text{mg/dL}$. That value is a fitting parameter of the minimal model and not a physiological setpoint. The model carries no counter-regulatory response, so there is no glucagon action, no hepatic glucose release and no floor under a low excursion. With the pump at zero the fasting steady state is 2394 mg/dL, which is not a physiological number and shows that the model is calibrated for a band around the basal point and not across its full mathematical range.

Stylized disturbances. The amplitude A is in the units of d and has no gram equivalent, and the ramp and decay shape is fixed, so a meal of the same carbohydrate content but a different fat content cannot be expressed. A single scaling of s_i captures the direction of the exercise effect and not its mechanism, and it applies instantly rather than building and decaying around the activity.

Fluxes that are not physiologically scaled. At the basal operating point the four terms of (2.1a) evaluate to $+6.81$, -73.67, $+0.20$ and $+66.67$ in mg/dL/min, and they sum to within 0.02 mg/dL/min of zero. Individual terms of that size are one to two orders of magnitude above physiological glucose turnover, so the parameter set reproduces the net dynamics around its own operating point rather than a flux balance whose individual terms can be read as rates of appearance and disappearance.

One patient. The parameters describe a single virtual subject. Between subject variability and within subject variability across a day are what in silico protocols exist to test [14], and neither is represented by a single parameter set.

2.4 Steady-State and Dynamic Analysis

2.4.1 Steady-State Map

Holding the infusion rate constant and setting the derivatives of Equation (2.1) to zero gives $i = x = u_i/k_e$, $q_1 = q_2 = u_i/k_{emp}$ and $g_{gut} = u_i/k_{abs}$, and substituting those into (2.1a) collapses the balance to a single algebraic relation, shown in Equation (2.4).

$$g_{ss}(u_i) = \frac{p_1 g_b + \dfrac{f}{v_g} u_i + \dfrac{f}{v_g} d}{p_1 + \dfrac{s_i}{k_e} u_i} \tag{2.4}$$

Equation (2.4) is a hyperbola in u_i, not a line, and its slope is $(BC - AD)/(C + Du_i)^2$ with $A = p_1 g_b + fd/v_g$, $B = f/v_g$, $C = p_1$ and $D = s_i/k_e$. At the fasting appearance rate $d = 1000$ the numerator of that slope is -24.45, so the tangent gain is $-24.5\,\text{mg/dL}$ per $\mu\text{U/min}$ at $u_i = 3.0\,\mu\text{U/min}$ and $-14.0\,\text{mg/dL}$ per $\mu\text{U/min}$ at $u_i = 4.0\,\mu\text{U/min}$, a factor of 1.75 across one unit of infusion rate. The map and the effect of a 30 % basal drift on it are shown in Figure 2.2.

Inverting Equation (2.4) for the infusion rate that holds a chosen glucose gives Equation (2.5), which is the bias a positional controller would need to hold that glucose without integral action.

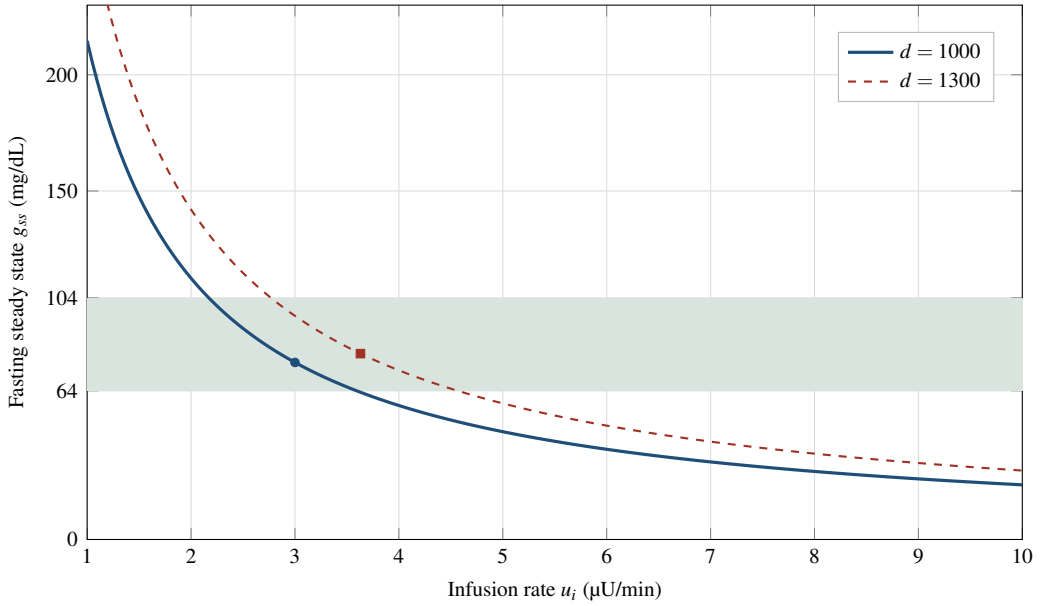


Figure 2.2 Fasting steady-state glucose against infusion rate from Equation (2.4). The shaded strip is the 64 mg/dL to 104 mg/dL teaching band of Section 2.6, the circle is the basal operating point and the square is the infusion rate that holds 80 mg/dL after a 30 % basal drift.

$$u_{bias}(g) = \frac{p_1(g_b - g) + \dfrac{fd}{v_g}}{\dfrac{s_i}{k_e}g - \dfrac{f}{v_g}} \tag{2.5}$$

At $g = 76.2\,\mathrm{mg/dL}$ and $d = 1000$ this returns 3.00 µU/min, at the default setpoint of 80 mg/dL it returns 2.85 µU/min, and after a 30 % basal drift it returns 3.63 µU/min. The application starts the patient at the pump rate of Equation (2.5) for the setpoint and disturbances in force when a run starts, and the controller takes that rate as its bias, 3.00 µU/min at 76.22 mg/dL and 2.85 µU/min at 80 mg/dL. A drift chosen before the run starts enters that calculation and returns 3.63 µU/min, but a drift that arrives during a run is unmeasured, so its value cannot be computed in advance and must be recovered by feedback.

2.4.2 Time Constants of the Linearized Model

Linearizing Equation (2.1) about the basal point separates the insulin path into three first-order lags in series. The remote action follows plasma insulin as an effect site follows plasma in Section 1.4.2. The plasma pool (2.1c) contributes $1/k_e = 11.1\,\mathrm{min}$, the remote action pool (2.1b) contributes $1/p_2 = 81.3\,\mathrm{min}$, and the glucose balance itself contributes $1/(p_1 + s_i\bar{x}) = 1.0\,\mathrm{min}$ at $\bar{x} = 33.33$. The dominant time constant is therefore an insulin compartment and not glucose, and the glucose pool is fast enough to be invisible in the response. An identification that fits a first-order plus dead-time form to this cascade should return a time constant close to 81 min and an apparent dead time close to 11 min, and Section 2.4.3 confirms both.

2.4.3 Step-Test Identification

A fasting step test runs from the reference operating point, which the application loads in manual mode with the scheduled meals off and the sensor noise at zero. The pump is held at 3.0 µU/min

with glucose flat at 76.22 mg/dL, and the operator moves it to 4.0 µU/min at $t = 2$ h, which the plant receives at the controller execution at 2.016 h. Glucose falls from 76.22 mg/dL to 57.67 mg/dL over the following twelve hours, which is a secant process gain of $K_p = -18.55$ mg/dL per µU/min. The analytic value from Equation (2.4) is $57.669 - 76.215 = -18.546$, so the integration reproduces the steady-state map to four significant figures. The case starts at the exact steady state, so the unstepped run does not move and no drift correction is needed. The application's own open-loop step case starts at 80 mg/dL with the pump at 2.85 µU/min and moves it by 2.0 µU/min at 2 h, and glucose settles at 47.8 mg/dL. Its secant gain of −16.1 mg/dL per µU/min is 60 % of the tangent gain of −27.0 mg/dL per µU/min at its starting rate, which is the curvature of Figure 2.2 measured directly.

The response reaches 63.2 % of its change at $t_{63} = 1.34$ h after the step. A two-point fit through the 28.3 % and 63.2 % crossings, which occur at 0.553 h and 1.340 h, returns $\tau_p = 1.18$ h and $\theta_p = 0.16$ h [1]. The fitted dead time of 0.16 h is 9.6 min against a plasma clearance time constant $1/k_e$ of 11.1 min, and the fitted time constant of 1.18 h is 70.8 min against a remote action time constant $1/p_2$ of 81.3 min. The apparent dead time of the identified model is not a transport delay. It is the second lag of the insulin chain reported as a delay by the first-order fit.

Internal model control correlations for a plant of this form give $K_c = \tau_p/(|K_p|(\tau_c + \theta_p))$ with $\tau_I = \tau_p$ [17]. A conservative closed-loop time constant $\tau_c = \tau_p$ returns $K_c = 0.048$, and an aggressive $\tau_c = \theta_p$ returns $K_c = 0.20$. The shipped default of $K_c = 0.3$ is more aggressive than either, and the shipped integral time $\tau_I = 3.0$ h, which is an integral gain $K_I = K_c/\tau_I = 0.1$ in µU/min per mg/dL per hour, is 2.5 times the identified time constant rather than equal to it. Section 2.7 shows why the integral time is deliberately slow.

Four models appear in the analysis and are kept distinct. The six-state set of Equation (2.1) is the simulation model. The first-order plus dead-time fit is the identified control model and is used only to produce tuning constants. The PID itself carries no internal model. Its starting bias is the pump rate in force when the run starts, read from the pump, which on a steady start is the value of Equation (2.5) at the setpoint in force. The predictive controller of Section 2.5.4 plans on the fourth, a linear model fitted separately at 80 mg/dL.

2.4.4 Actuator Authority

The manipulated variable is bounded on both sides, but the two bounds are not alike. The upper bound $u_{i,\max}$ is a device setting, and raising it provides faster rejection of a meal. The lower bound is zero and is a property of the physics, because a pump can add insulin and cannot remove it. Stopping delivery does not remove the insulin already in the plasma and action compartments, and that stored quantity decays with the time constants of (2.1c) and (2.1b). The slowest of those is $1/p_2 = 81.3$ min, the time constant that sets the slow way down of Equation (1.9), and it is the quantitative statement of why a low excursion is not recoverable by the actuator and why deployed systems suspend delivery on a prediction rather than on a measurement [10]. Constraining the controller by the insulin already on board addresses the same limit from the opposite side [11].

2.5 Controller Design

2.5.1 Velocity Form and Direct Action

The pump is driven by the velocity-form PID of Section 1.6.2, with the proportional action on the error and the derivative taken on the measurement. It executes every $\Delta t = 0.168$ h, and each execution computes an increment rather than the pump rate itself. The algorithm as coded for this loop is given by Equations (2.6) to (2.10).

$$e_k = a\left(g^{\mathrm{SP}} - g_{m,k}\right), \qquad a = -1 \tag{2.6}$$

$$\Delta u_k^P = K_c\left(e_k - e_{k-1}\right) \tag{2.7}$$

$$\Delta u_k^I = \frac{K_c}{\tau_I}\, e_k\, \Delta t \tag{2.8}$$

$$\Delta u_k^D = -K_c \tau_D\, a\, \frac{g_{m,k} - 2g_{m,k-1} + g_{m,k-2}}{\Delta t} \tag{2.9}$$

$$\begin{aligned} s_k &= s_{k-1} + \Delta u_k^P + \Delta u_k^D \\ c_k &= u_{bias} + s_k + I_{k-1} \\ u_k &= \min(u_{i,\max}, \max(0, c_k)) \\ I_k &= I_{k-1} + \Delta u_k^I + \frac{\Delta t}{\tau_I}\left(u_k - c_k\right) \end{aligned} \tag{2.10}$$

where a is the controller direction, s_k accumulates the proportional and derivative increments without limit, c_k is the infusion rate the algorithm demands, u_k is that demand clipped to the pump range and applied as u_i until the next execution, and I is the integral state. The integral used at execution k was accumulated through execution $k-1$, and u_{bias} is the infusion rate when the run starts. In the notation of Equations (1.13) to (1.15), s_k is S_k, c_k is v_k, u_{bias} is u_b and the tracking time T_t equals τ_I. The glucose loop is constructed with $a = -1$, so a measurement above setpoint raises the pump, which is direct action in the ISA sense, and every tuning constant entered in the application is a positive magnitude. Section 1.6.2 explains why a setpoint change kicks the demand by K_c times the change and why a change of K_c acts at once on the whole error.

🕮 **Pause.** Try Exercise 2.1 before reading on.

2.5.2 Initialization and Mode Transfer

A velocity-form controller has no absolute output of its own, so the value at which the recursion starts is the bias of the equivalent positional law, which the application sets to the infusion rate at the start of the run. A case that opens in automatic mode starts at the steady state of its setpoint with $I = 0$, $s = 0$ and a zero previous proportional term, so its first execution contributes the whole proportional term, as shown in Equation (2.11).

$$u_0 = \min(u_{i,\max}, \max(0, u_{bias} + K_c\, e_0)) \tag{2.11}$$

Away from the pump limits the demand then equals the positional law $u_{bias} + K_c e_k$ plus the accumulated integral and derivative contributions. Manual tracking makes transfer bumpless in both directions, as Section 1.6.2 describes. A mode or setpoint chosen before the run starts is part of the starting state rather than a step, and the same choice carried through Reset gives the same run. The closed-loop runs of Section 2.7 choose automatic mode and the 80 mg/dL setpoint in the reference case before the run starts, so they start at 80 mg/dL with the pump and the bias at the 2.85 µU/min that Equation (2.5) gives for that setpoint. With no sensor noise the starting error is zero, and nothing moves until the first meal.

2.5.3 Anti-Windup at the Actuator Limits

Anti-windup is the back-calculation term of Equation (2.10), which corrects the integral alone, as in Section 1.6.2 [3]. Volume 2, Chapter 3 studies the same scheme at a bound set by the supply rather than by a valve. At the shipped $\tau_I = 3$ h each execution removes 5.6 % of the clipped excess, so a demand that has run well past a limit takes several executions to return into range. With the integral switched off at the top of the τ_I range, I keeps its value and no back-calculation is applied.

Against a one-sided actuator this behavior is not symmetric in its consequences. At the upper limit the clamp is protective, because the pump is already delivering as fast as the device allows and the excess demand is discarded rather than stored. At the lower limit the clamp marks the point at which the loop has no remaining authority. The controller output is zero, the error is still telling it to reduce insulin delivery further, and there is no further reduction available. Around the deepest point of each noise-free meal day in Section 2.7 the pump has already been cut below 0.9 µU/min, less than a third of the basal rate, and after the exercise step it sits at zero for 0.84 h while glucose recovers. Once the command is at zero, no retuning of K_c, τ_I or τ_D can shorten the excursion. Tuning can only change how much insulin is already stored when the fall begins, and the quantities that set how quickly the stored insulin clears are plant parameters.

2.5.4 Model Predictive Control

In automatic mode the pump can be driven instead by a linear model predictive controller (MPC), chosen with the Controller option, which Section 1.6.3 describes in general, while the PID stays in charge by default in all cases but cases 9 and 10 of Table 2.3. At every execution, so with a decision period $T_s = \Delta t = 0.168\,\mathrm{h}$, the MPC plans the infusion rate over a horizon, keeps only the first move and replans at the next execution, the receding-horizon practice of industrial predictive controllers [18]. It plans on Equation (2.12), in deviations from the design point $\bar{g} = 80\,\mathrm{mg/dL}$, $\bar{u} = 2.853\,\mu\mathrm{U/min}$.

$$\begin{aligned} \tau_1 \frac{\mathrm{d}\xi_1}{\mathrm{d}t} &= -\xi_1 + K_u\,(u_i - \bar{u}) \\ \tau_2 \frac{\mathrm{d}\xi_2}{\mathrm{d}t} &= -\xi_2 + \xi_1 + \delta \\ \tau_m \frac{\mathrm{d}\xi_3}{\mathrm{d}t} &= -\xi_3 + K_m\,\hat{d}(t), \qquad \hat{g} = \bar{g} + \xi_2 + \xi_3 \end{aligned} \tag{2.12}$$

Here ξ_1 is the slow insulin state, ξ_2 the glucose deviation from insulin and from δ, an unmeasured disturbance in mg/dL that enters at glucose, where a meal acts, ξ_3 the glucose from announced meals of appearance rate $\hat{d}$, and $\hat{g}$ the predicted glucose. Pump steps of $\pm 0.2\,\mu\mathrm{U/min}$ about $\bar{u}$, meals and noise off, gave $K_u = -27.16\,\mathrm{mg/dL}$ per µU/min, $\tau_1 = 1.358\,\mathrm{h}$ and $\tau_2 = 0.1876\,\mathrm{h}$, which lie 0.5, 0.2 and 1.3 % above the tangent gain of Section 2.4.3, $1/p_2$ and $1/k_e$, and the model misses their averaged response by 0.012 mg/dL root mean square, 0.21 % of its span. One meal of amplitude 1800 with the pump held gave $K_m = 0.0703$ mg/dL per model unit and $\tau_m = 0.0155\,\mathrm{h}$, with a residual of 0.24 mg/dL on a 124.7 mg/dL rise. Both paths are frozen at 80 mg/dL, whereas the plant's insulin acts harder at higher glucose, Equation (2.4) giving tangent gains of −43.1 and −60.9 mg/dL per µU/min at 101 and 120 mg/dL, and a meal raises glucose less as insulin action grows.

A steady-state Kalman filter of Equation (2.12), sampled with a zero-order hold at Δt and extended with δ as a random walk, the disturbance arrangement that leaves a predictive controller without steady-state offset [19], is updated at each execution under either controller, as Equation (2.13) gives.

$$\begin{aligned} \bar{\xi}_k &= \Phi\hat{\xi}_{k-1} + \gamma\,(u_{k-1} - \bar{u}) + \gamma_\delta\,\hat{\delta}_{k-1} + \gamma_m \hat{d}_{k-1/2} \\ \begin{bmatrix} \hat{\xi}_k \\ \hat{\delta}_k \end{bmatrix} &= \begin{bmatrix} \bar{\xi}_k \\ \hat{\delta}_{k-1} \end{bmatrix} + \ell\,\left(g_{m,k} - \bar{g} - \bar{\xi}_{2,k} - \bar{\xi}_{3,k}\right) \end{aligned} \tag{2.13}$$

Here Φ, γ, γ_δ and γ_m are the sampled model, u_{k-1} the rate over the period just ended, $\hat{d}_{k-1/2}$ the announced appearance at its middle and $g_{m,k}$ the reading the PID also receives. For a measurement variance $r = \sigma^2 = 2.25\,(\mathrm{mg/dL})^2$, a state noise variance $10^{-6}r$ and a disturbance noise variance $q_\delta = 100\,r$ the gain ℓ has entries 0.974 for ξ_2, 1.62 for δ and below 10^{-7} otherwise, so each innovation is booked almost wholly as glucose and as a lasting offset.

Each plan minimizes Equation (2.14) over $N_p = 36$ steps, 6.05 h for the 6 h horizon setting T_p, with eleven infusion rates μ_j held over blocks of $(1,1,1,1,2,2,3,4,6,9,6)$ steps, subject to $0 \le \mu_j \le u_{i,\max}$

and $70-\varepsilon_- \le \hat{g}_k \le 180+\varepsilon_+$ with $\varepsilon_\pm \ge 0$.

$$J = \frac{T_s}{\tau_n}\sum_{k=1}^{N_p}\left(\frac{\hat{g}_k - g^{\mathrm{SP}}}{y_s}\right)^2 + \lambda\sum_{j=1}^{11}\left(\frac{\mu_j-\mu_{j-1}}{u_s}\right)^2 + \rho_+\varepsilon_+^2 + \rho_-\varepsilon_-^2 \tag{2.14}$$

Here μ_0 is the rate applied now, $\tau_n = 1.55\,\mathrm{h}$, $y_s = 20\,\mathrm{mg/dL}$, $u_s = 10\,\mu\mathrm{U/min}$, $\lambda = 3$ a dimensionless move suppression, and ε_+ and ε_- one slack for each side of the band over the horizon, priced $\rho_+ = 100$ and $\rho_- = 1000$ per $(\mathrm{mg/dL})^2$. The infusion limits are hard, with $u_{i,\max}$ read as the PID reads it, and the soft band cannot leave the program without a solution. Against a tracking weight of 2.7×10^{-4} per $(\mathrm{mg/dL})^2$ and step, these prices make the band nearly a limit. The tenfold price of the low side is a teaching choice after Section 2.6 and the zone controllers with asymmetric costs proposed for the artificial pancreas [20], not a clinical calibration. The decisions are departures from $\bar{u}$, a dual active-set method solves the program, and λ and q_δ were fixed before any comparison by the tuning rule shared by the seven single-loop studios of the series.

With Meals announced on, $\hat{d}(t) = \sum_k m(t-t_k, A_k)$ from Equation (2.2) over the meals eaten and those the case has planned, so a planned meal is known from the start, and each plan reads $\hat{d}$ at the middle of every step. The announcement is exact, a best case that a person's estimate would not reach, and its glucose comes from the fitted meal path, never from the plant. Hybrid closed-loop systems ask the user for such an announcement, and in a pediatric study small manual premeal boluses added to closed-loop control gave average postprandial peaks of 194 against 226 mg/dL without them [21].

A plan without an optimum, or with a first rate that is not finite or lies outside the pump range, is set aside, and the PID, which follows the applied rate, takes the pump without a bump, moving it by 0.002 µU/min when a failure is forced at 1.5 h of case 9 with the noise off. The MPC reads the CGM, its own rate, the setpoint, the infusion limits and the announced meals, and no state of the plant, so Equations (2.1), (2.3) and (2.12) remain three distinct models.

2.5.5 Moving Horizon Estimation

The state that opens each plan can come instead from a moving horizon estimator (MHE), a least-squares fit over a sliding window of readings with hard bounds on the estimated states [22]. The filter of Equation (2.13) runs under either choice and still supplies the logged $\hat{\delta}$. Writing $\zeta = [\xi_1\ \xi_2\ \xi_3\ \delta]^{\mathsf{T}}$ for the filter's augmented state, execution k refits the readings $j = s,\dots,k$ with $s = k - N_e$ by solving Equation (2.15).

$$\begin{aligned}\min_{\zeta_s,\omega}\ & \|\zeta_s - \bar{\zeta}_s\|^2_{\bar{P}^{-1}} + \sum_{j=s}^{k-1}\omega_j^{\mathsf{T}}Q^{-1}\omega_j \\ & + \sum_{j=s}^{k}\frac{\left(g_{m,j} - \bar{g} - \xi_{2,j} - \xi_{3,j}\right)^2}{r} \\ \text{s.t. }\ & \zeta_{j+1} = \Phi_e\zeta_j + \Gamma_e\left(u_j - \bar{u}\right) + \Gamma_m\hat{d}_{j+1/2} + \omega_j \\ & K_u\left(u_{i,\max,0} - \bar{u}\right) \le \xi_{1,j} \le -K_u\,\bar{u}, \quad \xi_{3,j},\, a_j \ge 0\end{aligned} \tag{2.15}$$

The matrices $\Phi_e = \begin{bmatrix}\Phi & \gamma_\delta \\ 0 & 1\end{bmatrix}$, $\Gamma_e = [\gamma^{\mathsf{T}}\ 0]^{\mathsf{T}}$ and $\Gamma_m = [\gamma_m^{\mathsf{T}}\ 0]^{\mathsf{T}}$ restate the prediction of Equation (2.13), u_j is the rate held after reading j, the decisions are ζ_s and a noise ω_j on every state, and $Q = \mathrm{diag}(10^{-6}r,\, 10^{-6}r,\, 10^{-6}r,\, q_\delta)$ and r are the variances from which the filter was designed. The prior $\bar{\zeta}_s$ is the estimate the MHE itself made at execution $s-1$, propagated one period, and its weight $\bar{P}$ is the steady prior covariance of the filter, held constant. With $N_e = 10$ the window spans 1.68 h and eleven readings, which gives the dual active-set method of Section 2.5.4 44 decisions and 33 bound rows. The bounds come from declared values only. The insulin state stays between −194.1 and 77.5 mg/dL, the steady values of ξ_1 for the declared pump maximum $u_{i,\max,0} = 10\,\mu\mathrm{U/min}$ and for a

stopped pump, and the announced-meal glucose ξ_3 cannot fall below zero because a meal only adds glucose. The glucose deviation ξ_2 and the offset δ are left free, since glucose is the sum $\bar{g}+\xi_2+\xi_3$ and δ may stand for a meal, exercise or drift of either sign. Without an active bound the window problem has the filter's model, weights and prior, so it returns the filter's estimate. If a window's solution is suboptimal, contains a non-finite value or leaves the bounds, the estimate becomes a Kalman correction of the MHE's own one-period prediction, projected onto the bounds, and the history records the substitution. A first window, also one opened by a change of estimator during a run, starts from the filter's own prediction.

The third option, MHE with a meal state, gives ζ an unknown input of known shape, the appearance a of a meal nobody announced [23], which reaches glucose as $\tau_m\dot{\xi}_3 = -\xi_3 + K_m(\hat{d}+a)$ and fades as $\tau_a\dot{a} = -a+\omega_a$ with the 1 h decay of Equation (2.2). The noise ω_a has a standard deviation of 201.6 per period, what the default meal adds in one period of its ramp. A meal only adds glucose, so $a_j \geq 0$, a bound no Kalman filter can hold because its estimate is linear in the readings. With $q_\delta = 3.11\,(\mathrm{mg/dL})^2$, which lets δ drift by the 21.1 mg/dL of the largest basal drift in a day, and the rest unchanged, the window has 55 decisions and 44 bound rows, and its arrival cost is the augmented filter's steady prior, so without bounds it returns that filter's estimate, to 1.2×10^{-9} of a standard deviation, and as ω_a vanishes at $q_\delta = 225$ the MHE above, to 3.3×10^{-6}. The plan predicts on the same model, so it expects $\hat{a}$ to fade. Estimates and plans used 0.62 % of a core in Chrome at 1×.

2.5.6 Relay Autotuning

PI constants can also be measured from a limit cycle by the relay feedback experiment of [24], which the reader starts from the PID in automatic and which no shipped case runs. In place of the PID, a relay holds the pump at $u_0+\delta_r$ until the error e_k of Equation (2.6) falls below $-\varepsilon_r$, then at $u_0-\delta_r$ until it rises above ε_r, and so on. The center u_0 is the mean rate applied over the last 14 executions (2.35 h), the amplitude δ_r is 5 % of the pump span, 0.5 µU/min, or 0.9 of the distance from u_0 to the nearer stop when that is smaller, and the band $\varepsilon_r = 3\sigma_0 = 4.5\,\mathrm{mg/dL}$ is three times the declared CGM noise σ_0. A cycle in which either level lasts fewer than four executions is discarded. Once the period and the swing $\pm\alpha$ of the reading repeat within 5 % over two consecutive accepted cycles, the first harmonic of the applied levels, corrected for the band, gives the ultimate gain K_{cu}, and the Tyreus–Luyben rule turns it and the period P_u into PI constants, as Equation (2.16) shows.

$$K_{cu} = \frac{4\delta_r \sin(\pi D)}{\pi\sqrt{\alpha^2-\varepsilon_r^2}}, \qquad K_c = \frac{K_{cu}}{3.2}, \qquad \tau_I = 2.2\,P_u \tag{2.16}$$

Here D is the share of the cycle during which the relay sits at its upper level, one half only when u_0 is the rate that holds the setpoint. The constants are entered to three significant figures with $\tau_D = 0$ and remain editable. An unconverged test stops at 30 cycles or 48 h, or earlier at the reader's request, and in every outcome the PID, which has tracked the applied rate, takes the pump back without a bump. The band delays each switch, so the cycle settles where the loop's phase is $\psi = \arcsin(\varepsilon_r/\alpha)$ short of -180°, and a result with $\psi \geq 10^\circ$ is reported as placed by the band rather than by the insulin path.

2.6 Asymmetric Performance Metrics

A symmetric error integral scores an excursion of a given size the same way whichever side of setpoint it falls on. That score is not appropriate for this loop, and the reason is clinical rather than mathematical. Glucose measurements are usually summarized by the fraction of time spent inside a target band, and the international consensus recommendation reports that fraction together with separate thresholds for time below 70 mg/dL and time below 54 mg/dL, because the two low thresholds carry different consequences and neither is the mirror of the corresponding high threshold [25]. Time in range has been validated against microvascular outcomes and is now used as a trial endpoint [26]. The asymmetry was recognized in the measurement scale itself long before it was recognized in the outcome measure, and a transformation that symmetrizes the glucose axis about its clinical center was proposed for exactly that purpose [27].

> ✦ DID YOU KNOW? **A Small Allowance Below the Range**
>
> An international consensus of 2019 set targets for most adults with type 1 or type 2 diabetes: more than 70 % of sensor readings between 70 and 180 mg/dL, under 25 % above 180 mg/dL, under 4 % below 70 mg/dL, and under 1 %, less than 15 min a day, below 54 mg/dL [25].

The application shades a band of $\pm 20\,\mathrm{mg/dL}$ about the setpoint on the glucose trend, 60 mg/dL to 100 mg/dL at the default setpoint, and flags glucose below 64 mg/dL as low and above 180 mg/dL as high. The analysis here uses a fixed teaching band from 64 mg/dL to 104 mg/dL, whose lower edge is the application's low flag. That band is narrower and lower than the 70 mg/dL to 180 mg/dL consensus target range, so it is a tighter test than clinical practice and not a clinical target. Both bands are reported in Section 2.7 to keep the two readings apart.

The results tables therefore carry four numbers rather than one. The integrated absolute error against setpoint is the symmetric score. The excursion areas $\int \max(0, g-104)\,\mathrm{d}t$ and $\int \max(0, 64-g)\,\mathrm{d}t$ are the two one-sided scores, and they are what an asymmetric objective would weight unequally. Any weighted combination $J = \int \max(0, g-104)\,\mathrm{d}t + w \int \max(0, 64-g)\,\mathrm{d}t$, the objective of Equation (1.10) with the band edges of this chapter, ranks the two sides differently when $w > 1$, and the results show that the ranking of two runs reverses at $w = 7.9$. A specification written as a band rather than a setpoint recurs in Volume 9, Chapter 9, where a comfort band leaves the zone setpoint free.

2.7 Results

2.7.1 Protocol

Every run in this section starts from the application's reference operating point, case 1 of Table 2.3, which loads in manual mode with the pump at 3.0 µU/min, the setpoint at 76.22 mg/dL, the scheduled meals off and the sensor noise at zero. The settings listed in Table 2.4 are chosen before the run starts, so they set the starting state, and every step and upset is made after the run has started. The runs were driven headless through the application's own simulation engine, one 7.2 sec step at a time, exactly as the browser drives it. Time fractions, excursion areas, integrated absolute error and pump travel are computed from the logged noise-free glucose and pump columns of the exported file, which carries one row per control period of 0.168 h. Glucose extremes are taken at the integration step, because the logged rows miss the sharp meal peaks by up to 5.5 mg/dL. The application's own scorecard computes the integrated absolute error on the measured glucose, which without sensor noise equals the true glucose, and for the noise-free closed-loop day it reports the same 95.2 mg/dL h, pump travel and time at a limit as computed here. The noisy run uses the default noise with the case's fixed seed and repeats exactly. The configurations are listed in Table 2.4 and the outcomes in Table 2.5.

The app ships the eleven scenarios of Table 2.3, each credited by a completion rule of the kind described in Section 1.6.4, and the runs below use several of them by number.

2.7.2 Open-Loop Reference Day

With the pump held at the basal 3.0 µU/min through three meals of amplitude 1200, glucose reaches 155.5 mg/dL after breakfast, 157.0 mg/dL after lunch and 155.7 mg/dL after dinner, each about one hour after the meal begins. It is still at 80.3 mg/dL when lunch begins four hours after breakfast, and it is back within 0.6 mg/dL of the basal value before dinner and at the end of the day. The day spends 78.9 % of its samples inside the 64 mg/dL to 104 mg/dL band and none of them below it, and the area above 104 mg/dL is 122.5 mg/dL h. The repeatability of the three peaks to within 1.5 mg/dL is a property of the model rather than of a patient, and it is what makes the open-loop day a usable reference case.

Table 2.3 Scenarios of the App and What Each Requires

#	Scenario	Kind	Required to complete
1	Reference operating point	Experiment	Run 6 h in Manual at the basal steady state (pump 3.0 µU/min, glucose 76.22 mg/dL, meals off, no sensor noise) and note which states stay stationary.
2	Open-loop actuator step	Experiment	Starting in Manual at 80 mg/dL with the pump at 2.85 µU/min, run through the pump step to 4.85 µU/min at 2 h and on to 14 h.
3	Closed-loop setpoint step	Experiment	Starting in Auto at 80 mg/dL, run through the setpoint step to 101 mg/dL at 2 h until glucose has stayed within $\pm 5\,\mathrm{mg/dL}$ of the new setpoint for 1.2 h.
4	Disturbance rejection	Experiment	Starting in Auto at 80 mg/dL, run through a meal of amplitude 1800 at 2 h and on to 14 h.
5	Recover by hand	Exploration	Starting in Manual at 80 mg/dL with the pump at 2.85 µU/min, move the pump by hand after the same meal and bring glucose back within $\pm 20\,\mathrm{mg/dL}$ of 80 mg/dL without letting it pass 180 mg/dL, with the pump at a rate that holds glucose inside that band once settled, and keep it there for 1.2 h.
6	Conservative tuning	Experiment	In Auto with $K_c = 0.12$, run through the setpoint step from 80 to 96.8 mg/dL at 2 h and on to 14 h.
7	Aggressive tuning	Experiment	As case 6 with $K_c = 0.6$.
8	Actuator limit	Demonstration	Starting in Auto at 80 mg/dL, run through the setpoint step to 120 mg/dL at 2 h until the pump reaches a limit, and on to 8 h.
9	MPC with an announced meal	Experiment	In Auto at 80 mg/dL under the MPC with meals announced, run from the start through the planned meal of amplitude 1800 at 2 h to 8 h, then switch to PID, press Reset and run to 8 h again.
10	Unannounced meal, MHE + meal or Kalman	Experiment	As case 9 with the meal not announced and the MHE + meal estimator, then with the estimator switched to Kalman before Reset.
11	Challenge	Challenge	After the meal of amplitude 1800 at 2 h, keep measured glucose within $\pm 40\,\mathrm{mg/dL}$ of 80 mg/dL from 2.64 h to 4 h, the last third of a 4 h run, with the model parameters and disturbances as given.[a]

[a] With the shipped tuning ($K_c = 0.06$, $\tau_I = 3\,\mathrm{h}$) glucose peaks at 150.3 mg/dL and the worst measured error over the judged window is 70.6 mg/dL, 30.6 mg/dL outside the band. The suggested tuning changes only the gain, to $K_c = 0.3$, and the worst error is then 33.4 mg/dL, inside the band with 6.6 mg/dL to spare, in one run whether the tuning is chosen before Run or carried through Reset.

Table 2.4 Configurations of the Reported Runs

Label	Configuration
Basal	Manual, pump held at 3.0 µU/min, meal schedule on with $A = 1200$, 24 h
PI day	Automatic, $K_c = 0.3$, $\tau_I = 3\,\mathrm{h}$, $\tau_D = 0$, $g^{\mathrm{SP}} = 80\,\mathrm{mg/dL}$, meal schedule on, 24 h, starting at 80 mg/dL
Noisy	As PI day with the default CGM noise $\sigma = 1.5\,\mathrm{mg/dL}$
Large meal	As PI day with $A = 2400$
Exercise	As PI day, exercise raised to 60 % at $t = 20\,\mathrm{h}$ (in effect at 20.16 h), run to 26 h
Step test	Manual, meals off, pump moved 3.0 µU/min to 4.0 µU/min at $t = 2\,\mathrm{h}$ (in effect at 2.016 h), run to 14.1 h
Drift, $\tau_I = 3\,\mathrm{h}$	Automatic at $g^{\mathrm{SP}} = 76.22\,\mathrm{mg/dL}$, meals off, basal drift +30 % at $t = 2\,\mathrm{h}$ (in effect at 2.016 h), run to 14.1 h
Drift, P only	As above with τ_I at its 48 h top, which switches integral action off

2.7.3 Closed-Loop Day

The shipped tuning of $K_c = 0.3$, $\tau_I = 3\,\mathrm{h}$ and $\tau_D = 0$ holds glucose between 67.2 mg/dL and 103.1 mg/dL across the same day, which is 100.0 % of the samples inside the 64 mg/dL to 104 mg/dL band against 78.9 % open loop, and it reduces the integrated absolute error against the 80 mg/dL setpoint from 345.5 to 95.2 mg/dL h. The three peaks fall from the open-loop range 155.5 mg/dL to

Table 2.5 Measured Outcomes of the Reported Runs. Glucose Extremes Are Taken at the Integration Step and the Other Statistics From the Logged Noise-Free Glucose. Excursion Areas and IAE Are in mg/dL h, With IAE Against 80 mg/dL Except in the Drift Runs, Which Use Their 76.22 mg/dL Setpoint. Pump Travel Is $\sum |\Delta u_i|$ in µU/min, and the Last Column Is the Percentage of Samples at Either Pump Stop

	Glucose (mg/dL)		Time (% of run)				Excursion area		Pump		
Run	min	max	64–104	70–180	< 64	< 54	above 104	below 64	IAE	travel	at limit
Basal	76.2	157.0	78.9	100.0	0.0	0.0	122.5	0.0	345.5	0.0	0.0
PI day	67.2	103.1	100.0	94.4	0.0	0.0	0.0	0.0	95.2	59.7	1.4
Noisy	67.0	103.4	100.0	93.7	0.0	0.0	0.0	0.0	104.5	124.9	2.8
Large meal	64.4	130.9	87.3	90.8	0.0	0.0	56.2	0.0	210.3	58.6	14.1
Exercise	47.8	103.1	97.4	91.6	2.6	1.3	0.0	7.1	118.5	73.0	4.5
Step test	57.7	76.2	23.8	17.9	76.2	0.0	0.0	59.9	—	1.0	0.0
Drift, $\tau_I = 3$ h	71.7	96.3	100.0	100.0	0.0	0.0	0.0	0.0	11.8	15.7	0.0
Drift, P only	73.1	96.3	100.0	100.0	0.0	0.0	0.0	0.0	33.7	16.1	0.0

157.0 mg/dL to 101.0 mg/dL to 103.1 mg/dL, and they occur 0.42 h to 0.45 h after each meal begins rather than an hour after. On the clinical 70 mg/dL to 180 mg/dL band the ordering reverses, 94.4 % closed loop against 100.0 % open loop, because each postprandial descent carries glucose below 70 mg/dL. The pump reaches its 10 µU/min limit on 1.4 % of the samples, one after breakfast and one after dinner, and its lowest commanded value is 0.63 µU/min.

The upper excursion at this tuning is shaped jointly by the pump limit and the proportional gain. The breakfast peak occurs with the pump at its limit and the lunch and dinner peaks with the pump at 9.93 and 9.31 µU/min, so the controller is close to asking for what the device can deliver. Raising K_c to 0.45 still lowers the highest peak from 103.1 mg/dL to 99.8 mg/dL and the integrated absolute error to 74.7 mg/dL h, at the cost of 4.9 % of the samples with the pump at zero and 6.3 % at its maximum.

> ✦ **DID YOU KNOW? Sensor Noise Doubles the Pump's Work**
>
> From the simulator: a sensor noise with a standard deviation of only 1.5 mg/dL leaves the closed-loop day 100.0 % inside the 64 mg/dL to 104 mg/dL band, yet the pump's travel over the day rises from 59.7 to 124.9 µU/min (Table 2.5).

2.7.4 Large Meal and Saturation at the Upper Limit

Doubling the meal amplitude to 2400 with the same tuning raises the peaks to 128.5 mg/dL to 130.9 mg/dL and leaves 12.7 % of the day above 104 mg/dL, with an area above the band of 56.2 mg/dL h. The pump now sits on its upper limit for seven consecutive samples after breakfast and lunch and six after dinner, which is 1.0 h to 1.2 h each time, and it falls to a minimum of 0.61 µU/min without reaching zero. In this case the loop is authority-limited rather than tuning-limited. Doubling K_c lowers the highest peak by only 3.3 mg/dL, to 127.6 mg/dL, while doubling the pump maximum to 20 µU/min lowers it by 8.3 mg/dL, to 122.6 mg/dL, and with the shipped settings the manipulated variable spends 14.1 % of the day at its upper stop.

2.7.5 Exercise Disturbance

Raising the exercise disturbance to 60 % at $t = 20$ h, in effect from the controller execution at 20.16 h, multiplies s_i by 1.6 with no change in any measured quantity. Glucose falls from 74.9 mg/dL to 47.8 mg/dL within 0.04 h, before the controller next executes, because the glucose balance has a time constant of about one minute. The next execution drives the pump to zero, where it stays for 0.84 h, and glucose returns above 64 mg/dL 0.80 h after the change. The run spends 2.6 % of its samples below 64 mg/dL and 1.3 % below 54 mg/dL, and the area below the band is 7.1 mg/dL h.

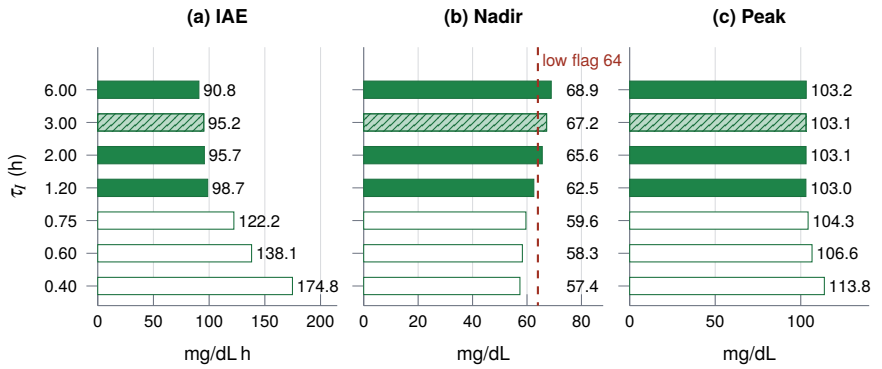


Figure 2.3 The integral gain sweep of Table 2.6: (a) integrated absolute error, (b) glucose nadir with the 64 mg/dL low flag (dashed), and (c) postprandial peak, for each integral time at $K_c = 0.3$. Hatched: the shipped 3.00 h. Solid: the other rows of the fivefold range from 6.00 to 1.20 h. White: the three faster rows. Notice that across the solid and hatched rows the IAE bars barely change and the peaks do not move, while the nadir walks down across the low flag.

Table 2.6 Integral Gain Sweep Over a Full Day at $K_c = 0.3$

K_I	τ_I (h)	nadir	peak	IAE	travel	% at zero
0.05	6.00	68.9	103.2	90.8	59.2	0.0
0.10	3.00	67.2	103.1	95.2	59.7	0.0
0.15	2.00	65.6	103.1	95.7	61.5	0.0
0.25	1.20	62.5	103.0	98.7	68.5	2.1
0.40	0.75	59.6	104.3	122.2	79.3	8.5
0.50	0.60	58.3	106.6	138.1	88.4	9.9
0.75	0.40	57.4	113.8	174.8	121.0	14.1

Pause. Try Exercise 2.2 before reading on.

The symmetric score misranks this run. Its integrated absolute error is 118.5 mg/dL h against 210.3 mg/dL h for the large-meal case, so the symmetric metric ranks it as the better of the two runs. The large-meal run never falls below 64 mg/dL, while the exercise run reaches 47.8 mg/dL and then holds the pump at zero for 0.84 h. Any asymmetric objective of the form given in Section 2.6 reverses the ranking once the low-side weight exceeds $w = 7.9$, which is the ratio of the two excursion areas 56.2 and 7.1.

2.7.6 Integral Gain Sweep

Repeating the closed-loop day at seven values of the integral time with $K_c = 0.3$ fixed separates the two costs, as listed in Table 2.6. Figure 2.3 draws the same rows as bars. Over a fivefold change in integral gain, from $\tau_I = 6$ h to 1.2 h, the integrated absolute error rises by only 8.7 %, from 90.8 to 98.7 mg/dL h, while the glucose nadir falls by 6.4 mg/dL, from 68.9 to 62.5 mg/dL, and the day begins to spend time below the 64 mg/dL flag. The postprandial peak moves by 0.15 mg/dL over the same range, which confirms that the high side is insensitive to the integral and the low side is not. Below $\tau_I = 1.2$ h both sides deteriorate, and at $\tau_I = 0.4$ h the peak reaches 113.8 mg/dL, the nadir 57.4 mg/dL, and the pump spends 14.1 % of the day at zero.

Within the sweep both objectives select the slowest integral, and switching integral action off at the 48 h top of the τ_I range lowers the integrated absolute error to 70.9 mg/dL h and raises the nadir to 72.1 mg/dL. On a day of meals with no unmeasured drift the integral action therefore carries no benefit, and its cost falls on the low side, where a faster integral keeps pushing insulin while glucose

is already falling and pulls the nadir below the band in the last four rows of Table 2.6. The integral is still needed, because Section 2.7.7 shows that only integral action returns glucose to setpoint after an unmeasured change in endogenous appearance. The shipped integral time of 3.0 h, 2.5 times the identified time constant, is a compromise between the two cases. It settles that drift within twelve hours while the day's nadir stays 3.2 mg/dL above the low flag.

Pause. Try Exercise 2.3 before reading on.

2.7.7 Basal Drift and Steady-State Offset

A 30 % step in the baseline appearance term with the meals switched off and the loop closed at the reference setpoint of 76.22 mg/dL raises the required infusion rate from 3.00 µU/min to 3.82 µU/min by Equation (2.5). Glucose rises to 96.3 mg/dL 0.17 h after the step, because the fast glucose balance reaches its new quasi-steady value before the insulin path can respond. With the shipped $\tau_I = 3$ h the loop returns to 76.25 mg/dL and 3.82 µU/min within the 12.1 h that follow, and to 76.22 mg/dL and 3.816 µU/min after 24 h, which matches Equation (2.5) to four significant figures. At the 48 h top of its range the integral-time setting switches integral action off, and the proportional-only loop settles at 78.55 mg/dL with the pump at 3.70 µU/min and stays there. The integrated absolute error over the 12.1 h after the step is 11.8 mg/dL h with $\tau_I = 3$ h and 33.7 mg/dL h without integral action.

The offset is small because the loop gain $K_c|K_p| = 5.6$ is well above unity. Equation (2.4) with the starting bias of 3.00 µU/min and $K_c = 0.3$ places the proportional-only equilibrium at 78.55 mg/dL and 3.70 µU/min, an offset of 2.33 mg/dL, which the run reproduces to four significant figures. The bias is set only when a run starts, so for a drift that arrives during the run the integral term is the only mechanism in the application that removes the offset.

2.7.8 Sensor Noise and the Pump

Turning the continuous glucose monitor noise back on at its default 1.5 mg/dL standard deviation changes the glucose statistics very little, because the noise is a small fraction of the postprandial excursion. The total variation of the pump rate roughly doubles, from 59.7 to 124.9 µU/min summed over the day, and the pump reaches zero on one sample, 0.7 % of the day, where the noise-free run never does. Eight other noise seeds give a pump travel of 109 to 124 µU/min and zero or one sample at zero, so the doubling is robust and the single sample at zero is within the spread between seeds. A proportional gain of 0.3 converts a 1.5 mg/dL measurement deviation into a 0.45 µU/min command deviation, which is 15 % of the basal rate. Adding derivative action with $\tau_D > 0$ magnifies the effect, because the second difference of the measurement is divided by $\Delta t = 0.168$ h.

2.7.9 Application

The app offers the manual and automatic modes, scorecard and data export of Section 1.6.4, and its scorecard reports the integrated absolute error, pump travel and time at a limit. The interface is shown in Figure 2.4, with a schematic beside the glucose reading and trends of glucose, infusion rate and insulin action.

2.8 Predictive Control and the Announced Meal

Table 2.7 runs cases 3, 4, 8 and 9 over 24 h and the challenge over 4 h, the controller chosen before each run, so the PID and the MPC face identical events and CGM noise within a case. Unlike Section 2.7, every value is read on the 0.168 h log, as the scorecard reads it. The PID keeps $K_c = 0.3$, $\tau_I = 3$ h and $\tau_D = 0$, the challenge shipping $K_c = 0.06$ and suggesting 0.3, while the MPC holds the values given in Section 2.5.4, with meals announced only in case 9. Cases 4 and 9 schedule the same meal of amplitude 1800 at 2 h, so case 4 is the like-for-like comparison and case 9 adds only the announcement.

At the 101 mg/dL of case 3 the tangent gain of Equation (2.4) is 1.6 times its value at 80 mg/dL, where every case starts and the MPC's model was fitted. The PID holds the pump at zero for 0.84 h,

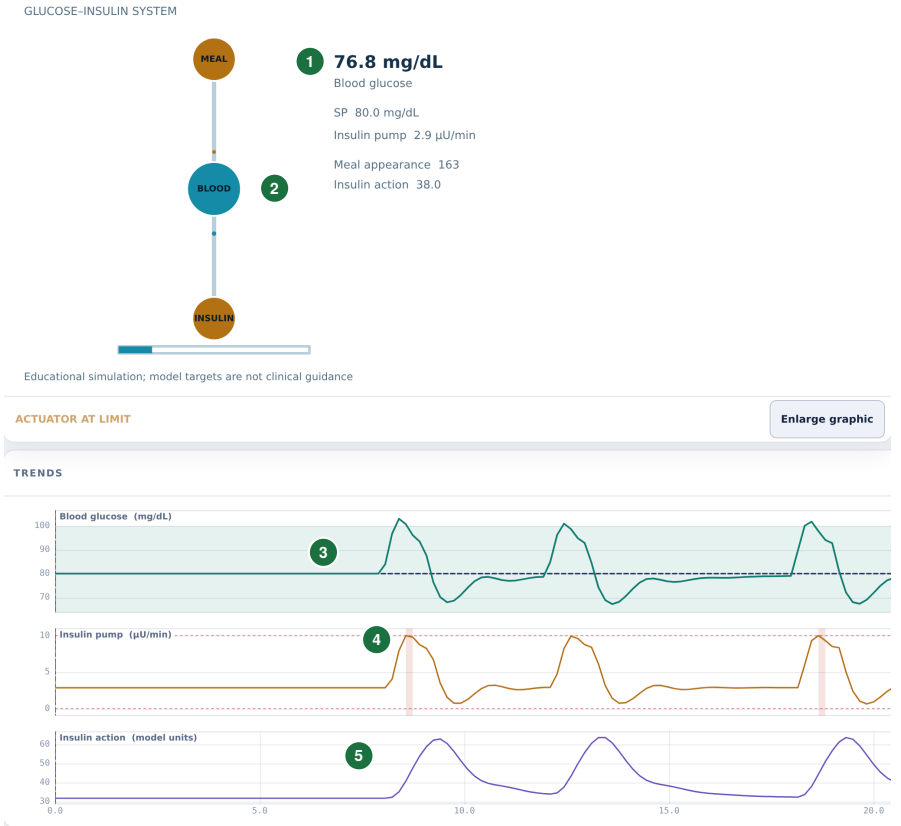


Figure 2.4 Closed-loop day under the shipped tuning, 21 h into the run with the 24 h trend window, from the reference case with the meal schedule on, the setpoint at 80 mg/dL and automatic control chosen before the run, so glucose holds 80 mg/dL until breakfast. The plant view and the trends are shown, cropped just before the end of the trends. (1) The glucose reading, with the setpoint and the pump rate. (2) The schematic of meal, blood and insulin. (3) The glucose trend with its shaded band. (4) The pump trend at its limit. (5) The insulin action. The three postprandial peaks rise just above the shaded ± 20 mg/dL band, the pump touches its 10 µU/min limit after breakfast and dinner, and the insulin action peaks about 0.9 h after the pump. At 21 h glucose is 76.8 mg/dL and the pump delivers 2.9 µU/min.

overshoots to 109.4 mg/dL and stays within 5 mg/dL only from 11.6 h after the step, while the MPC keeps the pump above 0.75 µU/min, peaks at 105.8 mg/dL and stays within 5 mg/dL from 0.69 h. At the 120 mg/dL of case 8 the gain is 2.25 times that value and the PID never settles, its pump at zero for 6.9 h and glucose still between 105.7 and 133.3 mg/dL after 8 h. The first three plans there bind the pump's floor, and the MPC spends 0.34 h at zero, peaks at 129.5 mg/dL, stays within 5 mg/dL from 1.36 h and over the last 12 h averages the 1.87 µU/min that Equation (2.5) requires, with $\hat{\delta}$ averaging 13.3 mg/dL.

The MPC loses case 4. The meal shows clearly in the reading of 2.184 h, from which the proportional term of the PID holds the pump at its maximum. The logged $\hat{\delta}$ climbs from 38 to 101 mg/dL by 3.19 h, but the plans, which expect it to persist and pay for each move, raise the pump through 5.4, 7.4, 9.1 and 9.6 µU/min to the maximum only from 2.856 h, and glucose peaks at 130.9 mg/dL against 116.8. On the descent $\hat{\delta}$ is still 85 mg/dL at 3.36 h, the MPC holds 6.7 and 5.7 µU/min where the PID is down to 4.4 and 2.3, and glucose falls to 64.9 mg/dL against 66.2, until two plans stop the pump on the low side of the band.

Table 2.7 Five Cases Under the Shipped PID and the Predictive Controller, With the Same CGM Noise in Every Run of a Case

Case	Controller	IAE (mg/dL h)	IAE_g	g_{max} (mg/dL)	g_{min}	$t_{<70}$ (h)	Travel	Zero (%)	Max (%)
3, setpoint	PID	67.65	58.31	109.4	79.1	0	125.1	3.5	0
	MPC	46.93	29.26	105.8	79.8	0	30.8	0	0
4, meal	PID	83.67	68.97	116.8	66.2	0.67	101.9	0.7	4.2
	MPC	98.98	82.54	130.9	64.9	0.84	51.1	1.4	1.4
8, limit	PID	164.17	160.28	133.3	79.1	0	219.2	28.9	0
	MPC	64.42	48.84	129.5	79.8	0	34.1	1.4	0
9, announced[a]	MPC	50.94	34.42	98.6	75.0	0	41.9	0	2.1
11, challenge	PID, $K_c = 0.06$	70.89	68.00	148.6	77.5	0	10.0	0	0
	PID, $K_c = 0.3$[b]	45.79	42.70	116.8	66.2	0.50	28.3	4.3	26.1
	MPC	56.50	53.12	130.9	64.9	0.34	20.2	4.3	8.7

IAE uses the CGM reading, as the scorecard does, and IAE_g the true glucose. g_{max} and g_{min} are the highest and lowest logged true glucose from the event at 2 h, $t_{<70}$ the logged time below 70 mg/dL, Travel the summed absolute change of the infusion rate in µU/min, and Zero and Max the time with the pump at zero and at its 10 µU/min maximum as shares of the run. No run exceeds 180 mg/dL. [a]PID as in case 4. [b]Suggested, $\tau_I = 3$ h.

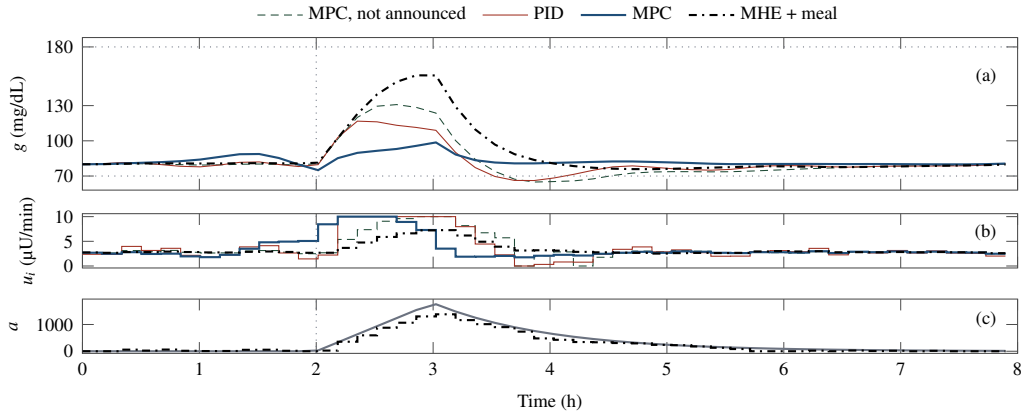


Figure 2.5 Case 9 run three ways from the same CGM noise, 8 h of the 0.168 h log, with the meal beginning at 2 h (dotted), and case 10, the unannounced run with the meal-state MHE in place of the Kalman filter. (a) True glucose and the 70 and 180 mg/dL sides of the band. (b) Infusion rate, each value drawn over the period it was held. (c) True meal appearance (gray) and $\hat{a}$, each drawn over the period after it was made.

> **TRY IT IN THE APP** **4 · Disturbance Rejection**
>
> Load scenario 4 and press Run. A meal of amplitude 1800 begins at 2 h. Watch the glucose trend: under the shipped PID it peaks at 116.8 mg/dL, and on the way down it falls to 66.2 mg/dL, below 70 mg/dL for 0.67 h (Table 2.7). Then watch the pump trend reach its maximum after the meal.

Told from the start of case 9 that the meal begins at 2 h, the twelve plans made before it each predict a lowest glucose of exactly 70 mg/dL, so the band decides how much insulin can go in early (Figure 2.5). They first lower the pump to 1.79 µU/min, letting glucose rise to 88.8 mg/dL, then raise it to 5.07 µU/min at 1.848 h, so glucose is 75.0 mg/dL in the first sample of the meal, which meets insulin already acting. Glucose peaks at 98.6 mg/dL, or 99.8 at the integration step, and stays above 78.7 mg/dL afterward. Over the first 8 h the MPC delivers 28.1 against 28.8 µU/min h of insulin, nearly the same amount given earlier, which costs 6.7 against 2.3 mg/dL h of true IAE before the meal and returns 18.7 against 52.2 over the next 6 h.

Table 2.8 The Meal of Cases 4 and 9 Under Each Estimator Over 24 h

	Case 4, not announced						Case 9	
Controller	g_{max}	g_{min}	$t_{<70}$	IAE	IAE_g	Area	IAE	IAE_g
PID	117.4	65.8	0.61	83.7	69.4	—	83.7	69.4
MPC, Kalman or MHE	131.1	64.9	0.89	99.0	82.6	—	50.9	34.1
MPC, MHE + meal	157.9	75.8	0	114.2	103.9	2730	78.6	72.4
$q_\delta = 12.4$	155.4	71.7	0	118.4	107.1	2292	67.4	57.7
$q_\delta = 44.5$	149.7	67.6	1.41	121.7	108.0	1668	60.8	48.6
$q_\delta = 225$	140.5	65.5	1.05	110.6	95.1	812	54.3	39.7
no sign bounds	153.4	72.7	0	138.5	131.6	298	49.1	31.9
MPC, Kalman[a]	142.1	61.6	2.20	123.3	106.8	—	48.4	31.9
PID, $\hat{a}$ fed forward	117.5	60.6	0.94	88.2	74.3	2951	88.8	75.0

One noise seed. Extremes (mg/dL) and $t_{<70}$ (h) at the integration step from 2 h, IAE as in Table 2.7, Area of $\hat{a}$ in units h (true 2700). [a] $q_\delta = 3.11$. No run leaves 54 mg/dL to 180 mg/dL, and in case 9 only the PID runs fall below 70.

> **TRY IT IN THE APP** **9 · MPC With an Announced Meal**
>
> Load scenario 9 and press Run. The predictive controller knows from the start that a meal begins at 2 h. Watch the pump before the meal: it first drops to 1.79 µU/min, then rises to 5.07 µU/min at 1.848 h, so insulin is already acting when the meal arrives. Glucose then peaks at 98.6 mg/dL.

With only the band removed, in a run outside the application, the announced plans give up to 7.3 µU/min before the meal without lowering the pump first, glucose is 67.6 mg/dL at the meal, below 70 for 0.17 h, and the IAE falls to 43.8 mg/dL h, while case 4 falls to 58.7 mg/dL with 2.18 h below 70. Pricing both sides alike, at 100 or 1000, reproduces both cases to every digit shown, since no plan finds the 180 mg/dL side active and every active low side has a slack of at most 1.1×10^{-4} mg/dL or the pump at zero.

Unaffected by K_c, the MPC repeats in the challenge its first 4 h of case 4, and its largest reading error from 2.64 to 4 h, 50.5 mg/dL, fails the $\pm 40\,\mathrm{mg/dL}$ test as the shipped gain does with 70.6, while the suggested $K_c = 0.3$ passes with 33.4. All plans were optimal, 143 per 24 h case, and their mean solve time, a wall-clock figure from one desktop computer, was 0.05 to 0.2 ms under Node.js and 216 and 206 µs in two Chrome runs of case 9, 0.13 % of one core at the default speed. The MPC wins cases 3, 8 and 9 and loses the two with a meal it is not told of.

2.9 Estimator Choice and Relay-Written Constants

The MHE without a meal state repeats the filter's run in every case, within 4.8×10^{-8} µU/min over 72 h, while the meal state changes it. Case 10 is case 4 with that estimator, and Table 2.8 compares the estimators on its meal, unannounced as in case 4 and announced as in case 9.

Unannounced, the meal-state MPC never falls below 70 mg/dL and pays at the peak, 157.9 against 131.1 mg/dL under the filter. Its $\hat{a}$ peaks at 1381 against a true 1800, a root-mean-square error of 77, and its plans, expecting the meal to pass, stop at 7.3 µU/min, while the filter books the rise as a lasting $\hat{\delta}$ and its plans keep insulin flowing as the meal fades (Figure 2.5). Over seeds 1001 to 1010 it stays above 74.8 mg/dL, against 0.90 to 1.38 h below 70 for the filter. Neither half suffices alone, since the meal state beside the quick $q_\delta = 225$ and the filter given the slow one both fall below 70, as does the filter's MPC at every λ and horizon tried, at best 0.67 h at $\lambda = 30$. The sign bounds buy the estimate, not the absence of the low, as unbounded $\hat{a}$ is negative at 111 of 142 estimates. Announced, it loses on every seed, a true IAE of 56.7 to 71.9 against 26.8 to 31.7, because $\hat{a}$ cannot fall below zero where the meal path overpredicts, and freeing a alone gives 55.3. The slow q_δ came after 225 was measured, and its 24 h scale is the trend window, so the table adds the rule over the 6 h horizon, which keeps the result, and the 1.68 h window, which loses it.

Over 74 runs at extreme settings of the controls, 24 after the meal of case 4 or 9, the meal-state MPC is worse on the low side than the filter's in 13, longer below 54 in one (0.43 against 0.39 h) and

Table 2.9 Shipped PID, Relay-Written PI and Two Counterfactual PI Loops on Cases 3 and 4 Over 72 h With One Noise Seed

Case	Controller	IAE	IAE_g	IAE_g^{24+}	$g_{\max}$	$g_{\min}$	$t_{<70}$	Travel
		(mg/dL h)			(mg/dL)		(h)	
3, setpoint	PID, shipped	192.7	172.8	114.5	110.3	78.8	0	379.9
	PI, relay	129.4	72.2	21.0	102.5	79.9	0	46.7
	PI[a]	186.3	165.3	109.1	109.8	78.9	0	374.0
	PI[b]	116.6	54.5	21.7	102.6	79.9	0	46.2
4, meal	PID, shipped	164.7	122.7	53.3	117.4	65.8	0.61	279.3
	PI, relay	205.0	150.1	16.2	152.5	70.2	0	52.0
	PI[a]	161.2	119.5	52.2	117.5	67.6	0.45	281.7
	PI[b]	210.4	151.4	15.6	151.5	66.2	2.28	51.8

PID, shipped: $K_c = 0.3$, $\tau_I = 3$ h. PI, relay: $K_c = 0.0578$, $\tau_I = 5.14$ h. [a]$K_c = 0.3$, $\tau_I = 5.14$ h. [b]$K_c = 0.0578$, $\tau_I = 3$ h. Neither counterfactual is shipped or written. IAE_g^{24+} is the true IAE after the first 24 h. Glucose extremes and $t_{<70}$ are taken at the integration step from the event at 2 h. Other columns as in Table 2.7.

above 180 in seven (up to 2.30 against 0.84 h), since a slow $\hat{\delta}$ and a nonnegative $\hat{a}$ correct a fall that is no meal late. In all twelve added runs after the unannounced meal it is safer, exercise at 40 % at 4 h leaving 56.1 against 47.0 mg/dL.

With the setpoint step of case 3 deleted, the relay test began at 2.4 h, glucose steady at 80 mg/dL and u_0 at 2.87 µU/min, against the 2.853 µU/min that holds that glucose, so δ_r took its full 0.5 µU/min. It converged after three cycles and 5.83 h with $\alpha = 5.64\,\text{mg/dL}$, $D = 0.457$, $P_u = 2.34$ h and $K_{cu} = 0.185$ in µU/min per mg/dL, and it wrote $K_c = 0.0578$ and $\tau_I = 5.14$ h, a fifth of the shipped gain and 1.7 times its integral time, flagged as band-limited at $\psi = 52.9°$. The ten seeds 1001 to 1010 converged as well, with K_{cu} from 0.114 to 0.176, P_u from 2.22 to 3.03 h and ψ from 38.8° to 52.1°.

That result reflects the switching band and the CGM noise rather than the insulin path, as the relay test of Volume 2, Chapter 2 does on a vehicle. The two lags of Equation (2.12) approach a phase of −180° without crossing it, and only the zero-order hold at Δt gives the sampled model an ultimate point, at a period of 0.77 h and a gain of 0.821. The measured cycle is three times slower, and its K_{cu} is 23 % of that gain. The same relay returns 0.279 at 2.69 h on the noise-free model and 0.278 at 2.69 h on the simulator with the CGM noise off, where $D = 0.527$ and $\psi = 63.1°$. Noise lowers K_{cu} because α lies close to ε_r. A swing 12 % wider, 5.64 against 5.05 mg/dL, makes $\sqrt{\alpha^2 - \varepsilon_r^2}$ in Equation (2.16) half as large again.

The written constants exchange the meal peak for the low side, as listed in Table 2.9. On the setpoint step the scorecard IAE falls from 193 to 129 mg/dL h and the true IAE from 173 to 72.2, mostly after the first day, where the shipped gain passes the CGM noise to the pump (travel 380 against 46.7) and the true glucose departs from 101 mg/dL by 2.4 mg/dL on average over the last 48 h against 0.44. On the meal the highest glucose rises from 117.4 to 152.5 mg/dL and the scorecard IAE from 165 to 205, while the lowest glucose rises from 65.8 to 70.2 mg/dL and the 0.61 h below 70 mg/dL vanishes. The counterfactual rows separate the two effects. The shipped K_c with the written τ_I keeps the peak at 117.5 mg/dL with 0.45 h below 70, and the written K_c with the shipped τ_I repeats the high peak, 151.5 mg/dL, with 2.28 h below 70. The peak follows the proportional gain, and the descent stays above 70 mg/dL only when the small gain is paired with the slow integral, an integral gain K_c/τ_I of 0.011 against the shipped 0.1, which is the low-side effect of Table 2.6 at a different gain. Over 21 relay tests (the shipped seed and 1001 to 1020) the written constants won the setpoint step every time, with scorecard IAE from 119 to 192, and lost the meal every time, with 186 to 235 mg/dL h against 165.

2.10 Technology in Practice

The symbols of Figure 2.1 stand, on a person, for a glucose sensor worn on the skin, a controller that runs in software, and an insulin pump that infuses under the skin. Figure 2.6 draws the loop with that hardware in place of the schematic's symbols.

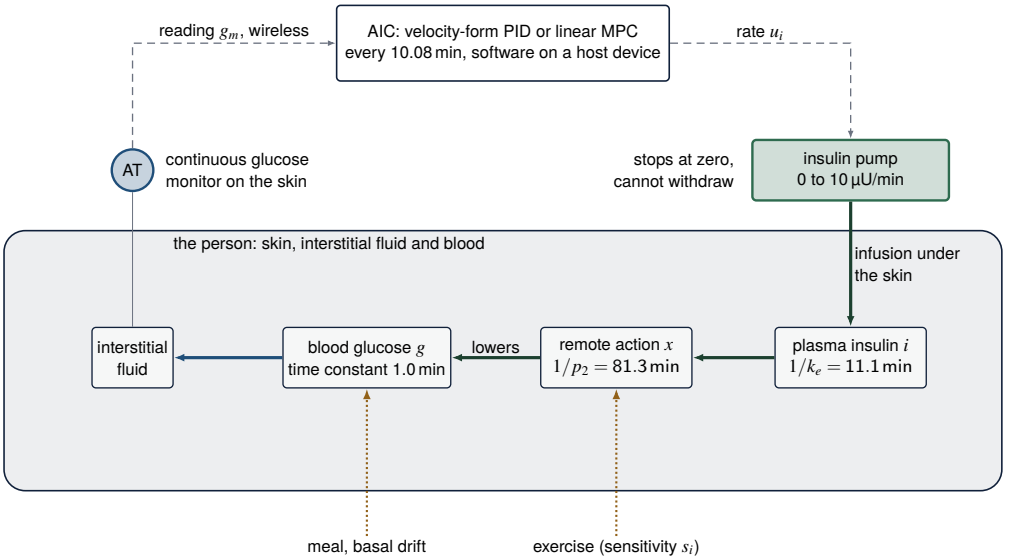


Figure 2.6 The loop of Figure 2.1 with the hardware of a wearable system. A continuous glucose monitor (AT) reads the interstitial fluid and sends its reading to the controller AIC, which every 10.08 min sets the rate of the insulin pump. Insulin reaches glucose through the plasma and remote action pools, whose time constants set the lag the loop must work through. Thick solid arrows carry insulin and glucose, thin dashed arrows signals, dotted arrows disturbances. In this model the pump feeds the plasma pool directly, and the sensor has no lag.

What Measures It

The controller reads one signal, the glucose that a continuous glucose monitor reports. A real monitor reads the interstitial fluid, and its data need filtering, denoising and artifact rejection before they can drive a controller [6]. A monitor meant to feed an automated system is held to accuracy limits written into regulation. Across its measuring range, the lower 95 % confidence bound on the share of readings within 20 % of blood glucose must exceed 87 %, and no reading above 180 mg/dL may correspond to a blood glucose below 70 mg/dL [28]. The chapter's monitor of Equation (2.3) keeps only the noise, with no lag and no calibration error.

> **TECHNOLOGY** **Sensor: The Continuous Glucose Monitor**
>
> A continuous glucose monitor places a small electrochemical sensor under the skin, where the enzyme glucose oxidase turns the glucose of the interstitial fluid into an electrical signal. It gives a reading every 1 to 5 min, and recent sensors reach a mean absolute relative difference below 10 % against a laboratory reference. The reading lags blood glucose by 5 to 10 min, the time glucose takes to pass from the blood into the interstitial fluid. Earlier sensors also suffered from imperfect calibration, from compression artifacts and from interference by drugs such as acetaminophen [29]. The chapter's monitor adds noise of 1.5 mg/dL standard deviation and has no lag.

What Moves It

The final element is a pump worn on the body that delivers insulin at set or variable basal rates and in boluses, and whose labeling states its minimum and maximum basal rates and its delivery increments [30]. Its fail position is the one the chapter studies: the pump can stop, and it cannot withdraw what it has given. The delay that a real pump adds sits in the subcutaneous depot, where insulin waits to be absorbed [6]. The chapter's pump has no depot, no delivery increment and no lag, as Section 2.11 states, so its insulin reaches the plasma pool of Equation (2.1c) at once.

TECHNOLOGY Final Element: The Insulin Pump

Continuous infusion of insulin under the skin has been an accepted therapy since the late 1990s. Pumps allow fine adjustment of the basal rate, and patch pumps are worn directly on the skin. The weak link is absorption. Regular insulin given under the skin peaks up to 120 min after a bolus, and altered absorption at the infusion site can underdeliver insulin, which may lead to ketoacidosis if it is not detected and corrected in time [6]. A pump that takes commands from an outside controller must detect occlusions and air in the line within a clinically relevant time and state its delivery accuracy over the first 1, 6 and 12 h [30]. The chapter's pump applies its rate at once between zero and 10 µU/min.

Where the Controller Runs

In current hybrid closed-loop systems the algorithm runs on one device of the system, and in one of them it is an app on a smartphone that talks to the sensor and the pump. The algorithms in use include model predictive control, PID control and fuzzy logic [31], and the regulation of an interoperable controller requires specifications for the device that hosts it [32]. A controller cannot act more often than its readings arrive, every 1 to 5 min from current sensors [29]. The chapter's controller acts every 10.08 min. Its structure is one block of software: the velocity-form PID with back-calculation of Equation (2.10), or the linear predictive controller with its Kalman filter, which hands the pump back to the PID when a plan fails.

IN PRACTICE How an Automated Insulin System Is Cleared

In the United States an automated insulin system can be assembled from three separately cleared devices: an integrated continuous glucose monitor (21 CFR 862.1355), an interoperable automated glycemic controller (21 CFR 862.1356) and an alternate controller enabled infusion pump (21 CFR 880.5730). The controller may take glucose only from such a monitor and command only such a pump. Each device must keep therapy safe when communication is interrupted or lost, for example by reverting to a preprogrammed safe delivery rate, and must log the commands and their confirmations, malfunctions such as a pump occlusion, alarms, and every loss of connection [30, 32].

What Is Modernizing It

Three developments are changing this loop. Interoperable devices let users combine a sensor, a pump and a controller into a closed-loop system of their own, and the three device classes of the in practice box above make that possible in regulation [31, 32]. Sensors now reach an accuracy close to that of fingerstick meters, and some can be used for ten days without calibration by the user [29]. Do-it-yourself closed-loop systems, built by user communities on open software without regulatory review, are used by several thousand people, and observational studies report gains in time in range, although no randomized trial had yet tested them at the time of the review [31].

2.11 Limitations

The application is an educational simulator. It is not a medical device, it is not validated against patient data, and no tuning obtained from it should be transferred to any clinical setting. The parameter set describes one virtual subject, and every result in Section 2.7 is a property of that parameter set.

The model omits the two largest sources of delay in a real system. There is no subcutaneous insulin depot between the pump and the plasma compartment, and there is no interstitial transport lag between plasma glucose and the reported measurement. Both delays are on the order of tens of minutes in a wearable device, and adding either would make the loop substantially harder to stabilize than the one measured here. The identified apparent dead time of 0.16 h is a consequence of the insulin cascade and not of any transport delay in the simulator.

The transit chain of Equation (2.1d) to Equation (2.1f) is driven by the infusion rate rather than by the meal, which follows the published assignment code. A model in which the meal drives the chain would place the meal disturbance behind two additional lags and would change both the shape and

the timing of the postprandial peak. The application's own model notes state the same routing, with the meal entering the blood directly.

The teaching band of 64 mg/dL to 104 mg/dL and the band of $\pm 20\,\mathrm{mg/dL}$ about the setpoint that the application shades are not the consensus clinical target range of 70 mg/dL to 180 mg/dL, and the low-glucose flag at 64 mg/dL is not either of the two clinical thresholds at 70 mg/dL and 54 mg/dL. Table 2.5 reports both bands so that the two readings cannot be confused.

The sensor noise comes from a generator reseeded with a fixed seed whenever a case starts, so a noisy run repeats exactly but represents one noise realization, and eight other seeds move the pump travel of the noisy day between 109 and 124 µU/min. The PID has no feedforward path and no meal announcement, and neither controller has an insulin-on-board constraint, all three of which are standard in deployed systems, so the closed-loop performance reported in Section 2.7 is that of feedback alone.

The predictive controller plans on a linear model frozen at 80 mg/dL, without the plant's insulin gain, 1.6 and 2.25 times larger at 101 and 120 mg/dL, or the smaller rise of a meal against insulin already acting, so its first plan in case 9 predicted a peak of 130.4 mg/dL where the plant, fed that plan open loop, peaks at 111.4 mg/dL. Its announcement is exact, its tenfold low-side price is never exercised by the compared cases, it has no insulin-on-board constraint, and it loses both cases with an unannounced meal.

The MHE constrains the states of the controller's linear model, never the plant's, so where the model is wrong the bounded estimate is wrong in the same way, and on this loop the filter crosses the one bound ever active by at most 6.2×10^{-7} mg/dL. The meal state knows one meal shape, a one-hour fade, so a longer meal is dosed as one about to pass, and its gain rests on a slow δ chosen after a first rule failed. The relay test reads a cycle placed by its 4.5 mg/dL band and by the CGM noise, at ψ of 38.8° to 52.9°, and converts it by one conservative rule, so its constants are a starting point that the meal case shows must still be judged against the low side.

> ⚠ **OPEN CHALLENGE** **Meals and Exercise Without Being Told**
>
> Rises in glucose after meals remain a challenge for closed-loop systems, because insulin given under the skin is absorbed slowly. Good control still needs the user to count carbohydrate and give a bolus before the meal, and attempts to lighten that burden with simplified boluses or fully closed loops have compromised glucose control. Exercise raises the risk of low glucose and changes insulin sensitivity, so users usually announce it to the algorithm in advance [31]. Cases 4 and 9 of this chapter measure the same gap on one virtual patient. Faster-acting insulins, and extra signals such as heart rate or body motion that reveal activity sooner than glucose alone, are the candidates for progress [31].

2.12 Discussion

Faster insulin, a shorter sensor lag and a larger pump all attack the delay between a decision and its effect, and none of them changes the sign structure of the problem. Insulin will still only lower glucose, the pump will still stop at zero, and a low will still cost more than a high of the same size. That is why deployed systems treat the insulin already delivered as a state that limits what the controller may ask for next [11], and why the outcome measure the field settled on reports the two sides of a band separately [25]. The lesson generalizes past this plant. Any loop with a one-sided actuator and an unequal cost of the two error signs, such as emergency cooling, pressure relief or an override acting only in the safe direction, shows the pattern of Table 2.6, a symmetric index nearly flat across tunings over which the constrained side moves a great deal, so tuning by that index selects a controller materially worse on the constraint.

2.13 Conclusion

This chapter studied Type 1 diabetic glucose regulation on the six-state expanded Bergman minimal model, three meals on a twenty-four hour clock, an exercise disturbance that scales insulin sensitivity

and a basal drift that shifts endogenous appearance, all stated in full for an independent rebuild. A fasting step test from 3.0 µU/min to 4.0 µU/min returns a process gain of −18.55 mg/dL per µU/min and a first-order plus dead-time fit of $\tau_p = 1.18$ h and $\theta_p = 0.16$ h, in which the apparent dead time is the plasma insulin clearance time constant rather than a transport delay. The shipped direct-acting velocity-form controller raises time inside the 64 mg/dL to 104 mg/dL band from 78.9 % to 100.0 % over a full day while reaching the pump limit on 1.4 % of the samples. A fivefold change in integral gain moves the integrated absolute error by 8.7 % and the glucose nadir by 6.4 mg/dL, which is the measured form of the claim that a symmetric metric cannot express the objective of this loop. Eleven shipped cases, from a reference operating point to a postprandial challenge, carry explicit completion requirements. Told of a meal, the predictive controller beside the PID scores an IAE of 50.9 against the PID's 83.7 mg/dL h, and 99.0 when not told. In place of its Kalman filter a moving horizon estimator moves no case by more than 4.8×10^{-8} µU/min, and one with a nonnegative meal state removes the 0.89 h below 70 mg/dL after the unannounced meal while the peak rises from 131.1 to 157.9 mg/dL. The relay-written constants, a fifth of the shipped gain, keep the meal case above 70 mg/dL while its peak climbs from 117.4 to 152.5 mg/dL.

Several things the simulator does not carry would change the numbers, including a subcutaneous insulin depot, an interstitial sensor lag, feedforward of an announced meal to the PID, an insulin-on-board constraint and more than one virtual subject. The app still shows direct action against a negative process gain, a one-sided actuator, a disturbance faster than the manipulated path, and a performance cost that is not symmetric, in a single run of a simulated day that takes 24 sec at the default speed.

KEY POINTS

- Insulin only lowers blood sugar and a pump cannot take it back, so the controller pushes one way only.
- The apparent dead time of the identified model is the plasma insulin lag, reported as a delay by the first-order fit.
- Once the pump is at zero, no retuning shortens a low.
- A fivefold change in integral gain moves the integrated absolute error by 8.7 % and the nadir by 6.4 mg/dL, so a symmetric score cannot express the objective of this loop.
- Told of a meal, the predictive controller scores an IAE of 50.9 against 83.7 for the PID.

Where else this applies. The one-way pump returns in Chapter 3, and anti-windup at a bound set by the supply in Volume 2, Chapter 3.

References

[1] D. E. Seborg, T. F. Edgar, D. A. Mellichamp, and F. J. Doyle, III, *Process Dynamics and Control*, 4th ed. Hoboken, NJ: John Wiley & Sons, 2016.

[2] B. W. Bequette, *Process Control: Modeling, Design, and Simulation*. Upper Saddle River, NJ: Prentice Hall, 2003.

[3] K. J. Åström and R. M. Murray, *Feedback Systems: An Introduction for Scientists and Engineers*, 2nd ed. Princeton, NJ: Princeton University Press, 2021.

[4] E. F. Pfeiffer, C. Thum, and A. H. Clemens, "The artificial beta cell - a continuous control of blood sugar by external regulation of insulin infusion (glucose controlled insulin infusion system)," *Hormone and Metabolic Research*, vol. 6, no. 5, pp. 339–342, 1974.

[5] A. M. Albisser, B. S. Leibel, T. G. Ewart, Z. Davidovac, C. K. Botz, and W. Zingg, "An artificial endocrine pancreas," *Diabetes*, vol. 23, no. 5, pp. 389–396, 1974.

[6] C. Cobelli, E. Renard, and B. Kovatchev, "Artificial pancreas: Past, present, future," *Diabetes*, vol. 60, no. 11, pp. 2672–2682, 2011.

[7] G. M. Steil, K. Rebrin, C. Darwin, F. Hariri, and M. F. Saad, "Feasibility of automating insulin delivery for the treatment of type 1 diabetes," *Diabetes*, vol. 55, no. 12, pp. 3344–3350, 2006.

[8] R. S. Parker, F. J. Doyle, III, and N. A. Peppas, "A model-based algorithm for blood glucose control in type i diabetic patients," *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 2, pp. 148–157, 1999.

[9] R. Hovorka, V. Canonico, L. J. Chassin, U. Haueter, M. Massi-Benedetti, M. O. Federici, T. R. Pieber, H. C. Schaller, L. Schaupp, T. Vering, and M. E. Wilinska, "Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes," *Physiological Measurement*, vol. 25, no. 4, pp. 905–920, 2004.

[10] B. W. Bequette, "A critical assessment of algorithms and challenges in the development of a closed-loop artificial pancreas," *Diabetes Technology & Therapeutics*, vol. 7, no. 1, pp. 28–47, 2005.

[11] C. Ellingsen, E. Dassau, H. Zisser, B. Grosman, M. W. Percival, L. Jovanovič, and F. J. Doyle, III, "Safety constraints in an artificial pancreatic β cell: An implementation of model predictive control with insulin on board," *Journal of Diabetes Science and Technology*, vol. 3, no. 3, pp. 536–544, 2009.

[12] R. Hovorka, "Closed-loop insulin delivery: from bench to clinical practice," *Nature Reviews Endocrinology*, vol. 7, no. 7, pp. 385–395, 2011.

[13] S. A. Brown, B. P. Kovatchev, D. Raghinaru, J. W. Lum, B. A. Buckingham, Y. C. Kudva, L. M. Laffel, C. J. Levy, J. E. Pinsker, R. P. Wadwa, E. Dassau, F. J. Doyle, III, S. M. Anderson, M. M. Church, V. Dadlani, L. Ekhlaspour, G. P. Forlenza, E. Isganaitis, D. W. Lam, C. Kollman, and R. W. Beck, "Six-month randomized, multicenter trial of closed-loop control in type 1 diabetes," *New England Journal of Medicine*, vol. 381, no. 18, pp. 1707–1717, 2019.

[14] B. P. Kovatchev, M. Breton, C. Dalla Man, and C. Cobelli, "In silico preclinical trials: A proof of concept in closed-loop control of type 1 diabetes," *Journal of Diabetes Science and Technology*, vol. 3, no. 1, pp. 44–55, 2009.

[15] R. N. Bergman, Y. Z. Ider, C. R. Bowden, and C. Cobelli, "Quantitative estimation of insulin sensitivity," *American Journal of Physiology-Endocrinology and Metabolism*, vol. 236, no. 6, p. E667, 1979.

[16] C. Dalla Man, R. A. Rizza, and C. Cobelli, "Meal simulation model of the glucose-insulin system," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 10, pp. 1740–1749, 2007.

[17] D. E. Rivera, M. Morari, and S. Skogestad, "Internal model control: PID controller design," *Industrial & Engineering Chemistry Process Design and Development*, vol. 25, no. 1, pp. 252–265, 1986.

[18] S. J. Qin and T. A. Badgwell, "A survey of industrial model predictive control technology," *Control Engineering Practice*, vol. 11, no. 7, pp. 733–764, 2003.

[19] G. Pannocchia and J. B. Rawlings, "Disturbance models for offset-free model-predictive control," *AIChE Journal*, vol. 49, no. 2, pp. 426–437, 2003.

[20] R. Gondhalekar, E. Dassau, and F. J. Doyle, "Periodic zone-MPC with asymmetric costs for outpatient-ready safety of an artificial pancreas to treat type 1 diabetes," *Automatica*, vol. 71, pp. 237–246, 2016.

[21] S. A. Weinzimer, G. M. Steil, K. L. Swan, J. Dziura, N. Kurtz, and W. V. Tamborlane, "Fully automated closed-loop insulin delivery versus semiautomated hybrid control in pediatric patients with type 1 diabetes using an artificial pancreas," *Diabetes Care*, vol. 31, no. 5, pp. 934–939, 2008.

[22] C. V. Rao, J. B. Rawlings, and D. Q. Mayne, "Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations," *IEEE Transactions on Automatic Control*, vol. 48, no. 2, pp. 246–258, 2003.

[23] H. Lee and B. W. Bequette, "A closed-loop artificial pancreas based on model predictive control: Human-friendly identification and automatic meal disturbance rejection," *Biomedical Signal Processing and Control*, vol. 4, no. 4, pp. 347–354, 2009.

[24] K. J. Åström and T. Hägglund, "Automatic tuning of simple regulators with specifications on phase and amplitude margins," *Automatica*, vol. 20, no. 5, pp. 645–651, 1984.

[25] T. Battelino, T. Danne, R. M. Bergenstal, S. A. Amiel, R. Beck, T. Biester, E. Bosi, B. A. Buckingham, W. T. Cefalu, K. L. Close, C. Cobelli, E. Dassau, J. H. DeVries, K. C. Donaghue, K. Dovc, F. J. Doyle, III, S. Garg, G. Grunberger, S. Heller, L. Heinemann, I. B. Hirsch, R. Hovorka, W. Jia, O. Kordonouri, B. Kovatchev, A. Kowalski, L. Laffel, B. Levine, A. Mayorov, C. Mathieu, H. R. Murphy, R. Nimri, K. Nørgaard, C. G. Parkin, E. Renard, D. Rodbard, B. Saboo, D. Schatz, K. Stoner, T. Urakami, S. A. Weinzimer, and M. Phillip, "Clinical targets for continuous glucose monitoring data interpretation: Recommendations from the international consensus on time in range," *Diabetes Care*, vol. 42, no. 8, pp. 1593–1603, 2019.

[26] R. W. Beck, R. M. Bergenstal, T. D. Riddlesworth, C. Kollman, Z. Li, A. S. Brown, and K. L. Close, "Validation of time in range as an outcome measure for diabetes clinical trials," *Diabetes Care*, vol. 42, no. 3, pp. 400–405, 2019.

[27] B. P. Kovatchev, D. J. Cox, L. A. Gonder-Frederick, and W. Clarke, "Symmetrization of the blood glucose measurement scale and its applications," *Diabetes Care*, vol. 20, no. 11, pp. 1655–1658, 1997.

[28] U.S. Food and Drug Administration, "Integrated continuous glucose monitoring system," Code of Federal Regulations, Title 21, Section 862.1355, revised as of April 1, 2024, U.S. Government Publishing Office, govinfo.gov, 2024.

[29] G. Cappon, M. Vettoretti, G. Sparacino, and A. Facchinetti, "Continuous glucose monitoring sensors for diabetes management: A review of technologies and applications," *Diabetes & Metabolism Journal*, vol. 43, no. 4, pp. 383–397, 2019.

[30] U.S. Food and Drug Administration, "Alternate controller enabled infusion pump," Code of Federal Regulations, Title 21, Section 880.5730, revised as of April 1, 2024, U.S. Government Publishing Office, govinfo.gov, 2024.

[31] C. K. Boughton and R. Hovorka, "New closed-loop insulin systems," *Diabetologia*, vol. 64, no. 5, pp. 1007–1015, 2021.

[32] U.S. Food and Drug Administration, "Interoperable automated glycemic controller," Code of Federal Regulations, Title 21, Section 862.1356, revised as of April 1, 2024, U.S. Government Publishing Office, govinfo.gov, 2024.

2.14 Exercises

2.1 Concept check · 2 min

A glucose loop is built on the insulin pump of this chapter. In the convention of Section 1.3, which controller action does the loop need, and why?

(a) Reverse acting, $a = +1$, because the pump must rise when the glucose rises above setpoint

(b) Reverse acting, $a = +1$, because the process gain from the pump to the glucose is positive

(c) Direct acting, $a = -1$, because insulin lowers glucose and the process gain is negative

(d) Either action, because a negative proportional gain entered in the app reverses the loop

2.2 Concept check · 1 min

In the noise-free closed-loop day, exercise raised to 60 % at 20 h drives glucose from 74.9 to 47.8 mg/dL within 0.04 h. The next execution sets the pump to zero, where it stays for 0.84 h. Which change to the controller would shorten the low from that moment on?

(a) A larger proportional gain, which cuts the pump harder for each mg/dL of fall

(b) A shorter integral time, which empties the stored integral faster at the floor

(c) Derivative action, which answers the rate of the fall before its size

(d) None of them, because the pump is at zero and the insulin stored clears at plant rates

2.3 Concept check · 2 min

Table 2.6 repeats the closed-loop day at $K_c = 0.3$ for integral times from 6 to 0.4 h. What does it show between $\tau_I = 6$ h and $\tau_I = 1.2$ h, a fivefold change in integral gain?

(a) The integrated error rises by under 9 % while the nadir falls 6.4 mg/dL and crosses the low flag

(b) The integrated error has a minimum near the shipped 3 h, which is why that value is shipped

(c) The meal peak falls by several mg/dL as the integral gets faster, at the price of pump travel

(d) The nadir rises as the integral gets faster, because glucose returns to its setpoint sooner

2.4 Theory and calculation · 6 min

Use Equation (2.4) with $g_b = 291\,\mathrm{mg/dL}$, $p_1 = 0.0317/\mathrm{min}$, $f = 0.8$, $v_g = 12\,\mathrm{L}$, $s_i = 0.029$, $k_e = 0.09/\mathrm{min}$ and the fasting appearance $d = 1000$.

(a) Find the fasting glucose at infusion rates of 3.0 and 4.0 µU/min, and the secant gain between them. *Checkpoint:* 76.22 and 57.67 mg/dL, and −18.55 mg/dL per µU/min.

(b) Find the tangent gain at each of the two rates and the ratio between them. *Checkpoint:* −24.5 and −14.0 in the same units, a ratio of 1.75.

(c) Find from Equation (2.5) the pump rate that holds 80 mg/dL at $d = 1000$ and after a 30 % basal drift, $d = 1300$. *Checkpoint:* 2.85 and 3.63 µU/min.

2.5 Theory and calculation · 6 min

The fasting step test of Section 2.4.3 moves the pump from 3.0 to 4.0 µU/min. The glucose response crosses 28.3 % and 63.2 % of its change at 0.553 and 1.340 h after the step, and the process gain is $K_p = -18.55$ mg/dL per µU/min.

(a) Fit a first-order plus dead-time model by the two-point rule $\tau_p = 1.5\,(t_{63} - t_{28})$ and $\theta_p = t_{63} - \tau_p$. *Checkpoint:* $\tau_p = 1.18$ h and $\theta_p = 0.16$ h.

(b) Compare both, in minutes, with $1/k_e$ and $1/p_2$ for $k_e = 0.09$/min and $p_2 = 0.0123$/min, and find the time constant of the glucose balance, $1/(p_1 + s_i\bar{x})$, at $\bar{x} = 33.33$. *Checkpoint:* 9.6 against 11.1 min, 70.8 against 81.3 min, and 1.0 min.

(c) Find the internal model control gain $K_c = \tau_p/(|K_p|(\tau_c + \theta_p))$ for $\tau_c = \tau_p$ and for $\tau_c = \theta_p$. *Checkpoint:* 0.048 and 0.20.

2.6 Theory and calculation · 8 min

The loop holds the reference setpoint of 76.22 mg/dL in Auto with $K_c = 0.3$, integral action off, and the bias at the 3.00 µU/min that holds that glucose with $d = 1000$. The basal drift then steps to 30 %, so $d = 1300$. Use the constants of the first theory exercise.

(a) Find the glucose at which the drift would settle with the pump held at 3.00 µU/min, and the pump rate that an integral would settle at to return glucose to 76.22 mg/dL. *Checkpoint:* 96.25 mg/dL and 3.82 µU/min.

(b) The proportional law is $u_i = 3.00 + 0.3\,(g - 76.22)$. Solve it together with Equation (2.4) for the equilibrium glucose, pump rate and offset. *Checkpoint:* 78.55 mg/dL, 3.70 µU/min and 2.33 mg/dL.

(c) Find the share of the open-loop rise of (a) that the proportional loop removes. *Checkpoint:* 88 %.

2.7 App challenge · 8 min

Load scenario 2, Open-loop actuator step, and change no setting. It starts in Manual at 80 mg/dL with the pump at 2.85 µU/min, and the pump steps by 2.0 µU/min at 2 h, in effect from the execution at 2.016 h. Use the default seed, run to at least 14 h at speed 5 and export the CSV file.

(a) Read the true glucose, column `true_pv`, in the row at 13.944 h. Compute the secant gain of the step and compare it with the tangent gain of −27.0 mg/dL per µU/min at the starting rate. *Checkpoint:* 47.77 mg/dL within 0.05, a secant gain of −16.1, 60 % of the tangent gain.

(b) Find the time after 2.016 h at which `true_pv` has covered 63.2 % of the change from 80 to the value of (a), interpolating between rows, and compare it with the 1.34 h of the fasting one-unit step. *Checkpoint:* 1.17 h, within 0.03 h.

2.8 App challenge · 8 min

Load scenario 1, Reference operating point. It starts in Manual with the pump at 3.0 µU/min, the glucose and its setpoint at 76.22 mg/dL and no sensor noise. Before Run, set the mode to Auto. Run at speed 1, pause between 1.5 and 2.5 h, set Basal drift to 30 %, run on past 14.2 h and export the CSV file.

(a) Read `true_pv` and `mv` in the row at 14.112 h, and the IAE tile of the scorecard. *Checkpoint:* 76.25 mg/dL, 3.82 µU/min and 11.8 mg/dL h, within 0.01, 0.01 and 0.1.

(b) Press the reload button (not Reset), set Integral time to its top, 48 h, which switches integral action off, set the mode to Auto and repeat the run. Read the same row. *Checkpoint:* 78.55 mg/dL and 3.70 µU/min, within 0.01.

Compare both rows with Equation (2.5) and with the proportional-only equilibrium of Equation (2.4), and say why the reload button is needed in (b).

2.9 App challenge · 8 min

Load scenario 9, MPC with an announced meal. It starts at 80 mg/dL in Auto under the MPC with Meals announced on, and a meal of amplitude 1800 arrives at 2 h. Use the default seed and run each case to 24 h at speed 10. For each run read the IAE tile of the scorecard, and find in the exported file the highest and lowest `true_pv` from 2 h on and the time below 70 mg/dL (the number of rows below 70 times 0.168 h).

(a) Run the scenario as shipped. *Checkpoint:* 50.9 mg/dL h, 98.6 and 75.0 mg/dL, 0 h.

(b) Set Meals announced to Off, press Reset and run again. *Checkpoint:* 99.0 mg/dL h, 130.9 and 64.9 mg/dL, 0.84 h.

(c) Set the Controller to PID, press Reset and run again. *Checkpoint:* 83.7 mg/dL h, 116.8 and 66.2 mg/dL, 0.67 h.

Accept an IAE within 0.1 and glucose within 0.1 mg/dL. Explain the order of the three runs.

2.10 Application area · 4 min

A real continuous glucose monitor reads interstitial glucose 5 to 15 min behind plasma, and the model has no such lag. Take the identified model of Section 2.4.3, $K_p = -18.55\,\mathrm{mg/dL}$ per µU/min, $\tau_p = 1.18\,\mathrm{h}$ and $\theta_p = 0.16\,\mathrm{h}$, and treat a 15 min sensor lag as added dead time.

(a) Find the new dead time and the internal model control gain for $\tau_c = \theta_p$. *Checkpoint:* 0.41 h and 0.078.

(b) Compare it with the shipped $K_c = 0.3$, and say what the Limitations section allows an engineer to take from the app to a device. *Checkpoint:* The shipped gain is 3.9 times larger.

2.11 Application area · 3 min

(a) The six-month trial cited in Section 2.2 reported time in the 70 mg/dL to 180 mg/dL range of 71 % under closed loop against 59 % in the control group. Express both in hours per day and give the difference. *Checkpoint:* 17.0 and 14.2 h, 2.9 h a day.

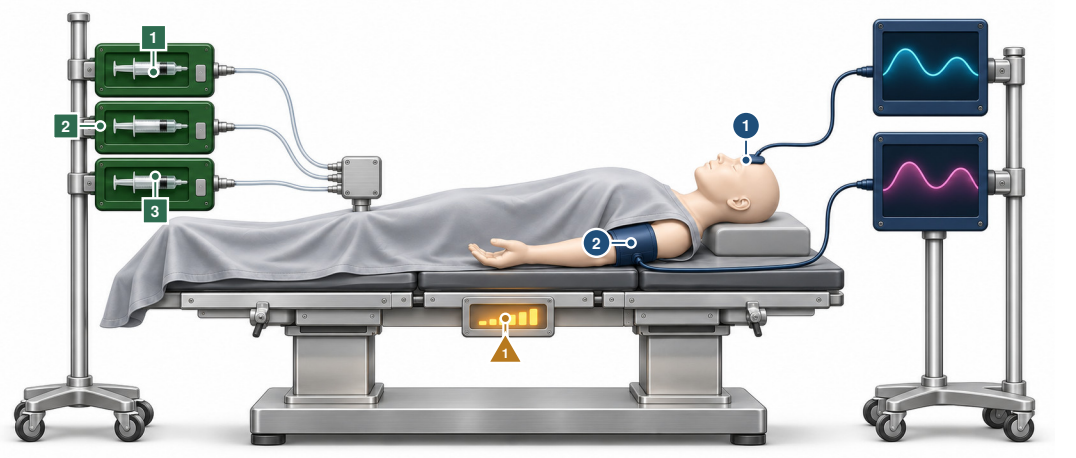
(b) The noise-free closed-loop day of Table 2.5 spends 94.4 % of its samples in the same range, none above 180 mg/dL and none below 64 mg/dL. Find the time it spends below 70 mg/dL and explain why the teaching band of 64 to 104 mg/dL reads 100 % for the same run. *Checkpoint:* 1.34 h.

2.12 Review the agent · 5 min

An AI assistant is asked to retune the glucose loop and replies as follows. “The step test gives $K_p = -18.55\,\mathrm{mg/dL}$ per µU/min, $\tau_p = 1.18\,\mathrm{h}$ and $\theta_p = 0.16\,\mathrm{h}$. The internal model control rule with $\tau_c = \theta_p$ gives $K_c = 0.198$ on the slider grid and $\tau_I = \tau_p = 1.18\,\mathrm{h}$. That is a third less proportional gain than the shipped 0.3, so the loop is more conservative and will lower the risk of low glucose after meals.” Find the flaw in the proposal and say how it could be checked.

CHAPTER 3

Closed-Loop Anesthesia Simulator for Model Inversion with Feedback Correction



The controller moves 1 propofol pump, the drug that brings sleep 2 remifentanil pump, an opioid against pain 3 vasopressor pump for blood pressure

The controller reads 1 depth monitor on the forehead 2 blood pressure on the arm

What disturbs it 1 surgical stimulation, up at the incision and down at the closing

Notice that the two drug pumps work from their own estimates of the drug in the brain, and only the forehead monitor shows how deep the sleep really is. Illustration, not to scale.

IN PLAIN TERMS **A Pump That Trusts Its Own Arithmetic**

The main idea. During an operation, two drugs given into a vein keep the patient asleep and free of pain. A pump works out how much drug has reached the brain, and a brain-wave monitor shows how deep the sleep is.

The challenge. The pump's arithmetic fits an average patient, but one patient can need three times the drug of another. The monitor can be fooled into showing a light sleep, and a pump can add drug but never take it back.

Why it matters. Too light a sleep risks memory of the operation, while too deep a sleep or too low a blood pressure does harm.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	The first page, the history boxes of Section 3.2, the fact boxes of Sections 3.1 and 3.6, the try-it boxes of Section 3.6, and the key points
Study	60 min	The whole chapter with its pause points and the twelve exercises
Practice	25 min	The first page, Section 3.6, Section 3.7, Section 3.8 and the open challenge

The control question An infusion pump holds the drug concentration its population model predicts, which is exact on the patient the model was built from and silently wrong on anybody else. What does an outer loop on a depth monitor add to that model inversion, and which structural choices matter more than its tuning?

Learning objectives

- Compute the rate constants of a three-compartment model from its covariates, and explain why the maintenance infusion falls for hours.
- Evaluate the response surface for a pair of effect-site concentrations, and find the process gain of the depth loop at an operating point.
- Explain why an inner loop closed on the pump's own model cannot detect a patient who differs from the population, and how an outer integral identifies that patient.
- Show by runs why signal validation, a pressure loop and the drug split are structural choices that no tuning replaces.
- Distinguish what a faster loop can and cannot change in the excursion at the end of a case.

Prerequisites Sections 1.4.2 (compartments and model inversion), 1.4.4 (variability between individuals), 1.4.6 (one-sided actuators) and 1.6.2 (velocity form). Section 1.6.3 is needed only for the predictive controller.

The app https://apmonitor.com/apps/anesthesia/

See also Volume 5, Chapter 10 and Volume 5, Chapter 4 for product quality inferred from a model and corrected by a late measurement. Volume 3, Chapter 2 for structure chosen before tuning, and Volume 3, Chapter 3 for how fast an inner loop must be.

3.1 Introduction

Every loop in an introductory course pairs a controlled variable with an instrument that measures it. Intravenous anesthesia has no such instrument. What decides whether a patient is unconscious is the concentration of the drug at its site of action in the brain, and there is no probe for it, no sample to draw and no prospect of either during an operation. The loop is closed regardless. A target-controlled infusion pump carries a published population pharmacokinetic model, feeds that model the infusion it is delivering, and controls the concentration the model predicts [1, 2]. The inner loop that results is exact, immediate and completely blind, because its plant is an equation rather than a patient.

DID YOU KNOW? Sixty Thousand Pumps

By the count of a review published in 2023, about 60 000 target-controlled infusion pumps had been sold in more than 90 countries, and they give intravenous anesthesia to millions of patients each year [3].

The misconception this plant corrects is that an exact loop is a correct loop. With the pump commanded to hold 1.7 µg/mL of propofol and 2 ng/mL of remifentanil, the modeled concentration follows the same trajectory on every patient and sits within 0.01 µg/mL of its target at forty minutes, while the depth index those concentrations actually produce settles anywhere between 52.1 and 70.4 as the patient's propofol requirement moves across the population range. The requirement itself is what varies, by a factor of three between a sensitive and a resistant adult, and no measurement available before induction reveals which one is on the table [4, 5]. Feedback from a real measurement is the only mechanism that finds the difference, and the simulator shows an integral finding it without anybody having told the controller that a parameter was missing.

The structure is general. A fast inner loop closed on a model-derived estimate, corrected by a slow outer loop closed on a real measurement, is the standard answer whenever the variable that matters cannot be measured quickly and a usable model of it exists, and it is the arrangement inferential control was introduced to formalize [6]. Distillation composition inferred from tray temperatures, reactor conversion inferred from a heat balance, a furnace outlet inferred from a firing model and any

modern soft sensor all have the same shape, and all of them fail in the same way when the model is fitted to a population and applied to an individual [7]. The milk powder dryer of Volume 5, Chapter 10 steers its product moisture by such a soft sensor and corrects its bias from a late laboratory sample. Anesthesia is the sharpest instance available because the inner measurement is not merely indirect, it is a calculation, so the inner loop cannot report an error even in principle.

A transfer function or a block diagram of a cascade draws two summing junctions and hides exactly that fact. However, a simulation can put the model concentration and the patient concentration on the same axis and let them separate, and it can move the plant rather than the output, corrupt the measurement in the direction that provokes an overdose, and shut the actuator off at the moment more actuator is not what is needed. This chapter works with a simulation of anesthetic depth in which the model-based inner loop and the feedback outer loop can be switched into service one at a time, so that what each half contributes is measured rather than asserted. It applies the compartment models, the open-loop error of model inversion and the one-sided actuator of Sections 1.4.2, 1.4.4 and 1.4.6 to a full-scale patient.

Section 3.2 traces how infusion pumps and depth monitors came to be combined. Section 3.3 gives the thirteen-state patient, with two independent copies of the pharmacokinetics, one for the patient and one inside the pump, down to every constant needed to rebuild it. Section 3.4 shows that the process gain of the depth loop rises from 8.8 to 18.8 index units per µg/mL as the opioid target moves from zero to 8 ng/mL, and why the maintenance rate never settles. Section 3.5 describes the four loops, the pump's model as a measurement and the optional predictive controller. Section 3.6 quantifies the open-loop error of model inversion against patient sensitivity, works through a thirteen-scenario sequence that separates model error, measurement error and actuator authority, ranks four structural choices by what each is worth on a two-hour case and compares the predictive controller with the closed loop. Section 3.7 describes the real monitor, pumps and controllers, and Sections 3.8 and 3.9 state the limitations and conclude. The pharmacokinetic models and the control algorithm are published ones.

3.2 Automation of Intravenous Anesthesia

Automatic control of anesthesia is older than most of the drugs it is now practiced with. A servo that read the amplitude of the electroencephalogram and moved the delivery of the anesthetic to hold it constant was built and used on patients in 1950, and it worked in the sense that the trace stayed where it was put [8]. What the arrangement lacked was a drug that could be titrated quickly enough to make the loop worth closing, and a measurement whose relationship to consciousness was understood well enough to choose a setpoint.

The problem that stayed was dosing, and it was solved by computing the infusion from a compartment model fitted to a population [1, 2, 9–11], as the first history box below tells. Section 1.4.2 derives the same inversion for two compartments.

> ⧖ HISTORY **From a Bolus to a Target (1992 to 1997)**
>
> An intravenous drug spreads into the tissues while it is being cleared, so the infusion that holds a steady concentration cannot be a steady rate. The first model for automated dosing gave a bolus sized from the target and the volume of distribution, followed by an infusion. Three-compartment models then described how the drug moves between the plasma, well-perfused tissue and fat, and an effect site joined to the plasma by a single rate constant let the target be set where the drug acts. A microprocessor in the pump now computes the rate from the model and the patient's covariates. The name target-controlled infusion appeared in 1992 and was agreed as the generic term in 1997. The first commercial systems entered practice in 1996 with a propofol model whose only covariate was weight [3].

Feedback returned when the measurement improved [12, 13], as the second history box tells, and the combination that this application implements is the one that box ends with, a bispectral index

controller sitting above a target-controlled infusion pump rather than driving the syringe directly. Later work replaced the tuning of that outer loop with a robust design based on the variability of the population rather than on a nominal patient, and evaluated it clinically [5]. Several hundred trials later the method is not standard of care anywhere, and the reasons are the ones this simulator reproduces. The measurement can be corrupted by electrocautery, by muscle activity and by warming devices, and it is corrupted upward, in the direction that makes a light patient of an artifact [14]. The actuator has authority in one direction only. The costs of the two error directions are not comparable, because awareness with recall occurs in roughly one general anesthetic in a thousand and is not undone when the trace recovers [15], while sustained electroencephalographic suppression on the other side is associated with delirium after surgery [16]. A regulator that scores the two directions with one symmetric error integral is answering the wrong question [17]. Equation (1.10) of Section 1.4.6 writes the asymmetric objective that such a loop needs.

> ⧖ HISTORY **Feedback Above the Pump (1987 to 2002)**
>
> In 1987 a team held methohexital anesthesia in 13 volunteers by feedback. The measured variable was the median frequency of the electroencephalogram, with a setpoint of 2 to 3 Hz, and the loop succeeded in 11 of them. The controller rested on a pharmacokinetic and pharmacodynamic model, and its authors saw feedback as a way to adapt population mean data to the individual [12]. A depth index computed from the electroencephalogram came into use in 1996 [14]. In 2002 a closed loop took that index as its controlled variable, a PID algorithm as its controller and a propofol target-controlled infusion system as its actuator. It gave adequate anesthesia in 9 of 10 patients, and in three of them the index oscillated around the setpoint [13].

3.3 Patient Model

The simulated patient is a 40 y old man of 70 kg and 170 cm, and every rate constant below is computed from those covariates rather than typed in. Two drugs are infused. Propofol is the hypnotic and is responsible for most of the depth of anesthesia and for essentially all of the loss of arterial pressure. Remifentanil is a short-acting opioid that contributes to depth mainly through its interaction with propofol and costs very little pressure. A third infusion, a vasopressor, exists only to defend the pressure. Figure 3.1 shows the compartments, the three infusions, the two effect sites, the response surface that produces the depth index, the arterial pressure state and the point at which the monitor is read.

3.3.1 What the Loops See

The controlled variable of the outer loop is the depth index as the monitor reports it, and the controlled variable of the pressure loop is the measured mean arterial pressure. The manipulated variables are the three infusion rates. Between them sits the propofol effect-site target, which is the output of the depth controller and the setpoint of the concentration controller, and which is the only place in the loop where a concentration is ever commanded. Table 3.1 lists the classification. The distinction between the true index and the reported index is load bearing throughout, because the controller acts on the reported one while every performance number in Section 3.6 is computed from the true one.

3.3.2 Pharmacokinetics

Propofol is described by the Schnider three-compartment model with an effect site hung behind the central compartment, as shown in Equation (3.1). The same four equations with the Minto coefficients describe remifentanil. They extend Equation (1.2) by a second peripheral compartment and are written in rate constants rather than clearances.

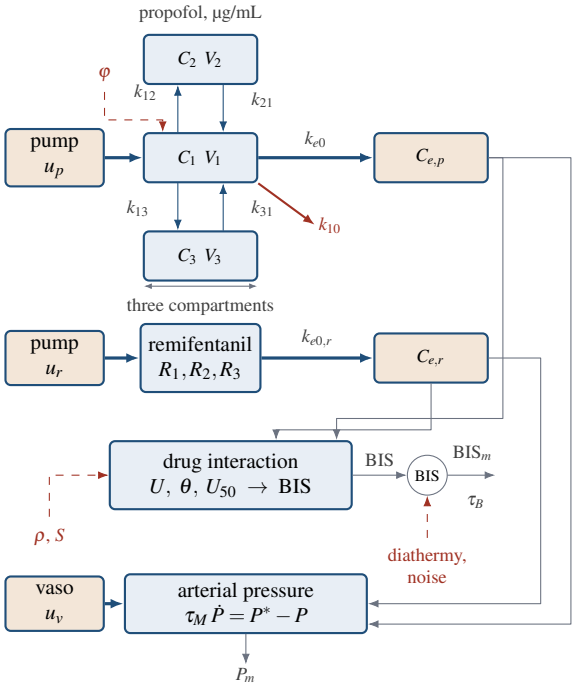


Figure 3.1 Plant of the anesthesia simulator. Propofol enters central compartment C_1 of volume V_1 from pump u_p, exchanges with peripheral compartments C_2 and C_3, is cleared at k_{10} and reaches effect site $C_{e,p}$ through k_{e0}. Remifentanil has the same structure and much faster constants. The two effect-site concentrations enter a response surface that returns the depth index BIS, which the monitor smooths with time constant τ_B and reports as BIS$_m$. Both concentrations, the surgical stimulation and the vasopressor effect site drive the arterial pressure state. Patient sensitivity ρ and stimulation S scale the concentrations at which the drugs are half maximally effective, blood loss φ shrinks V_1, and diathermy adds a positive offset to the monitor.

$$\frac{\mathrm{d}C_1}{\mathrm{d}t} = \frac{u_p W}{60\,\varphi V_1} - (k_{10} + k_{12} + k_{13})C_1 + k_{21}C_2 + k_{31}C_3 \tag{3.1a}$$

$$\frac{\mathrm{d}C_2}{\mathrm{d}t} = k_{12}C_1 - k_{21}C_2 \tag{3.1b}$$

$$\frac{\mathrm{d}C_3}{\mathrm{d}t} = k_{13}C_1 - k_{31}C_3 \tag{3.1c}$$

$$\frac{\mathrm{d}C_{e,p}}{\mathrm{d}t} = k_{e0}\,(C_1 - C_{e,p}) \tag{3.1d}$$

where C_1, C_2 and C_3 are the central, fast peripheral and slow peripheral concentrations in µg/mL, $C_{e,p}$ is the effect-site concentration, u_p is the commanded infusion in mg/(kg h), W is body weight in kg, V_1 is the central volume in L, φ is the central volume multiplier that represents blood loss, and the rate constants are in reciprocal minutes. The unit arithmetic of the infusion term is worth following once, because a factor of a thousand hides in it. Milligrams per kilogram per hour multiplied by kilograms is milligrams per hour, divided by sixty is milligrams per minute, and divided by a volume in liters is milligrams per liter, which is micrograms per milliliter. The remifentanil infusion u_r is in µg/(kg min), so its own term is $u_r W/(\varphi V_{1,r})$ with no factor of sixty, and micrograms per liter is

Table 3.1 Classification of Signals. Nominal Values Are Those of the Maintenance Start Used by Six of the Thirteen Scenarios.

Symbol	Role	Nominal	Range
BIS_m	CV, measured	50	0–100
P_m	CV, measured	79 mmHg	20–200 mmHg
$C_{e,\mathrm{SP}}$	outer MV, inner SP	1.70	0–8 µg/mL
u_p	MV, propofol	3.43	0–20 mg/(kg h)
u_r	MV, remifentanil	0.149	0–1.0 µg/(kg min)
u_v	MV, vasopressor	0	0–1.0 µg/(kg min)
ρ	DV, unmeasurable	1.0	0.5–2.0
ρ_r	DV, unmeasurable	1.0	0.6–1.6
S	DV, observed	0	0–1
φ	DV, unmeasured	1.0	0.6–1.1
σ	DV, monitor noise	1.0	0–6 index
ΔP_0	DV, baseline offset	0	−20 to +20 mmHg
C_1, C_2, C_3	states, propofol	1.70, 9.55, 13.6	≥ 0 µg/mL
$C_{e,p}$	state, effect site	1.70	≥ 0 µg/mL
R_1, R_2, R_3	states, opioid	4.00, 7.69, 4.20	≥ 0 ng/mL
$C_{e,r}$	state, effect site	4.00	≥ 0 ng/mL
P	state, pressure	79 mmHg	20–200 mmHg
$C_{e,v}$	state, vasopressor	0	0–1.2
x_p, x_r, x_v	states, pumps	3.43, 0.149, 0	at command

nanograms per milliliter. Blood loss enters only through φ, which concentrates the same delivered milligram into a smaller volume.

The Schnider volumes and clearances are functions of the covariates, as shown in Equation (3.2), with lean body mass from the James formula [10].

$$V_1 = 4.27\,\mathrm{L}, \quad V_2 = 18.9 - 0.391\,(A-53), \quad V_3 = 238\,\mathrm{L} \tag{3.2a}$$

$$\mathrm{Cl}_1 = 1.89 + 0.0456\,(W-77) - 0.0681\,(L-59) + 0.0264\,(H-177) \tag{3.2b}$$

$$\mathrm{Cl}_2 = 1.29 - 0.024\,(A-53), \quad \mathrm{Cl}_3 = 0.836 \tag{3.2c}$$

$$L = 1.1\,W - 128\,(W/H)^2 \quad \text{(male)} \tag{3.2d}$$

where A is age in years, H is height in cm, L lean body mass in kg, volumes are in L and clearances in L/min. The rate constants follow as $k_{10} = \mathrm{Cl}_1/V_1$, $k_{12} = \mathrm{Cl}_2/V_1$, $k_{13} = \mathrm{Cl}_3/V_1$, $k_{21} = \mathrm{Cl}_2/V_2$ and $k_{31} = \mathrm{Cl}_3/V_3$, with $k_{e0} = 0.456/\mathrm{min}$ fixed. The female lean body mass expression $1.07W - 148(W/H)^2$ is implemented and is reached through a hidden parameter. The remifentanil parameters are $V_1 = 5.1 - 0.0201(A-40) + 0.072(L-55)$, $V_2 = 9.82 - 0.0811(A-40) + 0.108(L-55)$, $V_3 = 5.42\,\mathrm{L}$, $\mathrm{Cl}_1 = 2.6 - 0.0162(A-40) + 0.0191(L-55)$, $\mathrm{Cl}_2 = 2.05 - 0.0301(A-40)$, $\mathrm{Cl}_3 = 0.076 - 0.00113(A-40)$ and $k_{e0,r} = 0.595 - 0.007(A-40)$ [11]. The two remifentanil volumes, and every volume that appears in a rate-constant denominator, are floored at 1 L so that an extreme covariate set cannot divide by zero. The resulting constants for the default patient are listed in Table 3.2.

The ratio $k_{13}/k_{31} = 55.7$ is the reason this plant has no steady state on the time scale of an operation. At equilibrium the slow compartment holds 55.7 times the central concentration, and it fills with a time constant of $1/k_{31} = 285\,\mathrm{min}$, so an anesthetic of ordinary length is a transient from beginning to end.

Table 3.2 Rate Constants Computed for the Default Patient, a 40 y Old Male of 70 kg and 170 cm With Lean Body Mass 55.3 kg. These Are Derived Quantities and Are Not Displayed by the Application.

Constant	Propofol	Remifentanil	Unit
V_1	4.27	5.12	L
V_2	23.98	9.85	L
V_3	238	5.42	L
k_{10}	0.3836	0.5088	1/min
k_{12}	0.3752	0.4003	1/min
k_{13}	0.1958	0.0148	1/min
k_{21}	0.0668	0.2081	1/min
k_{31}	0.00351	0.0140	1/min
k_{e0}	0.456	0.595	1/min

3.3.3 Response Surface and the Depth Index

The two drugs act on one endpoint and they are synergistic, so the index is computed from a response surface rather than from two independent contributions [18]. Each concentration is normalized by its own half-maximal concentration, the two normalized quantities are added, and the sum is divided by a normalizing term that depends on how the mixture is made up, as shown in Equation (3.3).

$$U_p = \frac{C_{e,p}}{C_{50,p}\,\rho\,(1+k_s S)}, \qquad U_r = \frac{C_{e,r}}{C_{50,r}\,\rho_r\,(1+k_s S)} \tag{3.3a}$$

$$\theta = \frac{U_p}{U_p + U_r}, \qquad U_{50} = 1 - \beta\,\theta\,(1-\theta) \tag{3.3b}$$

$$U = \frac{U_p + U_r}{\max(0.05,\ U_{50})} \tag{3.3c}$$

$$\mathrm{BIS} = E_0 - E_{\max}\,\frac{U^\gamma}{1+U^\gamma} \tag{3.3d}$$

where $C_{50,p}$ and $C_{50,r}$ are the population half-maximal concentrations, ρ and ρ_r are the patient's sensitivities relative to the population, S is surgical stimulation between zero and one, k_s is the stimulation gain, β is the synergy coefficient, γ is the Hill slope, E_0 is the index of an awake patient and $E_{\max}$ is the largest achievable depression. When either drug is absent θ is zero or one and U_{50} is unity, which is the additive case. At an even split $U_{50} = 0.65$ with the default $\beta = 1.40$, so the pair reaches an index of 50 on 35 % less total normalized drug than addition would predict. The index is returned as E_0 when both concentrations are zero.

Equation (3.3a) is where surgical stimulation acts, and the way it acts is the reason this disturbance behaves unlike a load. Stimulation does not add to the index. It multiplies the concentration the patient needs, so the operating point and the process gain move at the same instant. The incision event sets $S = 0.85$, the level the scenarios treat as full stimulation. At that level the effective propofol C_{50} of the default patient rises from 4.16 to 7.17 µg/mL, a factor of 1.72, and on the resistant patient of scenarios 3 and 4 it rises from 6.66 to 11.46 µg/mL.

3.3.4 Arterial Pressure

Arterial pressure is one first-order state driven by the two drug concentrations, the stimulation and the vasopressor, as shown in Equation (3.4).

$$P^* = P_0 - a\,C_{e,p} - b\,C_{e,r} + \Delta P_0 + c\,S\left[1 - 0.6\min\left(1,\ \frac{C_{e,r}}{\max(1, C_{50,r})}\right)\right] + 100\,d\,C_{e,v} \tag{3.4a}$$

$$\tau_M \frac{\mathrm{d}P}{\mathrm{d}t} = P^* - P, \qquad \tau_v \frac{\mathrm{d}C_{e,v}}{\mathrm{d}t} = u_v - C_{e,v} \tag{3.4b}$$

where P_0 is the awake pressure, a and b are the pressure lost per unit of propofol and of opioid, c is the pressure rise at $S = 1$, d is the vasopressor potency, $C_{e,v}$ is the vasopressor effect site and ΔP_0 is a baseline offset that represents a patient who starts low. The bracket in (3.4a) is the opioid blunting the pressor response to surgery, which is most of what the opioid is given for in practice. The coefficients $a = 6.4\,\mathrm{mmHg}$ per µg/mL and $b = 0.55\,\mathrm{mmHg}$ per ng/mL carry the whole argument about how to split the work between the two drugs, because holding an index of 50 on propofol alone requires 3.74 µg/mL against 1.69 µg/mL with the opioid at 4 ng/mL, and the difference is 11 mmHg taken off the patient.

3.3.5 Actuators and Instruments

All three pumps are first-order lags on the commanded rate with a common time constant, as shown in Equation (3.5), and the state is clipped after every step to the mechanical range of the device.

$$\tau_p \frac{\mathrm{d}x_i}{\mathrm{d}t} = u_{i,\mathrm{cmd}} - x_i, \qquad i \in \{p,\ r,\ v\} \tag{3.5}$$

The depth monitor is a first-order filter on the true index with additive noise and an artifact term, as shown in Equation (3.6).

$$y_k = \mathrm{BIS}_{k-1} + \sigma\,n_k + \delta_k\,(34 + 6\,n_k) \tag{3.6a}$$

$$\hat{y}_k = \alpha_k\,\hat{y}_{k-1} + (1 - \alpha_k)\,\mathrm{clamp}(y_k, 0, 100) \tag{3.6b}$$

$$\mathrm{BIS}_{m,k} = \begin{cases} \mathrm{BIS}_{m,k-1}, & v\,\delta_k = 1 \\ \hat{y}_k, & \text{else} \end{cases} \tag{3.6c}$$

where n_k is a first-order-filtered Gaussian sequence rescaled to a stationary standard deviation of one and advanced once per controller execution while the noise or the artifact is on, so the artifact's fluctuation is the same sequence scaled by six, δ_k is one while diathermy is in use, v is one when signal-quality validation is selected, σ is the noise setting in index units, Δt_c is the controller period and $\alpha_k = \exp(-\Delta t_c / \tau_B)$ with τ_B the monitor smoothing, or one while $v\,\delta_k = 1$, so the hold freezes the smoothing as well. The index the monitor filters is the one computed on the previous step, because everything downstream of the electrodes is at least one sample behind the patient. The artifact is a positive offset of 34 index units, which is the direction a real electromyographic or radio-frequency artifact takes the index [14], and the simulated signal quality indicator falls from 96 to 24 while it is present. The measured pressure is the pressure state through a first-order filter of time constant $\tau_{M,m}$ with no noise.

3.3.6 Parameters, Starting States and Numerics

Table 3.3 collects the constants. Four of them are hidden and never appear on screen. Every other entry is both displayed and adjustable over the range shown, and the derived rate constants of Table 3.2 are neither, because they follow from the covariates.

Two starting conditions are offered. An awake start puts every concentration at zero and the pressure at $P_0 + \Delta P_0$. A maintenance start puts the central compartment and the effect site at 1.7 µg/mL,

Table 3.3 Model Parameters. Adjustable Constants Appear in the Model Panel With the Ranges Shown. Hidden Constants Are Never Displayed.

Symbol	Meaning	Value	Unit	Range
A	age	40	y	18–90
W	weight	70	kg	40–140
H	height	170	cm	145–200
$C_{50,p}$	propofol C_{50}	4.16	µg/mL	1.5–9
$C_{50,r}$	opioid C_{50}	19.0	ng/mL	6–40
γ	Hill slope	1.43	—	0.8–4
β	synergy	1.40	—	0–3.6
E_0	awake index	93	—	85–100
k_s	stimulation gain	0.85	—	0–2
P_0	awake pressure	92	mmHg	60–130
a	propofol vasodilation	6.4	mmHg per µg/mL	0–15
b	opioid vasodilation	0.55	mmHg per ng/mL	0–2
c	pressor response	26	mmHg	0–60
d	vasopressor potency	0.42	mmHg per %	0.05–1.5
τ_M	circulatory lag	0.8	min	0.1–5
τ_v	vasopressor lag	1.6	min	0.2–8
τ_B	monitor smoothing	0.42	min	0.05–2
$\tau_{M,m}$	pressure line lag	0.15	min	0.02–1
τ_p	pump lag	0.12	min	0.02–1
BIS_{lo}	band bottom	40	—	25–55
BIS_{hi}	band top	60	—	45–75
BIS_s	suppression limit	30	—	15–40
P_{lo}	perfusion limit	60	mmHg	45–80
—	sex flag	1	—	hidden
$E_{\max}$	largest depression	93	—	hidden
—	propofol price	0.012	$/mg	hidden
—	remifentanil price	0.34	$/mg	hidden

the fast compartment at $C_1 k_{12}/k_{21} = 9.55\,\mu\mathrm{g/mL}$, the slow compartment at $8C_1 = 13.6\,\mu\mathrm{g/mL}$ and the pressure at the value of P^* in Equation (3.4) with no vasopressor on board, which is $P_0 - aC_1 - bR_1$ when the case starts unstimulated and without offset, with the opioid states set the same way and its slow compartment at $1.05R_1$. That state is not an equilibrium and cannot be, since the slow compartment starts at 8 times the central concentration against 55.7 times at equilibrium, so it is about a seventh full. What it is instead is about an hour into a case, which is what an anesthetist means by steady. One scenario overrides the start to $C_{e,p} = 4.3\,\mu\mathrm{g/mL}$ under full stimulation, which is the concentration a stimulated patient needs, with the concentration target at the same value, so that case starts on the setpoint at a pressure of 81.6 mmHg.

The engine of Section 1.6.1 integrates the model by the fourth-order Runge-Kutta method at a fixed step of 0.01 min (0.6 s). The controllers execute every 0.05 min (3 s), which is a typical update period for a published closed-loop system, and the historian logs twenty samples per simulated minute. The noise generator is seeded at 20260915. After every step the eight concentrations and the vasopressor effect site are floored at zero, the pressure is clipped to 20–200 mmHg, and the three pump states are clipped to their mechanical ranges of 40 mg/(kg h), 2.5 µg/(kg min) and 1.2 µg/(kg min), all of which are wider than the limits the controllers are allowed to command.

3.3.7 Assumptions

The model carries detail where the control problem is and very little elsewhere. Each simplification below breaks down at a stated place.

The pharmacokinetics are population means. Schnider and Minto were fitted to healthy adult volunteers and are known to describe the extremes of age and body habitus poorly [10, 11]. The application inherits that and then makes the residual variability a slider, because the variability rather than the mean is the control problem.

One hypnotic endpoint. The response surface returns a single index. Real anesthesia has at least three endpoints, hypnosis, analgesia and immobility, and the two drugs trade differently against each of them [18]. A controller tuned on one endpoint in this model would have to be re-examined against the others.

The index is algebraic in the concentrations. Equation (3.3d) is instantaneous and the only dynamics are the monitor filter. A real processed-EEG index is a proprietary computation over an epoch with dynamics of its own, it varies between patients at the same concentration, and its relationship to awareness is statistical rather than a threshold [14].

Stimulation is one number. A single quantity between zero and one scales both half-maximal concentrations. Real stimulation is a sequence of discrete events of very different intensities, it differs between the two drugs, and it depends on what is being done at that second.

The circulation is one state. There is no cardiac output, no heart rate, no fluid status, no baroreflex and no arrhythmia, and blood loss appears only as a change in one compartment volume. The application therefore cannot show the cases where a pressure problem is not a vasodilation problem.

The artifact is a step. Diathermy adds a fixed offset while it is on and the signal quality indicator switches between two values. A real artifact varies with electrode position and cautery mode, and a real quality index degrades gradually, which makes validation logic a threshold problem rather than the switch it is here.

Recovery is outside the model. Awakening, respiratory depression and postoperative pain are absent, which is precisely where the cost of a heavy opioid technique appears. The model can therefore show what moving depth onto the opioid buys and cannot show what it spends.

3.4 Dynamic Analysis

3.4.1 Process Gain of the Depth Loop

The gain the outer loop sees is the slope of the response surface at the operating point, and it moves for two independent reasons. Differentiating Equation (3.3d) numerically at the concentration pair that holds an index of 50 gives -8.8 index units per µg/mL with no opioid, -10.0 at 4 ng/mL and -18.8 at 8 ng/mL. The opioid target is an operator setting rather than a disturbance, so a decision taken for a reason that has nothing to do with tuning changes the loop gain by a factor of 2.1. Patient sensitivity moves the same gain again. The propofol concentration that holds an index of 50 at 4 ng/mL of opioid is 0.85 µg/mL at $\rho = 0.5$ and 3.38 µg/mL at $\rho = 2.0$, and over the narrower range $0.6 \leq \rho \leq 1.8$ the requirement still varies by a factor of three. Equation (1.6) gives the open-loop error that such a spread leaves under model inversion alone. A fixed-gain outer loop has to be stable across all of it, which is the argument for tuning the loop against the population rather than against a nominal patient [5].

3.4.2 Synergy and the Choice of Split

The set of concentration pairs that produce a given index is the isobologram, and its curvature is what makes the split between the two drugs a decision rather than a habit. Figure 3.2 draws the curve for an index of 50 on the population patient, on a resistant patient and under full surgical stimulation, against the straight line that additive drugs would give.

Two features of Figure 3.2 matter for control. The curve bows well inside the additive line, so at 8 ng/mL the propofol requirement is 0.84 µg/mL against the 1.99 µg/mL that addition predicts, a factor of 2.4. The two ways the curve moves are not the same shape. Patient sensitivity scales $C_{50,p}$ alone in (3.3a), so the whole curve stretches along the propofol axis by the factor ρ and by nothing else. Stimulation scales both half-maximal concentrations, so the curve moves outward along both

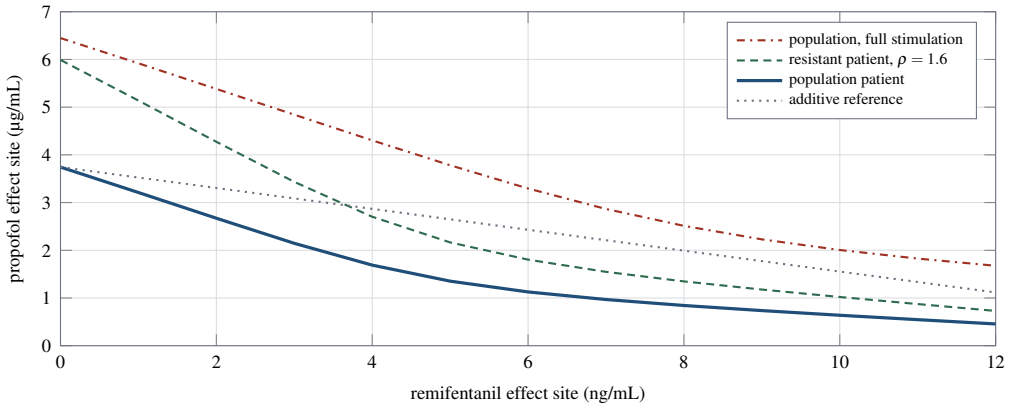


Figure 3.2 Isobologram of concentration pairs holding a depth index of 50, computed from the implemented response surface. The population patient is the thick solid curve, the resistant patient densely dashed and full stimulation dash-dotted. The dotted line joins the population curve at zero opioid to the additive prediction, and the bow of the solid curve below it is the synergy. Moving the opioid from zero to 4 ng/mL cuts the propofol requirement from 3.74 to 1.69 µg/mL, and every unit of that reduction is worth 6.4 mmHg of arterial pressure.

axes by $1 + k_s S$ together and the same opioid concentration buys proportionally less depth than it did before the incision.

🕮 **Pause.** Try Exercise 3.2 before reading on.

3.4.3 Why the Maintenance Rate Never Settles

Holding the effect site at 1.7 µg/mL from the maintenance start requires 3.43 mg/(kg h) at the beginning of the case. With the target-controlled pump in service and nothing disturbed at all, the pump delivers that rate at one minute, and the rate falls to 3.23 mg/(kg h) at one hour and to 2.84 mg/(kg h) at four hours, heading for the 2.39 mg/(kg h) that clearance alone would require once the slow compartment is full. The decay is the third compartment filling, the three-compartment form of the falling rate of Equation (1.3) and Figure 1.3, and it is the clearest illustration in the application of what a model buys. A controller with no model has to discover a requirement that falls by 17 % over four hours through its integral, continuously, for the whole case. A controller with a model gets it for free and spends its integral on the patient instead. With the depth loop closed over the same hour, the rate wanders between 2.7 and 3.8 mg/(kg h), because the outer loop passes the monitor noise through to the target.

3.4.4 Lags and the Speed of the Outer Loop

The propofol effect site is behind the plasma by $1/k_{e0} = 2.19$ min, the monitor adds $\tau_B = 0.42$ min of smoothing, and the pump adds $\tau_p = 0.12$ min. The outer loop therefore faces about two and a half minutes of lag that no tuning removes, and it also faces an inner loop that is fast enough to be treated as instantaneous because that loop works on an equation. The fermenter of Volume 3, Chapter 3 measures what an inner loop that is not fast enough costs a cascade. The default depth integral time of 3.5 min is a little longer than the total lag, which is why identification of an unusual patient takes half an hour rather than a few minutes. Anything that feels responsive on this loop is too fast.

3.4.5 One-Sided Authority

The propofol pump can add drug and cannot remove it. Shutting it off leaves whatever is already in the effect site to leave by redistribution and metabolism, and the effect site can only follow the

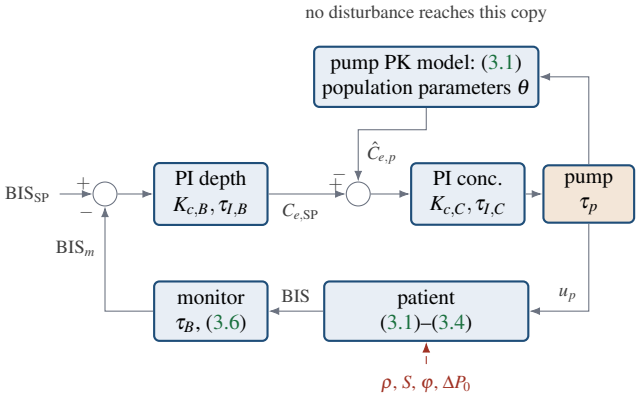


Figure 3.3 Two nested loops with one real measurement. The inner loop is closed on the pump's own integration of Equation (3.1) with population parameters. No disturbance reaches that copy, so it holds its target whatever the patient does, and everything the model gets wrong appears in the outer loop's error where an integral can absorb it.

central compartment down at k_{e0}. The deep excursions in Section 3.6 are of two kinds. In scenario 7 the propofol pump has just been commanded to zero when the index reaches its lowest value, so a different choice of gains trims the excursion only at its margins. At the end of a case the index reaches its lowest value within half a minute of the stimulation going, while the pump is still running, and in that time no change of infusion can reach an effect site that lags the plasma by 2.19 min. Stopping the propofol pump outright at the instant of closing changes the lowest index by less than 0.1 in both scenarios that end a case. A faster response after the closing, from a higher depth gain, a stopped pump or a lower opioid target, ends the excursion sooner while leaving its depth nearly unchanged. What makes it shallower is the moment at which the operator starts lightening, which is decided before the excursion rather than in response to it. Section 1.4.6 gives the time such a plant takes to fall to a limit, and the insulin pump of Chapter 2 meets the same asymmetry.

3.5 Control Structure as Implemented

3.5.1 Four Loops and Four Structures

Figure 3.3 shows the arrangement, a cascade whose inner loop is closed on a model. The depth controller reads the monitor and writes the propofol effect-site target. The concentration controller reads the pump's own copy of the pharmacokinetics and writes the propofol infusion rate. A second concentration controller does the same for remifentanil against an operator-set target, and a pressure controller writes the vasopressor rate. The control structure selector chooses how many of them are in service. In MANUAL the operator sets all three infusion rates directly. In target concentration the two inner loops run and the operator sets the concentration targets. In closed loop the depth controller sets the propofol target as well. A fourth setting replaces the depth and propofol concentration loops with the predictive controller of Section 3.5.4. The vasopressor loop and the signal-quality logic are independent switches, so the four structural choices can be made one at a time.

3.5.2 The Discrete Algorithm

All four controllers are instances of the velocity-form proportional-integral-derivative algorithm of Section 1.6.2, Equations (1.12) to (1.15), executed every $\Delta t_c = 0.05$ min. This chapter writes the action as $\mathscr{A}$, the a of Chapter 1, with $+1$ for a reverse-acting loop whose process gain is positive and -1 for a direct-acting one, since a here is the pressure coefficient of Equation (3.4). Anti-windup is back-calculation with the tracking constant $T_t = \tau_I$ for all four loops, so each scan removes $\Delta t_c / \tau_I$ of

Table 3.4 The Four Loops as Configured. Every Loop Is Proportional-Integral With Derivative Action Disabled, a 3 s Execution Period and No Rate Limit.

Loop	PV	MV (range)	K_c	τ_I
Depth ($\mathscr{A} = -1$)	BIS_m	$C_{e,\mathrm{SP}}$ (0–8)	0.045	3.5
Propofol ($\mathscr{A} = +1$)	$\hat{C}_{e,p}$	u_p (0–20)	5.0	2.5
Opioid ($\mathscr{A} = +1$)	$\hat{C}_{e,r}$	u_r (0–1.0)	0.16	4.0
Pressure ($\mathscr{A} = +1$)	P_m	u_v (0–1.0)	0.012	4.0

the clipped excess, and it runs only when there is integral action to unwind. The clamp acts on the emitted output alone.

Three properties of this implementation matter for the results of this chapter. Derivative action is taken on the measurement rather than on the error, and τ_D is zero for all four loops and is not exposed as a setting, so every loop in the application is a PI controller. No output carries a rate limit, and the only actuator bounds are the absolute ones in Table 3.4. MANUAL mode sets the integral so that the output reproduces the operator's setting exactly, so transfer in either direction is bumpless with no logic of its own. The inner loops are held in that tracking state whenever target control is out of service, and the depth controller is held in it whenever the loop is open, so switching structures mid-case does not bump the pumps.

The depth loop is the only direct-acting one, because more propofol produces a lower index, so the gain from this manipulated variable to this controlled variable is negative. Setting it reverse acting instead sends the target to zero on a light patient. The pressure loop is reverse acting because more vasopressor produces more pressure, and the opposite choice drives the syringe to its stop whenever surgical stimulation pushes the pressure above setpoint, which is most of an operation.

The opioid loop's integral time of 4 min is a little longer than the 1.7 min remifentanil effect-site lag plus the pump. After the awake start of scenario 2 the modeled concentration rises past its 4 ng/mL target to 4.45 at 3.2 min, dips to 3.87 at 5.8 min and stays within 5 % of the target from 4.1 min on, and the true index stays between 49.5 and 49.8 from twenty to forty minutes. From a maintenance start the same loop holds its target without moving. Setting the integral time to 0.7 min on its slider shows what too short an integral time does on this loop: the concentration overshoots to 7.3 ng/mL at about three minutes and then rings with a period of about 5.3 min.

3.5.3 The Measurement That Is a Calculation

The pump's model is a separate integration of Equation (3.1) advanced by explicit Euler at the controller period, driven by the infusion the pump is actually delivering, and built from the population parameters of Equation (3.2). It never sees ρ, S or φ. Explicit Euler at 3 s is adequate for rate constants of a few per minute and is what a device does. The inner loop is a PI controller on that model rather than the effect-site targeting algorithm of commercial pumps [1], so after an awake start the modeled effect site overshoots a 1.7 µg/mL target by 0.19 µg/mL at five minutes, stays within 0.043 µg/mL of it from ten minutes on and is 0.008 µg/mL above it at forty minutes. The consequence of a model that sees no disturbance is stated in the application's own verification suite and is worth restating here. With the central volume reduced by 20 %, the modeled effect-site concentration holds its target to better than 0.05 µg/mL after twenty-five minutes while the patient is more than 0.25 µg/mL past it (0.002 and 0.33 µg/mL in the current build). The inner loop cannot be wrong about its target and cannot be right about the patient.

🕮 **Pause.** Try Exercise 3.1 before reading on.

Signal-quality handling is four lines of configuration and it is the single most consequential switch in the application. With the controller set to trust the monitor, Equation (3.6c) passes the filtered value through whatever the quality indicator says. With validation selected, the controller holds its

last good measurement while the indicator is poor. No tuning substitutes for it, because a corrupted trace and a light patient are the same number.

3.5.4 Model Predictive Control

The fourth value of the structure selector gives the propofol infusion to a receding-horizon optimization of the kind proposed for hypnosis under total intravenous anesthesia [19, 20]. It has the move suppression, hard input box and soft output bands of Section 1.6.3, and it departs from the linear model of that section in two ways. The index enters through the nonlinear response surface, and one parameter of that surface is learned online. It writes u_p itself, while the depth and propofol concentration loops track the applied infusion at every scan with no preload, so a change of structure in either direction is free of a bump. The opioid loop, the pressure loop, the signal-quality logic and the pump's range act as they do in closed loop. Where the cascade corrects the response surface with an integral, the predictive controller corrects it with a parameter identified online [21].

Controller model. The pharmacokinetic half is the pump's own model, the delivered rate and the four concentrations of Equation (3.1) with $\varphi = 1$, built from the covariate entries the pump uses and the declared lag $\tau_p = 0.12\,\text{min}$ of Equation (3.5), discretized with an exact zero-order hold and started at each decision from the pump model's state. Ten minutes of 6 mg/(kg h) from empty give an effect site of 1.9089 µg/mL through the hold and through a fourth-order Runge-Kutta integration of the plant at $\varphi = 1$, and 1.9097 µg/mL through the pump's Euler copy, a residual of 0.044 %. The response half is Equation (3.3) at the declared defaults of Table 3.3 with $\rho_r = 1$, $S = 0$ and an estimate $\hat{\rho}$ in place of ρ, so the monitor is predicted as $\hat{y} = h(\hat{C}_{e,p}, \hat{C}_{e,r}; \hat{\rho}) + \hat{b}$, with h the index of Equation (3.3d) and $\hat{b}$ an output bias. The opioid effect site is assumed to approach its target with $\tau_R = 1.5\,\text{min}$. The surface is not linearized, because a linearization at an awake start is flat. Since h is static and monotone in $\hat{C}_{e,p}$, the setpoint is mapped by bisection into an effect-site reference C_k^* at every step k of the horizon,

$$h(C_k^*, \hat{C}_{e,r,k}; \hat{\rho}) = \text{BIS}_{\text{SP}} - \hat{b} \tag{3.7}$$

capped at 8 µg/mL, and the band edges 40 and 60 are mapped the same way into C_k^+ and C_k^-. At an index of 50 with 4 ng/mL of opioid and $\hat{\rho} = 1$ the reference is 1.6893 µg/mL and the slope of h is $K = 10.04$ index units per µg/mL, against 10.31 identified from target steps in closed loop.

Learned potency. The estimate is a one-parameter recursive least-squares fit of the monitor against the model's own prediction of it, in the potency $g = 1/\hat{\rho}$ bounded so that $0.5 \le \hat{\rho} \le 2$, the declared range of sensitivity. The prediction h_k and the regressor $\partial h_k / \partial g$, both evaluated at the pump model's effect sites, pass through the declared monitor smoothing as F_k and ϕ_k, with $\alpha_B = 1 - \exp(-\Delta t_c / \tau_B)$, as shown in Equation (3.8).

$$z_k = \text{BIS}_{m,k} - F_k + \phi_k\, \hat{g}_{k-1}, \quad L_k = \frac{P_{k-1}\,\phi_k}{\lambda_k + \phi_k^2\, P_{k-1}} \tag{3.8a}$$

$$\hat{g}_k = \text{clamp}(\hat{g}_{k-1} + L_k\,(z_k - \phi_k\,\hat{g}_{k-1}),\ 0.5,\ 2) \tag{3.8b}$$

$$P_k = \min[(P_{k-1} - L_k\,\phi_k\,P_{k-1})/\lambda_k,\ 0.5] \tag{3.8c}$$

Here z_k is the regression target, L_k the gain and P_k the covariance, which starts at 0.5. Three guards keep a quiet signal from walking the estimate away. The forgetting factor λ_k is 0.98 per scan only while the pump model's effect site has moved by more than 0.1 µg/mL in 3 min, and one otherwise. The covariance returns to 0.5 when the depth setpoint changes. Samples are skipped while the signal quality indicator is below 50 and for 1.5 min afterward. Over a quiet maintenance hour of scenario 5 the estimate moves by 0.14, 0.83 and 0.17 % at ρ of 0.5, 1 and 2 and P falls. The bias is $\hat{b}_k = \hat{b}_{k-1} + \alpha_D(\text{BIS}_{m,k} - F_k - \hat{b}_{k-1})$ with $\alpha_D = 1 - \text{e}^{-\Delta t_c/\tau_D}$ and $\tau_D = 1\,\text{min}$, an output disturbance model [22] frozen while validation holds the measurement, which leaves a mean $|\text{BIS} - 50|$ of 0.006 and 0.079 over the last twenty minutes of scenario 5 at ρ of 0.5 and 2 with the learning and the noise off.

Decision and objective. A plan is made every $T_s = 0.25$ min, five scans, over $N = 40$ steps (10 min) in ten move blocks of $1, 1, 1, 2, 2, 3, 4, 6, 8, 12$ steps, and the pump applies the first block's infusion for the 15 s until the next plan. Each plan minimizes, over the ten block moves,

$$J = \sum_{k=1}^{N} K^2 \left(\hat{C}_{e,p,k} - C_k^*\right)^2 + \mu K^2 g_u^2 \sum_{j=1}^{10} \Delta u_{p,j}^2 + \sum_{i=1}^{4} \left(w_i s_i^2 + v_i s_i\right) \quad (3.9)$$

subject to $C_k^- - s_1 \le \hat{C}_{e,p,k} \le C_k^+ + s_2$, $\hat{C}_{e,p,k} \le \hat{C}_{e,p,0} + (P_m - P_{lo})/a + s_3$, $\hat{C}_{e,p,k} \le 8 + s_4$, $s_i \ge 0$ and the hard box $0 \le u_p \le 20$, the clamp of the concentration loop. Here $g_u = 0.318$ µg/mL per mg/(kg h) is the model's effect site after ten minutes of a unit infusion, so a move is priced at μ times the square of the index shift it would cause if held, and each slack is one number for the whole horizon. The move weight is $\mu = 0.1$. The light side of the band carries $w_1 = 10^3 K^2$ and $v_1 = 100K$, and the deep side ten times as much, because the pump can always add drug and can never take any back (Section 3.4). The pressure floor, written through the declared $a = 6.4$ mmHg per µg/mL without the circulatory lag, the opioid or the vasopressor, carries $w_3 = 10^3 a^2$ and $v_3 = 100a$, and the ceiling $w_4 = 10^7$ and $v_4 = 10^6$. The limits are read from the run's parameters as the existing logic reads them. The weight μ was chosen once on scenario 12 from 0.03 to 3 as the smallest value whose infusion travel stays within about 15 % of the closed loop's (76 against 68 mg/(kg h)), and the horizon was not tuned.

Acceptance and fallback. A plan takes effect only with an optimal status, every planned infusion inside the box, the first planned infusion equal to the applied one and a finite cost, slacks and predicted index. Otherwise the historian is marked and the closed loop continues from its tracked state until a valid plan returns. Seven kinds of forced failure on scenario 12 handed over with a step of at most 0.018 mg/(kg h), except 0.657 one minute into the induction. The first plan is solved at $t = 0$. In Chrome the 360 plans of scenario 12 cost a mean of 601 and 610 µs in two runs, which is 0.24 % of one core at the 1 min/s playback speed.

Three models kept apart. The plant is the thirteen-state model at the run's patient, the instrument is Equation (3.6) with its noise and artifact, and the controller's model is the pump's pharmacokinetics, the population surface with $\hat{\rho}$ and $\hat{b}$, and the slope a. It never reads ρ, ρ_r, S, φ or the true index, so $\hat{\rho}$ is the potency of the drug pair as the population surface reads the monitor, which Section 3.6.9 shows is not the patient's sensitivity.

3.6 Scenarios and Measured Results

The thirteen scenarios in Table 3.5 are arranged so that each one adds a single mechanism to the one before. Every scenario fixes the starting state, the control structure, the four tunings, the patient and the timed events, and carries the briefing and completion rule of Section 1.6.4. Table 3.6 reports what the application measures over each run, using its own scorecard quantities. Peak values, minutes outside the band and delivered dose are quoted rather than a bare integrated error, because integrated absolute error keeps accruing after a transient ends and is comparable only within a fixed window.

3.6.1 A Rate Is Not a Trajectory

Scenario 1 makes the point before any controller is involved. Setting the propofol pump to 3.43 mg/(kg h) and the remifentanil pump to 0.149 µg/(kg min), the rates that hold the maintenance state still, and waiting, gives an index of 84.2 at two minutes, 69.5 at five, 61.5 at ten, 56.2 at twenty and 52.4 at forty, still falling. The effect site fills through a plasma that is itself filling a 4.27 L central compartment while two peripheral compartments draw off it, and the sum of the two transfer rate constants $k_{12} + k_{13} = 0.571/\text{min}$ is larger than the clearance $k_{10} = 0.384/\text{min}$. Arriving in two minutes requires overshooting the plasma deliberately and coming off again before the effect

Table 3.5 The Scenario Ladder and What Each Step Introduces

#	Configuration	What it shows
1	MANUAL, awake start	a rate is not a trajectory
2	Target control, $\rho = 1$, $C_{e,\mathrm{SP}} = 1.7$, $C_{e,r} = 4$	model inversion on the patient the model fits
3	Target control, $\rho = 1.6$, $C_{e,r} = 2$	exact inner loop, wrong depth
4	Closed loop, $\rho = 1.6$, incision at 40 min	feedback as identification, then a target at its bound
5	Closed loop, maintenance start, nothing happens	a plant with no steady state
6	Closed loop, incision at 8, closing at 40 min	a disturbance that moves the plant
7	Closed loop, opioid stepped 0.5, 8, 4 ng/mL	synergy and the price of the split
8	Closed loop, diathermy 10–16 and 30–38 min	an artifact in the dangerous direction
9	Target control, 20 % volume loss at 10, loop closed at 45 min	a disturbance the model cannot see
10	Closed loop, $\rho = 0.7$, $\Delta P_0 = -12$ mmHg, vasopressor loop in service	the constraint is on another loop
11	Closed loop, $S = 0.85$ from the start, closing at 20 min	one-sided authority at the end of a case
12	Predictive control, awake start, $\rho = 0.5$, incision at 30, closing at 60 min	a plan that learns the gain it lacks
13	Challenge: 120 min, $\rho = 1.45$, $\rho_r = 0.85$, five events	structure before tuning

Table 3.6 Measured Outcomes Over the Window Shown, Read From the Application's Own Scorecard. The Index Extremes Are of the True Index and Not of the Monitor. In the Six Scenarios That Start Awake the Maximum Is the Awake Value of 93 and the Minutes Above the Band Include the Induction Transit.

#	Configuration	Window	True index		In band	Above 60	Below 30	Lowest P	Propofol
		(min)	lowest	highest	(%)	(min)	(min)	(mmHg)	(mg)
1	MANUAL, maintenance rates	40	52.4	93.0	70	12.0	0.0	80	160
2	Target control, population patient	40	48.0	93.0	95	2.1	0.0	78	215
3	Target control, resistant patient	40	65.9	93.0	0	40.0	0.0	79	215
4	Closed loop, resistant patient	40	51.0	93.0	94	2.5	0.0	65	481
4	Closed loop, through the incision	90	51.0	93.0	87	11.5	0.0	61	1415
5	Closed loop, undisturbed	90	49.5	50.7	100	0.0	0.0	79	339
6	Closed loop, incision and closing	70	37.1	66.8	90	3.7	0.0	69	483
7	Closed loop, opioid swept	110	29.0	57.1	94	0.0	1.3	68	456
8	Diathermy, controller trusts monitor	45	31.8	50.2	63	0.0	0.0	62	333
8	Diathermy, last good value held	45	48.9	50.2	100	0.0	0.0	78	178
9	Blood loss, loop closed at 45	90	43.2	50.7	100	0.0	0.0	76	275
10	Depth against pressure	90	31.1	49.0	29	0.0	0.0	67	340
11	End of case, opioid at 4	90	35.5	50.6	95	0.0	0.0	67	427
11	End of case, opioid at 8	90	33.0	50.8	95	0.0	0.0	73	185
12	Predictive control, sensitive patient	90	35.8	93.0	92.5	3.6	0.0	81	322
13	Challenge as shipped	120	60.0	93.0	17	99.7	0.0	72	482
13	Challenge, structure configured	120	33.5	93.0	91	4.4	0.0	59	795

site catches up. With the target-controlled pump in service that trajectory is computed rather than remembered, and scenario 2 delivers 10.2 mg/(kg h) at one minute, 7.3 at three, 4.8 at ten and 3.8 at thirty, reaching an index of 50 at 3.1 min and dipping to 48.0 at 4.2 min as both inner loops carry their effect sites about a tenth past target before settling.

Table 3.7 Settled Index at 40 min Under Model Inversion and Under Closed-Loop Control, Against Patient Sensitivity. The Modeled Concentration Is the Same in Every Row of the Left Half, 0.008 µg/mL Above Its Target. The Closed-Loop Values Are Read Before the Incision of Scenario 4.

	Target control		Closed loop		
ρ	index	P (mmHg)	index	$C_{e,\mathrm{SP}}$	P (mmHg)
0.70	52.1	80	50.6	1.90	79
0.85	56.3	80	50.6	2.29	77
1.00	59.6	80	50.6	2.68	74
1.30	64.2	80	50.8	3.43	69
1.60	67.4	80	51.1	4.15	65
2.00	70.4	80	51.7	5.03	59

3.6.2 What Model Inversion Cannot Do

Table 3.7 runs the scenario 3 configuration, a commanded 1.7 µg/mL of propofol with the opioid at 2 ng/mL, across the range of patient sensitivity, and runs the scenario 4 configuration, which is the same case with the depth loop closed, beside it.

The left half of Table 3.7 spans 18.3 index units while the pump reports the same modeled concentration in every row, and only the three lowest sensitivities land inside the surgical band, the population patient at 59.6 with 0.4 to spare. The right half spans 1.1 index units, and the mechanism is entirely in the fourth column, where the commanded concentration moves from 1.90 to 5.03 µg/mL. That column is a parameter estimate produced by an integral, on a plant where nobody told the controller which parameter was missing. The last column is the price. Because the depth is bought with a vasodilator, the settled pressure falls from 79 mmHg at $\rho = 0.70$ to 65 mmHg at $\rho = 1.60$ and to 59 mmHg at $\rho = 2.0$, which is one millimeter of mercury below the perfusion limit the application marks.

Identification is not fast, and it cannot be. In scenario 4 the commanded concentration is 3.13 µg/mL at five minutes, 3.83 µg/mL at twenty and 4.07 µg/mL at thirty-nine, against the 4.27 µg/mL this patient requires at 2 ng/mL of opioid. The index stays within five units of setpoint from 9.4 min onward. Half an hour to identify a patient 60 % away from the population is consistent with what published closed-loop systems achieve [5, 13], and it follows directly from the effect-site lag of Section 3.4.

> **TRY IT IN THE APP 4 · Closing the Loop**
>
> Load scenario 4 and press Run. Watch the Propofol target trace under the Depth of hypnosis strip rather than the index itself. The target climbs as the outer integral finds this resistant patient: 3.13 µg/mL at five minutes, 3.83 at twenty and 4.07 at thirty-nine, and the index stays within five units of 50 from 9.4 min onward.

3.6.3 A Disturbance That Moves the Plant

The incision in scenario 6 takes the index from 50.2 to 66.8 within one integration step, because the index is algebraic in the stimulation, and the controller answers by raising the target, which is ordinary. What is not ordinary is that the effective C_{50} moved by a factor of 1.72 at the same instant, so the gain and the operating point changed together and the loop is at its slowest exactly when it is being asked to work hardest. Closing at forty minutes is the same disturbance reversed, and it is the one that costs. The index falls to 37.1 half a minute after the closing and spends 10.3 min below 45. The propofol pump is still delivering 7.2 mg/(kg h) at the lowest index and 4.0 mg/(kg h) two minutes after the closing, because the depth loop lowers the target by only 0.7 µg/mL in the first minute, but stopping the pump outright at the closing would leave the lowest index within 0.1 of the same value.

Scenario 4 shows the other end of the same mechanism. On the resistant patient at full stimulation the depth controller runs its output up over the forty-eight minutes after the incision and reaches

8 µg/mL, the upper bound on the concentration target, at 88.5 min. At 90 min the index is 52.0 and the pressure is 60.6 mmHg. The controlled variable is close to setpoint and the loop has no authority left. The honest reading of that trend is not a tuning problem, because a resistant patient at full stimulation needs more than the hypnotic alone can safely give, and the remedy is a higher opioid target.

3.6.4 The Cost of Believing the Monitor

Scenario 8 runs two episodes of electrocautery totaling fourteen minutes. With the controller trusting the monitor, the displayed index rises while the true index falls to 31.8, the pressure reaches 62 mmHg and the patient receives 333 mg of propofol in forty-five minutes. With signal-quality validation selected and nothing else changed, the lowest true index is 48.9, the lowest pressure is 78 mmHg and the dose is 178 mg. Over the same forty-five minute window the integrated absolute index error falls from 367 to 14 index-minutes, a factor of 26, and the time inside the surgical band rises from 63 % to 100 %. The rejected samples never enter the monitor's smoothing, so after each episode the reading resumes within 1.8 index units of the true index. Every second of the trusting run is invisible while it happens, because a corrupted trace and a light patient produce the same number on the same screen.

> **TRY IT IN THE APP 8 · The Measurement Lies, Upward**
>
> Load scenario 8 and press Run with Signal quality at Trust the monitor. During the cautery from 10 to 16 and 30 to 38 min, watch the monitor line rise on the Depth of hypnosis strip while the true index falls, to 31.8. Reset, set Signal quality to Hold on low quality and run again: the lowest true index is now 48.9.

3.6.5 A Disturbance the Model Cannot See

Scenario 9 removes a fifth of the central volume from a patient on target control. The pump's model reports 1.702 µg/mL against a target of 1.700 µg/mL because it is holding a number it computed itself, while the patient reaches 2.05 µg/mL and the index falls to 43.2. Nothing anywhere in the pump indicates that anything has happened. Closing the loop at forty-five minutes brings the index within one unit of 50 from 63.8 min, about nineteen minutes later, by lowering the commanded concentration. At 89 min the target is 1.18 µg/mL, the model holds 1.09 µg/mL and the patient holds 1.32 µg/mL. The model is still wrong. It is being asked for the wrong number in exactly the way that makes the patient right, which is a more honest description of what an outer loop does than the phrase error correction suggests.

3.6.6 The Constraint Is on Another Loop

Scenario 10 gives a sensitive patient a baseline pressure 12 mmHg low and a surgeon who wants twelve more index units of depth. Taking the setpoint from 50 to 38 moves the propofol effect site from 1.27 to 2.22 µg/mL, and the pressure holds 70 mmHg only because the vasopressor loop takes itself to 0.147 µg/(kg min) by fifty minutes. Moving the same depth onto the opioid brings the vasopressor back to 0.067 µg/(kg min) by ninety minutes, less than half, at the same depth and the same pressure. The loop with the setpoint everyone is discussing is not the loop that limits the plant, and the only way to see that is to have the other one on the screen.

3.6.7 The End of the Case

Scenario 11 ends a case run at full stimulation. It starts with the propofol effect site and the concentration target both at 4.3 µg/mL, the requirement under stimulation, and the index stays between 49.8 and 50.2 until the closing. Closing at twenty minutes takes the index from 50.1 to 35.5 a quarter of a minute later, and it spends 13.0 min below 45 while the propofol rate falls from 8.5 to 3.8 mg/(kg h) over two minutes, because the drug is in the effect site and the only way out is redistribution and metabolism. The dressings at 42 min take the remaining stimulation away and add a second, shallower dip to 44.4. Running the same case with the opioid target at 8 ng/mL instead of 4 ng/mL deepens the dip to 33.0, although the propofol pump is commanded to zero 1.8 min after the

closing. The time below 45 after the closing is shorter, 11.5 min instead of 13.0, because the opioid has taken over part of the depth and less propofol has to wash out, but the higher target costs another 8.9 min below 45 at the start, where the opioid climbs to its target faster than the depth loop takes propofol off and the index reaches 40.8. A deeper dip is the opposite of what the pharmacology of a short-acting opioid suggests. The reason is that nothing in the arrangement asks for less of it, because the target-controlled pump goes on holding 8 ng/mL after the stimulation has gone. A short-acting drug is only short-acting if something turns it down.

> ✦ **DID YOU KNOW? More Opioid, Less Propofol, a Deeper Dip**
>
> From the simulator: raising the opioid target of scenario 11 from 4 to 8 ng/mL cuts the propofol given over 90 min from 427 to 185 mg, yet the lowest true index after the closing falls further, to 33.0 against 35.5 (Table 3.6).

Turning it down works. Cutting the opioid target to 1 ng/mL at the moment closing begins leaves the depth of the dip unchanged at 35.5, because that part is the stimulation falling in one step, and takes the time below 45 from 13.0 to 3.8 min. Lightening six minutes before closing instead, to an index setpoint of 55 with the opioid target cut to 1 ng/mL at the same time, stops the dip at 43.3 with 1.8 min below 45 and no time above 60. A faster depth loop would shorten the time below 45 as the opioid cut does and leave the lowest index where it is. Only anticipation raises the bottom of the dip, and it is feedforward from an event that nobody has instrumented.

🕮 **Pause.** Try Exercise 3.3 before reading on.

3.6.8 Structure Before Tuning

The challenge is a case run to 120 min on a patient who needs 45 % more propofol than the model expects and is slightly more sensitive to the opioid than the population ($\rho_r = 0.85$), with a baseline pressure 8 mmHg low, monitor noise of 1.5 index units, an incision at eight minutes, cautery from twenty-two to twenty-nine, blood loss at forty-two and closing at sixty-two. It loads on target control with weak tuning ($K_c = 0.008$ and $\tau_I = 20$ min on the depth loop), a trusting monitor, the vasopressor by hand and an opioid target of 2 ng/mL. The application judges the case as given on the structure the challenge names and on three quantities of the true index. For the whole run the case must stay in Auto with the depth loop closed, the signal-quality hold selected and the vasopressor loop in service, and the opioid target must be above the shipped 2 ng/mL when the run starts. The largest deviation of the index from 50 after 79.2 min, which the application calls the last third of the run, must not exceed 6, and over the whole run the index must spend less than 2 min below 30 and less than 10 min above 60. Hypotension is displayed beside the verdict and is not judged. Table 3.8 adds the four structural choices one at a time to the shipped configuration and reports the judged quantities beside the pressure and the dose.

Reading Table 3.8 downward separates four different kinds of help. Closing the loop with the default tuning raises the time in the surgical band from 17 to 82 %, cuts the time above 60 from 99.7 to 12.55 min and the late deviation from 10.1 to 5.9, and it is the only change that addresses the patient. On its own it creates a hypotension problem that did not exist before, because it buys the depth with a vasodilator and spends 74.0 min below 60 mmHg with a lowest pressure of 38 mmHg. Signal validation is worth 17.1 min of that hypotension and 86 mg of propofol, and it addresses the measurement rather than the patient. The vasopressor loop is worth another 54.5 min, addresses the constraint and leaves the index untouched. Raising the opioid target addresses the split, and it is the change that takes the time above 60 under its limit and reduces the propofol dose by 43 %, from 1399 to 795 mg. None of the four is a tuning constant, and the tuning is identical in the last five rows. The level loop of Volume 3, Chapter 2 makes the same point on a tank, where each structure rejects the disturbance that enters its own measurement path whatever the tuning. Figure 3.4 draws three columns of the table as bars.

Table 3.8 Structural Choices Added One at a Time on the 120 min Challenge. Each Row Above the Rule Keeps Everything From the Row Above It. $|e|_{\max}$ Is the Largest $|\text{BIS} - 50|$ From 79.2 min, Light Counts Minutes Above an Index of 60, Low P Counts Minutes Below 60 mmHg, $P_{\min}$ Is the Lowest Pressure, and No Row Spends Time Below an Index of 30. The Last Row Adds Only the Opioid Target to the Closed, Tuned Loop. Pass Is the Application's Verdict, Which Also Requires the Three Switches for the Whole Run and an Opioid Target Above 2 ng/mL at the Start.

Added	$\|e\|_{\max}$	Light (min)	In band (%)	Low P (min)	$P_{\min}$ (mmHg)	Propofol (mg)	Pass
As shipped	10.1	99.7	17	0.0	72	482	no
Closed loop, tuned	5.9	12.55	82	74.0	38	1486	no
Signal validation	5.2	12.55	84	56.9	40	1399	no
Vasopressor loop	5.2	12.55	84	2.35	56	1399	no
Opioid target 5	3.3	4.4	91	1.18	59	795	yes
Opioid 5 only	4.4	4.4	88.5	11.3	53	928	no

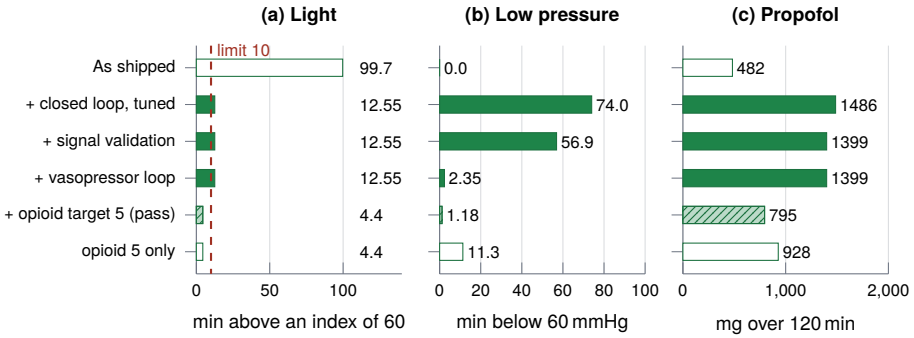


Figure 3.4 The structural choices of Table 3.8 on the challenge: (a) minutes above an index of 60 against the limit of 10 (dashed), with the values in a column at the right, (b) minutes below 60 mmHg and (c) propofol given. White: the shipped configuration and the opioid target raised alone. Solid: the cumulative rows that fail. Hatched: the configured row that passes. Closing the loop cuts the light time and creates hypotension, the vasopressor loop removes most of the hypotension, and only the opioid target brings the light time under its limit and the dose down.

The configured run passes with a late deviation of 3.3, 4.4 min above 60 and none below 30, margins of 2.7 index units, 5.6 min and 2 min, and it gives the same result whether its settings are made before Run or carried through Reset. The shipped configuration lacks all three switches, keeps the opioid at 2 ng/mL and misses the index limits by 4.1 index units and by 89.7 min. It is not a clean case. Of the time above 60, 1.55 min is the induction transit and 2.85 min follows the incision, and the index also spends 6.7 min below the surgical band, 1.3 min of it in an undershoot to 38.8 during induction and 5.4 min after the closing, where it reaches 33.5. That deep excursion is the end-of-case mechanism of scenario 11 arriving on a case that also carries blood loss.

The verdict judges what the challenge text states. The challenge asks for a control structure, a signal-quality gate, a pressure loop and a drug split to be chosen before any tuning, its briefing calls the gate and the closed outer loop structural requirements, and its explanation names the vasopressor loop in service and an opioid target above the shipped one. The last row of Table 3.8 closes the loop and raises the opioid target with the monitor trusted and the vasopressor by hand. Its index meets the three index limits, with a late deviation of 4.4 and 4.4 min above 60, but the pressure spends 11.3 min below 60 mmHg and falls to 53 mmHg, and the verdict fails it on the two missing switches. The index limits alone reject the weaker configurations as well. Target control fails them in each

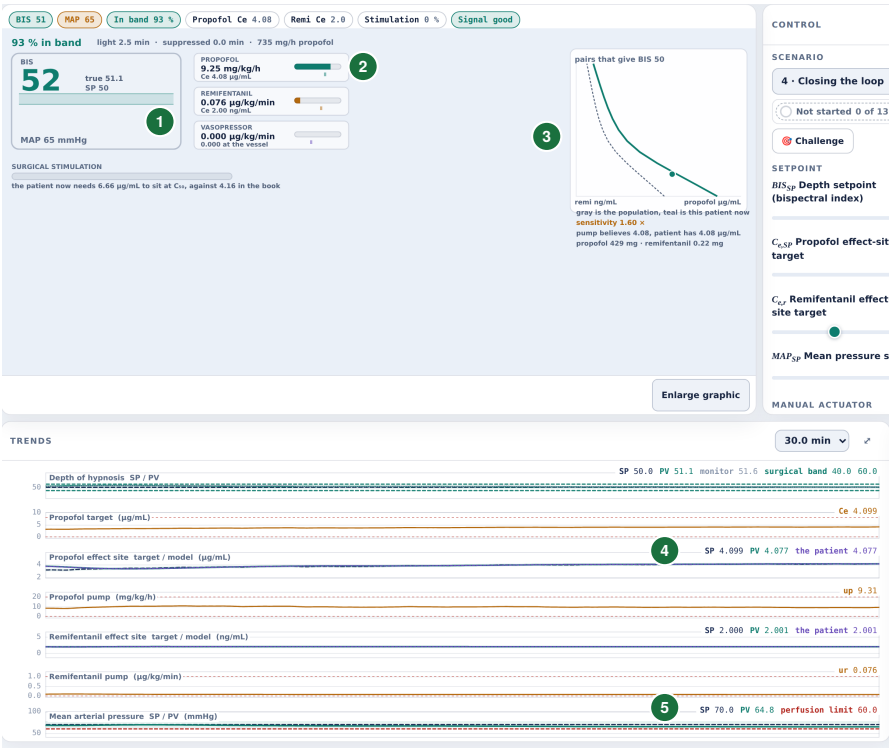


Figure 3.5 Depth held on a patient who needs 60 % more propofol than the population model expects, scenario 4 at 35 min, with the trends covering the last thirty minutes. The true index is 51.1 against a setpoint of 50 while the propofol target below it has been driven to 4.10 µg/mL, where the population model asks for 2.67 at the same opioid concentration of 2.00 ng/mL, and the separation between those two numbers is the patient. The isobologram inset shows the operating point sitting on the patient's own curve rather than on the population curve, the arterial pressure strip shows what the extra propofol costs (64.8 mmHg), and the remifentanil traces show the opioid loop holding its target. Callouts: (1) the depth readout with the true index and the setpoint, (2) the pump tiles, (3) the isobologram inset, (4) the propofol effect-site target and model trend, (5) the mean arterial pressure strip. The crop removes the header, the scorecard and most of the control panel.

of six combinations of targets tried, the full structure with the opioid at 5 ng/mL and the shipped weak tuning spends 46.25 min above 60, and the full structure with the opioid left at 2 ng/mL spends 12.55 min above 60. The ±6 band applies only from 79.2 min, seventeen minutes after the last event, so tracking through the incision, the cautery and the blood loss is judged through the two time limits and the structure, and time between an index of 30 and the bottom of the surgical band at 40 is not counted. The pressure loop is required as a switch rather than through a pressure limit, so a pass says that the structure was in place and the depth held, and the hypotension column of Table 3.8 still has to be read beside it. The application is shown in Figure 3.5 running the scenario that carries the argument, the resistant patient of scenario 4 with the depth loop closed.

3.6.9 Predictive Control Against the Closed Loop

Table 3.9 runs scenarios 4, 5, 6, 8 and 9 for 90 min from their own settings with only the structure changed and the patient sensitivity at 0.5, 1 and 2, and the challenge for 120 min on its suggested settings, with the same seed and instruments in every row. Scenario 8 keeps its trusting monitor, scenario 9 closes the depth loop from the start, and the predictive controller uses scenario 12's settings everywhere. Five rows are compared: the closed loop with its shipped tuning, the predictive controller,

Table 3.9 Integrated Absolute Index Error in BIS·min, With the Depth Below an Index of 40 in BIS·min in Parentheses, Over 90 min for Scenarios 4 to 9 and 120 min for the Challenge (13, Suggested Settings), at Patient Sensitivity ρ. The Predictive Controller Uses the Settings of Scenario 12 Throughout. Scenario 8 Trusts the Monitor, and Scenario 9 Closes the Depth Loop From the Start.

Scenario	ρ	Closed loop	Predictive	Predictive, no learning	Retuned loop	Learned-gain loop
4	0.5	254 (18.8)	155 (0.0)	157 (6.5)	288 (41.8)	333 (15.8)
4	1	297 (0.0)	161 (0.0)	146 (0.0)	216 (4.7)	237 (0.0)
4	2	695 (0.0)	622 (0.0)	632 (0.0)	537 (0.0)	561 (0.0)
5	0.5	82 (2.9)	66 (2.9)	67 (2.9)	120 (2.9)	132 (3.8)
5	1	19 (0.0)	32 (0.0)	29 (0.0)	31 (0.0)	19 (0.0)
5	2	134 (0.0)	51 (0.0)	45 (0.0)	59 (0.0)	74 (0.0)
6	0.5	214 (9.6)	182 (10.1)	162 (10.3)	212 (9.4)	221 (10.2)
6	1	315 (5.8)	138 (6.9)	126 (7.2)	141 (8.4)	218 (6.4)
6	2	548 (0.0)	216 (2.4)	215 (2.3)	297 (8.4)	370 (5.3)
8	0.5	523 (176.8)	666 (324.0)	673 (328.2)	682 (302.4)	576 (110.2)
8	1	454 (88.6)	350 (124.9)	348 (124.9)	484 (203.1)	452 (87.7)
8	2	445 (13.9)	179 (0.0)	174 (0.3)	403 (50.3)	431 (37.7)
9	0.5	103 (2.9)	75 (2.9)	78 (2.9)	204 (2.9)	171 (3.8)
9	1	53 (0.0)	41 (0.0)	41 (0.0)	50 (0.0)	57 (0.0)
9	2	103 (0.0)	59 (0.0)	58 (0.0)	82 (0.0)	86 (0.0)
13	1.45	479 (20.8)	251 (10.7)	242 (10.8)	358 (22.8)	354 (13.9)

Table 3.10 Three Decisive Cases Under the Five Rows. Scenario 12 ($\rho = 0.5$) and Scenario 4 ($\rho = 2$) Over 90 min, the Challenge on Its Suggested Settings Over 120 min. t_b Is the Time to the Band, Below 40 the Minutes and the Depth Under It in BIS·min, IAE in BIS·min, P_{lo} the Minutes Under 60 mmHg, Plans the Optimizations Solved and Rejected.

Row	t_b	Below 40		IAE	Propofol	P_{lo}	Plans
	(min)	(min)	depth		(mg)	(min)	
Scenario 12, sensitive patient							
Closed loop	1.20	6.39	38.2	341	330	0.00	—
Predictive	1.20	3.11	7.5	211	322	0.00	360/0
No learning	1.19	4.86	17.2	207	323	0.00	360/0
Retuned loop	1.20	7.60	60.7	352	336	0.00	—
Learned gain	1.22	7.72	40.2	423	339	0.00	—
Scenario 4, resistant patient							
Closed loop	3.04	0.00	0.0	695	1562	15.65	—
Predictive	2.99	0.00	0.0	622	1595	15.98	360/0
No learning	3.12	0.00	0.0	632	1587	20.40	360/0
Retuned loop	3.00	0.00	0.0	537	1667	44.47	—
Learned gain	3.00	0.00	0.0	561	1638	37.87	—
Challenge, suggested settings							
Closed loop	1.55	6.74	20.8	479	795	1.18	—
Predictive	1.55	3.18	10.7	251	815	0.00	480/0
No learning	1.56	3.21	10.8	242	809	0.00	480/0
Retuned loop	1.56	7.11	22.8	358	828	1.03	—
Learned gain	1.57	4.88	13.9	354	822	0.51	—

the same with the learning off ($\hat{\rho} = 1$, bias only), the closed loop retuned once by the simple internal model control rule with the closed-loop time constant equal to the dead time, and the closed loop with its depth gain scheduled on the learned sensitivity, $K_c = 0.045\,\hat{\rho}$. The retuned loop rests on a first-order-plus-dead-time fit of $\pm 0.2\,\mu g/mL$ target steps from the maintenance start with the noise off, a gain of -10.31 index units per µg/mL, a lag of 1.13 min and a dead time of 1.33 min with a residual of 3.6 % of the response, and gives $K_c = 0.0413$ and $\tau_I = 1.13$ min. The time to the band differs by at most 0.63 min among the rows of any case, and the 19,440 plans include no fallback.

Scenario 12 induces a patient who needs half the propofol the model expects ($\rho = 0.5$) from an awake start, with the opioid target at 4 ng/mL, an incision at 30 min and closing at 60 min. The plan opens the pump to 20 mg/(kg h) and cuts the infusion to 8.1 at 1.25 min and to zero by 1.55 min, because its model already holds enough drug in the plasma to carry the effect site to the reference, while the closed loop stays at or above 18 until 1.45 min and still delivers 4.6 at 2.75 min. The index falls to 27.7 under the closed loop and turns at 39.5 under the plan, with 31.5 against 0.2 BIS·min below the band in the first half hour, as shown in Figure 3.6a. In minutes below 40 over that half hour the closed loop spends 4.03, the plan 0.68 and the plan with the learning off, which turns at 33.9, spends 2.38. Over the case the index error falls from 341 to 211 BIS·min and the closed loop's 1.6 min below an index of 30 disappear (Table 3.10). The closing costs the same under every structure, 2.43 min below 40 against 2.36, because that drug is in the effect site before the stimulation goes.

Switching the corrections off one at a time apportions the induction. With the learning off and the bias frozen at its first value, the plan holds the sensitive patient at the population's concentration and spends 28.2 min of the first thirty below 40 (82.2 BIS·min), worse than the closed loop. Either correction then suffices and the two substitute. Adding the learning first removes 81.9 of the 82.2 BIS·min, while adding the bias first removes 72.6 and the learning then 9.3. Over the whole case only the bias handles the incision (164 against 65 BIS·min from 30 to 50 min), and with the learning off the whole-case error is 207 against 211, a tie, so the learning pays only below the band, 17.2 against 7.5 BIS·min. Given to the closed loop as a gain schedule, the learned number raises its error from 341 to 423, because a lower gain on a sensitive patient slows the recovery without changing the first minutes on the pump's ceiling.

The learned sensitivity reads 0.535 at 2 min and 0.502 at 5 min against the patient's 0.50, and after the incision it climbs to 1.26 at 50 min although the patient has not changed, as shown in Figure 3.6b. The equivalent value, at which the population surface returns the patient's true index at the true effect sites, is 1.28 there. Stimulation multiplies both half-maximal concentrations by 1.72 in Equation (3.3a), and one potency can express that only as a less sensitive patient (on propofol alone the equivalent would be 0.86). After the closing the estimate returns to 0.62 at 80 min against an equivalent 0.60.

Across Table 3.9 the predictive controller lowers the index error on scenarios 4, 6 and 9 at every sensitivity and on scenario 5 at 0.5 and 2. With the hold selected it also lowers the error on scenario 8, 58, 23 and 43 against 81, 24 and 114 BIS·min at ρ of 0.5, 1 and 2. It loses scenario 5 at $\rho = 1$, 32 against 19 BIS·min, because the 1 min bias passes the monitor noise to the plan (2.2 against 0.9 with the noise off, 25.4 with a 3 min bias). It also loses scenario 8 with the monitor trusted, 324.0 against 176.8 BIS·min below 40 at $\rho = 0.5$ with a lowest index of 15.6 against 24.5, and 124.9 against 88.6 at $\rho = 1$ although its index error is lower there. The bias follows the 34-unit artifact within a minute, and on a sensitive patient the inverted surface turns that apparent lightness into a large reference. Freezing the bias on a poor signal would credit the plan with the validation this row has switched off, so it is not done.

The retuned loop beats the plan on scenario 4 at $\rho = 2$, 537 against 622 BIS·min, at a cost of 44.5 min below the 60 mmHg floor against 16.0, all of the plan's minutes falling in the induction and 10.5 min of the retuned loop's after the incision. Its faster integral goes deeper on sensitive patients, 41.8 against 18.8 BIS·min below the band on scenario 4 at $\rho = 0.5$. On the challenge's suggested settings the plan lowers the index error from 479 to 251 BIS·min with a largest late deviation of 1.7 against 3.3 and no time under the pressure floor, yet the verdict fails it because the challenge requires the closed depth loop for the whole run, and the rules were left as they are. From the shipped challenge settings with only the structure changed, the pressure spends 30.6 min below the floor with the vasopressor by hand, so the soft pressure row does not replace the pressure loop.

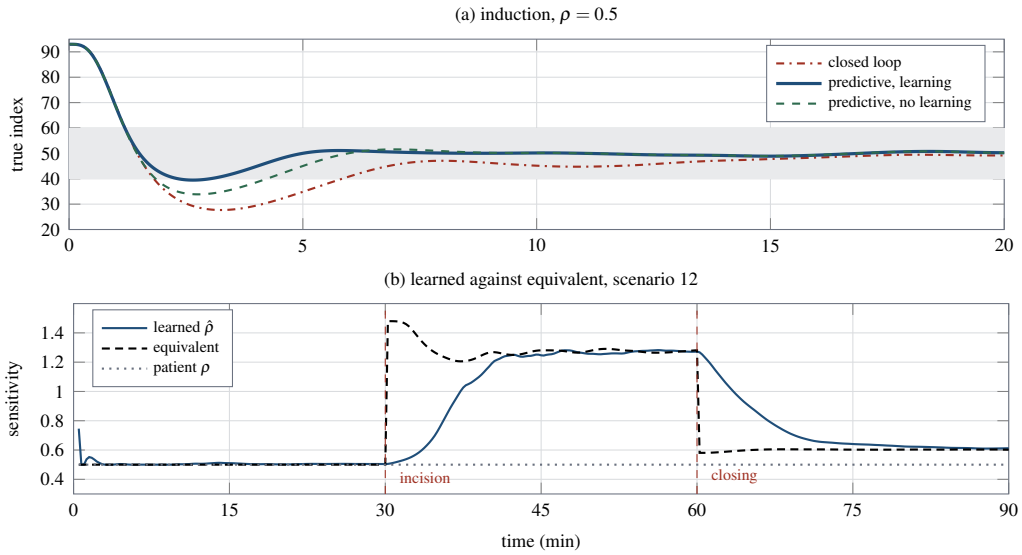


Figure 3.6 Scenario 12 under the closed loop and the predictive controller, from engine data. (a) The true index in the first 20 min against the shaded surgical band, under the closed loop (dash-dotted), the plan (thick solid) and the plan without learning (dashed). The plan stops the pump at 1.55 min and the index turns at 39.5, while the closed loop and the plan without learning fall to 27.7 and 33.9. (b) The learned sensitivity (solid) against the equivalent value (densely dashed) at which the population surface reads the patient's true index at the patient's true effect sites. The patient's sensitivity (dotted) stays at 0.50, and the estimate follows the equivalent to 1.26 under the incision.

3.7 Technology in Practice

The monitor, the pumps and the pressure measurement of Figure 3.1 stand, in an operating room, for a processed electroencephalogram monitor, syringe pumps that carry the pharmacokinetic models, and a blood pressure monitor. Figure 3.7 draws the four loops with that hardware in place of the schematic's symbols.

What Measures It

The depth loop reads a processed electroencephalogram. Its index is computed by the monitor's own algorithm from a stretch of signal, which is why Section 3.3.7 credits it with dynamics of its own. The chapter's monitor of Equation (3.6) stands for it with a first-order filter of $\tau_B = 0.42$ min, noise, and a positive offset of 34 index units during diathermy.

> **TECHNOLOGY** **Sensor: The Processed EEG Depth Monitor**
>
> A depth monitor records the electroencephalogram through electrodes on the head, the function of a Class II electroencephalograph in United States regulation [23], and reduces it to one number. The most common index analyzes the signal in the frequency and phase domains and reports a value from 0 to 100, with 40 to 60 generally taken as right for surgical anesthesia. Electrode placement alone can shift the reading by up to 9 %, and drugs such as ketamine and nitrous oxide give a wake-like signal in a sedated patient [24]. High muscle activity and interference from electrical devices can pass for brain activity and raise the index without being flagged as artifact [14]. On a trend that failure is an index climbing while nothing has changed in the patient.

The pressure loop reads mean arterial pressure. A basic monitoring standard requires that every patient under anesthesia have arterial blood pressure and heart rate determined and evaluated at least

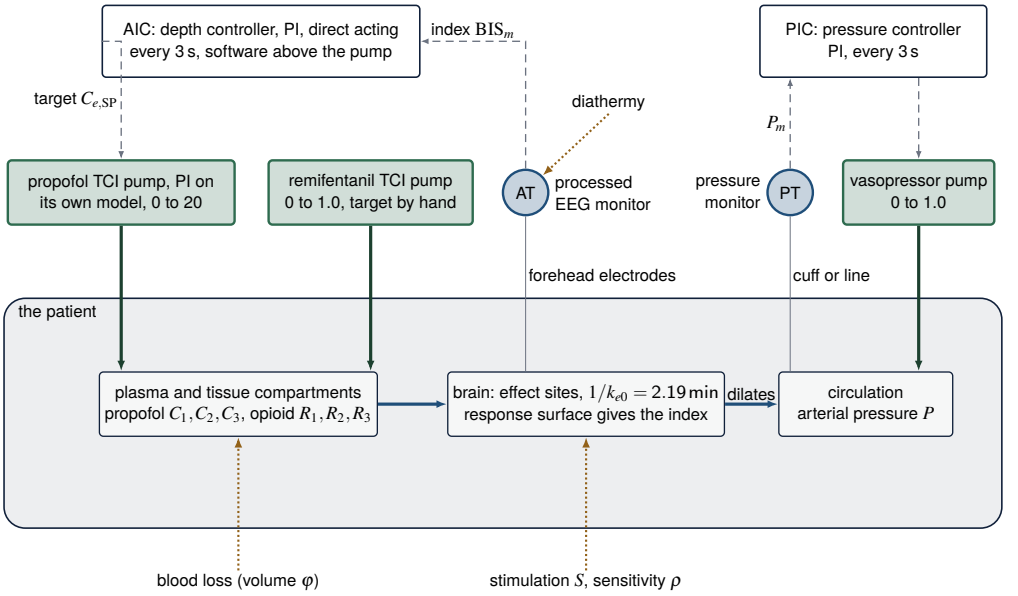


Figure 3.7 The four loops of Figure 3.3 with the hardware of an operating room. The processed EEG monitor (AT) reads the forehead and gives the index to the depth controller AIC, which writes the propofol effect-site target to a target-controlled infusion (TCI) pump. That pump closes its own loop on its internal model, which no disturbance reaches. The pressure monitor (PT) feeds the pressure controller PIC and the vasopressor pump. Pump ranges are those of Table 3.4, in mg/(kg h) for propofol and µg/(kg min) for the others. Thick solid arrows carry drug and its effect, thin dashed arrows signals, dotted arrows disturbances. Every pump fails stopped and none can remove drug.

every five minutes, and it lists a tracing of intra-arterial pressure among the ways to evaluate the circulation continually [25]. The chapter's pressure measurement, a first-order filter with no noise, is closer to a continuous arterial tracing than to a cuff read every few minutes.

What Moves It

Three syringe pumps deliver propofol, remifentanil and the vasopressor. The two anesthetic pumps run target-controlled infusion, in which the processor in the pump computes the rate from its model and the patient's covariates [3]. The inner concentration loops of Figure 3.3 are the chapter's version of that calculation. The fail position of every pump is stopped, and a stopped pump cannot remove drug, which is the one-sided authority of Section 3.4. Equation (3.5) models each pump as a first-order lag of $\tau_p = 0.12$ min with no bolus function and no dead volume.

TECHNOLOGY Final Element: The Target-Controlled Syringe Pump

In United States regulation an infusion pump moves fluid into a patient in a controlled manner by a piston, roller or peristaltic mechanism, may detect air in or blockage of the line and sound an alarm, and is a Class II device [26]. A target-controlled pump adds a microprocessor that holds a pharmacokinetic model. The clinician enters a target for the plasma or the effect site and the patient's covariates, and the processor computes the infusion, because neither concentration can be measured during the case. Such pumps alarm on high line pressure, a stopped infusion, an empty syringe, a power failure and a low battery, and some detect a disconnection from a drop in line pressure [3].

Where the Controller Runs

The inner loop runs inside the pump, on the processor that carries the model [3]. Research systems have placed the depth controller above such a pump: in 2002 a PID algorithm read the depth index and wrote the target of a propofol target-controlled infusion system [13]. Closed-loop target-controlled infusion on a pharmacodynamic endpoint such as the depth index or blood pressure is still experimental, and in routine practice the clinician watches the clinical response and the depth of anesthesia while the pump runs [3], which closes the outer loop by hand. The chapter's four controllers run every 3 s, as Section 3.5 describes.

IN PRACTICE Setting Up an Intravenous Anesthetic

A national audit named a poor grasp of the pharmacology and failed drug delivery as the two most common causes of awareness during total intravenous anesthesia. The setup therefore matters as much as the model. The drug goes into a large peripheral vein that stays visible, through a line free of dead space, kinks, air and leaks, with locking connectors, an anti-siphon valve against overdose from a damaged syringe, and an anti-reflux valve where several infusions share a line. The pressure cuff goes on another limb, a keep-vein-open mode is not used for propofol or remifentanil, and several pumps sit on the pole in a fixed order [3].

What Is Modernizing It

Three developments bear on this loop. Commercial pharmacokinetic models were built from selected populations that often left out children, obese and elderly patients. General-purpose models fitted to pooled data now cover far wider ranges: one propofol model rests on 15 433 concentration measurements from 1033 people, from 27 weeks postmenstrual age to 88 years and from 680 g to 160 kg [27]. Such a model still predicts a population mean, and no measurement before induction reveals where the patient on the table sits in the spread of Table 3.7. Closed loops are reaching beyond hypnosis: a closed-loop vasopressor controller studied in 40 patients after cardiac surgery reduced postoperative hypotension [3]. Research on artificial intelligence in anesthesia now centers on multimodal depth monitoring and automated drug infusion [24].

3.8 Limitations

The most important limitation is what the application is for. It is a teaching model of a control structure, and closed-loop control of anesthetic depth is not standard of care in any jurisdiction [17]. Nothing in the application has been validated against patient data, no result in Section 3.6 describes a person, and no configuration found here should be used to plan or guide the care of a patient. The pharmacokinetic models and the response surface are taken from published work, which makes the trajectories plausible and does not make them clinical. A reader who takes anything from this chapter into an operating room has misread it.

The technical limitations follow the assumptions of Section 3.3.7. The response surface has one endpoint, so the application cannot show a patient who is unconscious and moving, which is the situation the opioid and the neuromuscular blocking agents exist to handle. The circulation is one first-order state, so a pressure problem is always a vasodilation problem and never a volume or

a rhythm problem. The depth index is a deterministic function of two concentrations plus noise, whereas real indices differ between patients at the same concentration and occasionally move in the wrong direction. Recovery and respiratory depression are outside the model, which is exactly where a heavy opioid technique presents its bill, so scenario 7 can price the split in millimeters of mercury and cannot price it in minutes of apnea. The pump is a first-order lag with no bolus function and no mechanical dead volume. Repeatability is itself unrealistic. A fixed integration step and a fixed controller period give the same trajectory on every machine, and neither a real pump nor a real monitor behaves that way.

Two elements are deliberately stylized. Surgical stimulation arrives as a step and leaves as a step, which no operation does, although a step is the one disturbance shape that can be read off a trend without fitting anything. A patient sensitivity that the operator can set on a slider does not exist in an operating room either, although it is the only way to show the quantity the outer loop is estimating while the estimate is being made.

Everything that survived the simplification survived for one reason, which is that a control engineer has to answer for it. What survived is a manipulated variable that acts through two lags in series, a plant gain uncertain by a factor of three before the case starts, a multiplicative disturbance that moves the gain and the operating point together, an inner measurement that is a computation, a measurement corrupted in the direction that provokes an overdose, an actuator with authority in one direction, and a constraint that lives on a different loop from the setpoint under discussion.

The predictive controller's surface is the population one, so its single learned parameter absorbs everything the surface leaves out. Under the incision it reads 1.26 on a patient at 0.50, and on a resistant patient under stimulation the equivalent value of 4.29 lies beyond the bound of 2, so the estimate cannot be used to infer the patient. Each soft constraint has one slack for the whole horizon, the pressure row is a static slope with no circulatory lag or vasopressor, and the opioid is not a decision variable. The comparison is between control structures on a teaching simulator and makes no clinical claim.

> ⚠ **OPEN CHALLENGE** **One Number for Several Endpoints**
>
> A processed index reflects only statistical features of cortical electrical activity, while consciousness arises from networks that reach below the cortex. Indices respond differently to different drugs, proprietary algorithms make monitors disagree, and neuromuscular blockade and the balance of analgesia also move the reading. An index is therefore not a measure of overall adequacy when hypnosis, analgesia and paralysis change together [24]. The single response surface of Section 3.3.7 makes the same simplification. Progress would be monitoring that joins the electroencephalogram with heart rate variability, the pulse wave, muscle activity and near-infrared spectroscopy [24], and a controller that can tell which endpoint has moved.

3.9 Conclusion

This chapter used depth of anesthesia under propofol and remifentanil as a laboratory for a control structure whose inner measurement is a calculation. The simulator integrates thirteen states, two three-compartment pharmacokinetic models with an effect site behind each, an arterial pressure state, a vasopressor effect site and three pump lags, and turns the two effect-site concentrations into a depth index through a response surface with one synergy coefficient. The pump carries its own copy of the pharmacokinetics, built from population parameters and shielded from every disturbance, and that copy is within 0.01 µg/mL of its target at forty minutes on a patient whose requirement is anywhere between 0.85 and 3.38 µg/mL.

Under model inversion alone the settled index moves across 18.3 units as patient sensitivity spans its range, and closing an outer loop on the depth monitor holds the same range of patients within 1.1 units by moving the commanded concentration from 1.90 to 5.03 µg/mL. Signal-quality validation changes the lowest true index during electrocautery from 31.8 to 48.9 and the delivered dose from 333 to 178 mg, and no tuning substitutes for it. At the end of a case the index falls to 35.5 a quarter

of a minute after the stimulation goes, before any change of infusion can reach the effect site, so a higher gain or a lower opioid target can shorten the dip, while lightening before the closing raises its bottom.

The graded challenge judges the structure its text names as well as the index, so a configuration that trusts the monitor and leaves the pressure to the operator, 11.3 min below 60 mmHg, fails it although its index passes. A predictive controller on the pump's own model with one learned potency holds the induction of a patient twice as sensitive as the model at a lowest index of 39.5 against 27.7, although without the learning it ties on index error (207 against 211 BIS·min). A fuller model would add a second endpoint to the response surface, a subcutaneous or intramuscular route with real transport delay, a signal quality index that degrades gradually, and a population of virtual patients rather than one covariate set with a slider. What the application already shows is the division of labor between a model and a measurement.

KEY POINTS

- The pump holds the drug level its average-patient model predicts, so it cannot be wrong about its target and cannot be right about the patient.
- Under model inversion alone the settled index spans 18.3 units across patient sensitivity, and the closed depth loop holds the same patients within 1.1 units by moving the commanded concentration.
- Signal-quality validation raises the lowest true index during electrocautery from 31.8 to 48.9, and no tuning substitutes for it.
- At the end of a case a higher gain or a lower opioid target shortens the dip, while lightening before the closing raises its bottom.

Where else this applies. A model-based estimate corrected by a late measurement returns in Volume 5, Chapter 10, and the one-way pump in Chapter 2.

References

[1] S. L. Shafer and K. M. Gregg, "Algorithms to rapidly achieve and maintain stable drug concentrations at the site of drug effect with a computer-controlled infusion pump," *Journal of Pharmacokinetics and Biopharmaceutics*, vol. 20, no. 2, pp. 147–169, 1992.

[2] M. M. R. F. Struys, T. De Smet, J. I. B. Glen, H. E. M. Vereecke, A. R. Absalom, and T. W. Schnider, "The history of target-controlled infusion," *Anesthesia & Analgesia*, vol. 122, no. 1, pp. 56–69, 2016.

[3] P. U. Bidkar, A. Dey, P. Chatterjee, R. Ramadurai, and J. J. Joy, "Target-controlled infusion – past, present, and future," *Journal of Anaesthesiology Clinical Pharmacology*, vol. 40, no. 3, pp. 371–380, 2024.

[4] T. W. Schnider, C. F. Minto, S. L. Shafer, P. L. Gambus, C. Andresen, D. B. Goodale, and E. J. Youngs, "The influence of age on propofol pharmacodynamics," *Anesthesiology*, vol. 90, no. 6, pp. 1502–1516, 1999.

[5] K. van Heusden, G. A. Dumont, K. Soltesz, C. L. Petersen, A. Umedaly, N. West, and J. M. Ansermino, "Design and clinical evaluation of robust PID control of propofol anesthesia in children," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 2, pp. 491–501, 2014.

[6] B. Joseph and C. B. Brosilow, "Inferential control of processes: Part I. steady state analysis and design," *AIChE Journal*, vol. 24, no. 3, pp. 485–492, 1978.

[7] D. E. Seborg, T. F. Edgar, D. A. Mellichamp, and F. J. Doyle, III, *Process Dynamics and Control*, 4th ed. Hoboken, NJ: John Wiley & Sons, 2016.

[8] R. G. Bickford, "Automatic electroencephalographic control of general anesthesia," *Electroencephalography and Clinical Neurophysiology*, vol. 2, no. 1–4, pp. 93–96, 1950.

[9] E. Krüger-Thiemer, "Continuous intravenous infusion and multicompartment accumulation," *European Journal of Pharmacology*, vol. 4, no. 3, pp. 317–324, 1968.

[10] T. W. Schnider, C. F. Minto, P. L. Gambus, C. Andresen, D. B. Goodale, S. L. Shafer, and E. J. Youngs, "The influence of method of administration and covariates on the pharmacokinetics of propofol in adult volunteers," *Anesthesiology*, vol. 88, no. 5, pp. 1170–1182, 1998.

[11] C. F. Minto, T. W. Schnider, T. D. Egan, E. Youngs, H. J. M. Lemmens, P. L. Gambus, V. Billard, J. F. Hoke, K. H. P. Moore, D. J. Hermann, K. T. Muir, J. W. Mandema, and S. L. Shafer, "Influence of age and gender on the pharmacokinetics and pharmacodynamics of remifentanil," *Anesthesiology*, vol. 86, no. 1, pp. 10–23, 1997.

[12] H. Schwilden, J. Schüttler, and H. Stoeckel, "Closed-loop feedback control of methohexital anesthesia by quantitative EEG analysis in humans," *Anesthesiology*, vol. 67, no. 3, pp. 341–347, 1987.

[13] A. R. Absalom, N. Sutcliffe, and G. N. Kenny, "Closed-loop control of anesthesia using bispectral index," *Anesthesiology*, vol. 96, no. 1, pp. 67–73, 2002.

[14] A. A. Dahaba, "Different conditions that could result in the bispectral index indicating an incorrect hypnotic state," *Anesthesia & Analgesia*, vol. 101, no. 3, pp. 765–773, 2005.

[15] P. S. Sebel, T. A. Bowdle, M. M. Ghoneim, I. J. Rampil, R. E. Padilla, T. J. Gan, and K. B. Domino, "The incidence of awareness during anesthesia: A multicenter United States study," *Anesthesia & Analgesia*, vol. 99, no. 3, pp. 833–839, 2004.

[16] B. A. Fritz, P. L. Kalarickal, H. R. Maybrier, M. R. Muench, D. Dearth, Y. Chen, K. E. Escallier, A. Ben Abdallah, N. Lin, and M. S. Avidan, "Intraoperative electroencephalogram suppression predicts postoperative delirium," *Anesthesia & Analgesia*, vol. 122, no. 1, pp. 234–242, 2016.

[17] G. A. Dumont and J. M. Ansermino, "Closed-loop control of anesthesia," *Anesthesia & Analgesia*, vol. 117, no. 5, pp. 1130–1138, 2013.

[18] C. F. Minto, T. W. Schnider, T. G. Short, K. M. Gregg, A. Gentilini, and S. L. Shafer, "Response surface model for anesthetic drug interactions," *Anesthesiology*, vol. 92, no. 6, pp. 1603–1616, 2000.

[19] Y. Sawaguchi, E. Furutani, G. Shirakami, M. Araki, and K. Fukuda, "A model-predictive hypnosis control system under total intravenous anesthesia," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 3, pp. 874–887, 2008.

[20] C. M. Ionescu, R. De Keyser, B. C. Torrico, T. De Smet, M. M. Struys, and J. E. Normey-Rico, "Robust predictive control strategy applied for propofol dosing using BIS as a controlled variable during anesthesia," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 9, pp. 2161–2170, 2008.

[21] E. Furutani, Y. Sawaguchi, G. Shirakami, M. Araki, and K. Fukuda, "A hypnosis control system using a model predictive controller with online identification of individual parameters," in *Proceedings of 2005 IEEE Conference on Control Applications (CCA 2005)*. Toronto, Canada: IEEE, 2005, pp. 154–159.

[22] G. Pannocchia and J. B. Rawlings, "Disturbance models for offset-free model-predictive control," *AIChE Journal*, vol. 49, no. 2, pp. 426–437, 2003.

[23] U.S. Food and Drug Administration, "Electroencephalograph," Code of Federal Regulations, Title 21, Section 882.1400, revised as of April 1, 2024, U.S. Government Publishing Office, govinfo.gov, 2024.

[24] H. Yu, B. Gao, and Y. Li, "Depth of anesthesia monitoring: Current evidence, clinical impact, and future directions," *Journal of International Medical Research*, vol. 54, no. 6, p. 03000605261457781, 2026.

[25] American Society of Anesthesiologists, "Standards for basic anesthetic monitoring," Standard of the American Society of Anesthesiologists, approved 1986, last amended October 15, 2025, asahq.org, 2025.

[26] U.S. Food and Drug Administration, "Infusion pump," Code of Federal Regulations, Title 21, Section 880.5725, revised as of April 1, 2024, U.S. Government Publishing Office, govinfo.gov, 2024.

[27] O. Vandemoortele, L. N. Hannivoort, F. Vanhoorebeeck, M. M. R. F. Struys, and H. E. M. Vereecke, "General purpose pharmacokinetic-pharmacodynamic models for target-controlled infusion of anaesthetic drugs: A narrative review," *Journal of Clinical Medicine*, vol. 11, no. 9, p. 2487, 2022.

3.10 Exercises

3.1 ○ Concept check · 2 min

In scenario 3 the target-controlled pump holds its model's propofol effect site within 0.01 µg/mL of the 1.7 µg/mL target, while the true index settles near 67, above the surgical band. Why does the propofol concentration loop not act on this error?

(a) Its gain is too low to remove an error that builds up over forty minutes

(b) Its measurement is the pump's own model, which no property of the patient reaches

(c) The monitor noise hides the error from the concentration controller

(d) Its integral is held at the pump's upper limit by the anti-windup

3.2 Concept check · 2 min

The incision of scenario 6 sets the surgical stimulation to $S = 0.85$. What does it do to the plant that the depth loop sees?

(a) It adds a fixed number of index units, so it acts like a load at the output

(b) It speeds the effect-site link k_{e0}, so the index answers the pump sooner

(c) It multiplies both half-maximal concentrations, so the operating point and the gain move together

(d) It shrinks the central volume, so the same infusion gives a higher plasma level

3.3 Concept check · 2 min

At the closing of scenario 11 the true index falls from 50.1 to 35.5 within a quarter of a minute. Which statement is right?

(a) A doubled depth loop gain would keep the index above 40, since the pump would be cut sooner

(b) Stopping the propofol pump at the instant of closing would keep the index above 40

(c) An opioid target of 8 instead of 4 ng/mL makes the dip shallower, since the opioid is short-acting

(d) The lowest index comes before an infusion change can reach the effect site, so only earlier lightening raises it

3.4 Theory and calculation · 5 min

The default patient is a man of 40 y, 70 kg and 170 cm. Use Equation (3.2) with $V_1 = 4.27\,\mathrm{L}$, $V_3 = 238\,\mathrm{L}$ and $\mathrm{Cl}_3 = 0.836\,\mathrm{L/min}$.

(a) Find the lean body mass L, the clearance Cl_1 and the rate constant k_{10}. *Checkpoint:* 55.3 kg, 1.638 L/min and 0.384/min.

(b) Find the ratio k_{13}/k_{31} and the time constant $1/k_{31}$, and say what the ratio means. *Checkpoint:* 55.7 and 285 min.

(c) Find the infusion that clearance alone needs to hold $C_1 = 1.7\,\mu\mathrm{g/mL}$ once every compartment is full. *Checkpoint:* 2.39 mg/(kg h).

3.5 Theory and calculation · 6 min

Use the response surface of Equation (3.3) with the defaults of Table 3.3: $C_{50,p} = 4.16\,\mu\mathrm{g/mL}$, $C_{50,r} = 19.0\,\mathrm{ng/mL}$, $\gamma = 1.43$, $\beta = 1.40$ and $E_0 = E_{\max} = 93$, with $\rho = \rho_r = 1$.

(a) Find the value of U that gives an index of 50, and the propofol effect site that gives it with no opioid. *Checkpoint:* 0.900 and 3.74 µg/mL.

(b) Find U_{50} at an even split, $\theta = 0.5$, and say how much less total normalized drug the pair needs than addition would predict. *Checkpoint:* 0.65, or 35 % less.

(c) The pair $C_{e,p} = 1.69\,\mu\mathrm{g/mL}$ and $C_{e,r} = 4\,\mathrm{ng/mL}$ holds an index of 50. Find the index at the same concentrations at $S = 0.85$, within 0.1. *Checkpoint:* 66.7.

3.6 Theory and calculation · 6 min

Scenario 3 runs target control on a patient with $\rho = 1.6$ and $\rho_r = 1$, with no stimulation. At 40 min the model holds 1.708 µg/mL of propofol and 2.00 ng/mL of remifentanil at the effect sites, and the patient's pharmacokinetics are those of the model.

(a) Find the true index from Equation (3.3), within 0.1. *Checkpoint:* 67.4.

(b) The population patient needs 2.67 µg/mL of propofol for an index of 50 at 2.00 ng/mL of opioid. Find what this patient needs, and its ratio to the target the pump holds. *Checkpoint:* 4.27 µg/mL and 2.5.

(c) The depth loop has $K_{c,B} = 0.045\,\mu g/mL$ per index unit. The process gain is -10.0 index units per µg/mL at 4 ng/mL of opioid and -18.8 at 8 ng/mL. Find the magnitude of the loop gain at each. *Checkpoint:* 0.45 and 0.85.

3.7 App challenge · 8 min

Load scenario 3, The same pump, a different patient. Use the default seed, run at speed 5 to 40 min and export the CSV file.

(a) Read the true index, column `bis`, and the model's propofol effect site, column `ce_model`, in the row at 40 min. *Checkpoint:* 67.4 within 0.2, and 1.708 µg/mL within 0.005.

(b) Press the reload button, set Patient sensitivity to propofol to 1.00 before Run, run to 40 min and read `bis` again. *Checkpoint:* 59.6 within 0.2.

(c) Load scenario 4, Closing the loop, run it as shipped to 39 min and read the propofol target, column `ce_target`, at 39 min. *Checkpoint:* 4.07 µg/mL within 0.02.

3.8 App challenge · 8 min

Load scenario 8, The measurement lies, upward. Use the default seed and run each case at speed 5 to 45 min, then export the CSV file.

(a) Run with Signal quality at Trust the monitor, as shipped. Read the lowest true index, column `bis`, the IAE tile and the propofol delivered, column `metric_prop_mg`, in the last row. *Checkpoint:* 31.8 within 0.2, IAE 367 within 3, and 333 mg within 2.

(b) Press Reset, set Signal quality to Hold on low quality before Run and run again. Read the same three values. *Checkpoint:* 48.9, 14 and 178 mg, with the same tolerances.

3.9 App challenge · 8 min

Load the challenge, A whole case. Use the default seed and run each case at speed 20 to 120 min. After each run read the challenge verdict (pass or fail and the worst index error from 79 min), the Light tile and the Hypotension tile.

(a) Run the case as shipped. *Checkpoint:* Fail, worst error 10.1, Light 99.7 min, Hypotension 0.0 min.

(b) Press the reload button, press Reveal suggested tuning and run. *Checkpoint:* Pass, worst error 3.3, Light 4.4 min, Hypotension 1.2 min.

(c) Press the reload button, press Reveal suggested tuning, then set Signal quality to Trust the monitor and Vasopressor to By hand before Run. *Checkpoint:* Fail, worst error 4.4, Light 4.4 min, Hypotension 11.3 min.

Accept the worst error within 0.2 and each tile within 0.3 min. Say what each run shows about structure against tuning.

3.10 Application area · 3 min

An index of 50 needs 3.74 µg/mL of propofol alone, or 1.69 µg/mL with 4 ng/mL of remifentanil. Take $a = 6.4$ mmHg per µg/mL and $b = 0.55$ mmHg per ng/mL from Equation (3.4).

(a) Find the steady difference in arterial pressure between the two splits with no stimulation. *Checkpoint:* About 10.9 mmHg higher with the opioid.

(b) Name two costs of an opioid-heavy split that the model does not price.

3.11 Application area · 4 min

A distillation column holds its top composition with a fast inner loop on a composition inferred from tray temperatures through a model fitted to past operation. A slow outer loop corrects the inferred value from an online analyzer.

(a) Name the counterpart in the anesthesia loop of the inference model, of the analyzer and of a change of feed that the model was not fitted to.

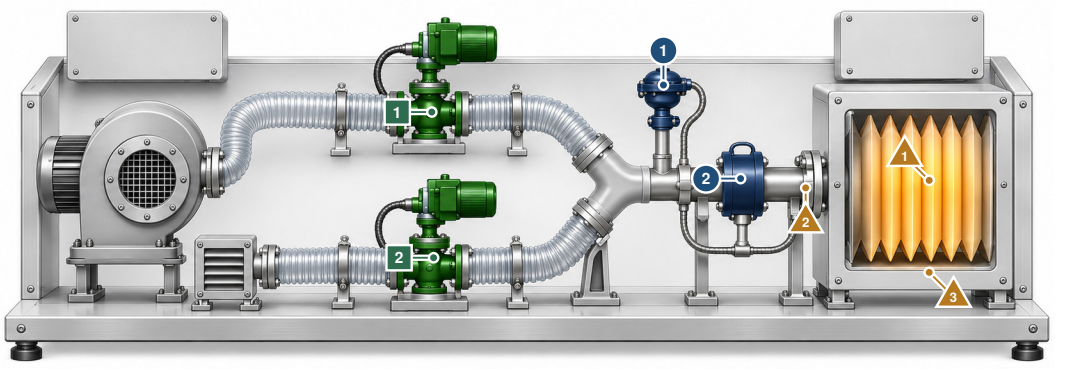
(b) The analyzer reads high for ten minutes during its own calibration. Say what the outer loop does with it and what the anesthesia chapter recommends.

3.12 Review the agent · 5 min

An AI assistant comments on scenario 11 as follows. “The index drops to 35.5 at the closing because the depth loop is too slow: it lowers the propofol target by only a fraction of a microgram per milliliter in the first minute. Double the depth loop gain to 0.090 µg/mL per index unit and the loop will cut the pump at once and keep the index inside the 40 to 60 band through the closing.” Find the flaw in the proposal and say how it could be checked.

CHAPTER 4

Mechanical Ventilator Simulator for Patient-Triggered Cycles and Breath-by-Breath Identification



The controller moves 1 inspiratory valve, lets gas in from the blower
2 expiratory valve, lets gas out and holds the end pressure

The controller reads 1 airway pressure at the wye 2 flow at the wye

What disturbs it 1 a lung that stiffens over the hour 2 a leak beyond the flow sensor, at the mask or cuff
3 the patient's own breathing effort

Both sensors sit between the valves and the lung, so a leak beyond them, a stiffer lung or the patient's own pull reaches the controller only through what they read. Illustration, not to scale.

IN PLAIN TERMS **A Breathing Machine That Sees Only the Tube**

The main idea. A ventilator pushes air into the lungs of a patient who cannot breathe well enough alone. It measures only the pressure and the flow where its tubes join the patient's tube. It shapes each breath from those two readings.

The challenge. A leak at the mask looks to the machine like air going into the lungs, and the patient's own breathing pulls on the same tube. The lungs can also stiffen within an hour, so a breath that was safe can grow too forceful or too small.

Why it matters. Too forceful a breath injures the lung, and too small a breath lets carbon dioxide build up in the blood.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	The first page, the fact box of Section 4.1, the first history box in Section 4.2, the try-it boxes, the second history box and the second fact box of Section 4.6, and the key points
Study	50 min	The whole chapter with its pause points and the twelve exercises
Practice	20 min	The first page, Section 4.6, Section 4.7, Section 4.8 and the open challenges

The control question A ventilator measures pressure and flow at the wye, while the leak, the patient's effort and the lung stay hidden. What does each mode hold, and when do the breath the machine reports and the breath the lung receives part company?

Learning objectives

- Compute the plateau of volume control and the volume of pressure control, and name the variable each mode lets go.
- Predict auto-PEEP from the expiratory time in time constants, and name the settings that change it.
- Explain what a leak past the flow sensor does to triggering, cycling and the measured volume, and find the largest leak the valve can hold PEEP against.
- Judge when a per-breath fit of resistance and compliance can be trusted, and what an effort does to it.
- Choose a mode, leak compensation and adaptive support that meet an hour-long target scored on the hidden truth.

Prerequisites Sections 1.4.2 (one compartment), 1.4.4 (identifying the individual), 1.4.6 (actuators that push one way), 1.6.2 (velocity-form PID) and 1.6.4 (scenarios and the challenge).

The app https://apmonitor.com/apps/mechanical-ventilator/

See also Volume 8, Chapter 7 for a cycle that hands one valve between two loops, and Volume 8, Chapter 15 for a hybrid cycle held one impact at a time. Volume 1, Chapter 12 for least squares on inputs that move together, and Volume 1, Chapter 11 for fitted coefficients that absorb what the model leaves out.

4.1 Introduction

Mechanical ventilation is the life support around which modern intensive care units grew [1], and the way a ventilator is set changes survival. A trial stopped after 861 patients with acute lung injury or the acute respiratory distress syndrome (ARDS) gave one arm breaths of 6 mL/kg of predicted body weight with the plateau pressure held at or below 30 cmH_2O, the other 12 mL/kg with a plateau of up to 50 cmH_2O, and mortality was 31.0 % in the first arm against 39.8 % in the second [2]. Pooled data from 3562 patients in nine trials later found the driving pressure, the tidal volume divided by the compliance of the respiratory system, more closely tied to survival than the volume or the PEEP alone [3]. Both findings concern the breath the lung receives, while a ventilator controls the breath its sensors report.

> ✦ **DID YOU KNOW? Ventilators at Home**
>
> By about the year 2000, mechanical ventilation at home was in wide clinical use for chronic respiratory failure, at a prevalence of 6.6 per 100 000 people, with marked variation between countries [4].

The ventilator's view of that breath is narrow. It measures pressure and flow at the wye, where its two limbs join the patient's tube, so a leak past the sensor counts as flow into the lung, the patient's own inspiratory effort is a pressure nobody measures, and the resistance and compliance of the lung are unknown and, in the simulator's cases, halve or double within an hour. A common reading of the screen takes its tidal volume for the lung's, and this simulator shows where that reading fails. In its hour-long challenge, run as shipped in volume control, the monitor reads 418 mL at minute 45 while the lung receives 147 mL, the identified compliance reads 70.6 against a true 31.25 mL/cmH_2O, and the end-tidal CO_2 has risen from 35.5 to 79.3 mmHg, because a 14 L/min leak opened at minute 30 trips the flow trigger 0.44 to 0.48 s into every expiration between minutes 31 and 39. Two neighboring beliefs fail as plainly. Volume control keeps 420 mL going into a lung stiffened to 18 mL/cmH_2O only by letting the plateau climb to between 31.3 and 31.5 cmH_2O, past its 30 cmH_2O limit. Per-breath mechanics are estimates rather than measurements, and a 5 cmH_2O effort pulls the fitted resistance of every breath the patient triggers 21.6 to 21.8 % low and the compliance 10.8 to 12.0 % high on a lung that, relaxed, was identified within 0.5 and 0.2 %.

The class of problem is a hybrid one. Inside each phase of a breath the gas flows continuously, and at each phase boundary a discrete decision is taken, some on a clock and some when a measured flow or pressure crosses a threshold, which is the mixture of flows and jumps that the theory of hybrid dynamical systems treats in general [5]. The taxonomy of ventilator modes uses the same terms, defining a breath by the events that start and stop inspiration, by whether the patient or the machine initiates them, and by the control variable that the equation of motion leaves free [6]. An injection-molding machine that hands its screw from a velocity servo to a pressure servo at a transfer position (Volume 8, Chapter 7), a walking robot whose swing ends at ground contact, and an artificial pancreas dosing around meals (Chapter 2) share that shape, with a cycle whose events the plant partly decides, an estimate refitted every cycle and a limit that counts on every cycle rather than on average. A transfer function between valve opening and airway pressure carries none of it.

The engineering literature reports ventilator controllers on physical lung models, in simulation and in small feasibility studies [7, 8], and the clinical literature finds asynchrony by inspecting recorded pressure and flow [9]. Neither shows the hidden effort, alveolar pressure and leak beside what the machine measured on the same breath. This chapter tells apart, on a simulated ventilator whose patient is hidden from its controller, the breath the lung actually receives and the breath the machine reports, mode by mode, threshold by threshold and estimate by estimate. The modes, the fit, the leak estimate, the adaptive support and the caps all follow the clinical and engineering literature. The chapter gives every equation, constant and rule of the simulator, whose controller reads only two noisy sensors at the wye while an audit classifies every patient effort against the truth. Its nine cases and an hour-long challenge each isolate one way in which the measured and delivered breaths part. All of the application's limits, alarm thresholds and patient values are teaching values. Real ones come from trials such as the ARDS Network's, from the particular standard for critical care ventilators [10], from the asynchrony literature and from the clinician at the bedside, and neither the simulator nor this text offers clinical guidance.

Section 4.2 reviews how ventilation came to be controlled and which variable each generation chose to hold. Section 4.3 gives the lung, the effort, the circuit with its valves and the carbon dioxide store, with the assumptions behind them. Section 4.4 derives four closed forms, for the plateau and volume of a mode, the trapped pressure of an expiration cut short, the edge of the actuator against a leak and the bias that an effort puts into the fit, and checks each against the engine. Section 4.5 describes the breath sequencer, the three valve servos, the per-breath least-squares fit, the volume loop with its protective caps and the alarms. Section 4.6 runs the nine cases and the challenge, Section 4.7 describes the real sensors, valves and controllers, and Sections 4.8 and 4.9 state the limitations and the conclusions.

4.2 Breathing Machines and the Patient Behind Them

Machines breathed through the chest wall before they breathed through the airway. Negative-pressure ventilators appeared in the early 1800s, positive-pressure devices around 1900, and the intensive care ventilator of today began to be developed in the 1940s and has passed through four distinct generations since [11]. The tank respirator of 1929 and the turn to positive pressure in the poliomyelitis epidemics are told in the first history box of this section, and the synchrony problem its authors reported [12] in the second, at the case of Section 4.6 where a patient breathes.

The positive-pressure modes that followed differ in which variable they hold, and the taxonomy now used to classify them starts from the equation of motion, defines mandatory and spontaneous breaths by whether the ventilator or the patient starts and stops them, and combines control variables and breath sequences with targeting schemes [6]. Pressure support, a pressure assist that augments each spontaneous effort with a clinician-selected airway pressure, was studied in 1986 on 15 stable patients and on a mechanical model of them, and was found to improve comfort and reduce the patient's work [13]. Proportional assist went further in 1992 by applying pressure according to the equation of motion so that the ventilator amplifies the patient's instantaneous effort while the

patient keeps control of the volume and both durations [14]. Closed loops on the ventilator's own settings followed. An adaptive lung ventilation controller of 1994 adjusted the rate and the inspiratory pressure toward a target alveolar ventilation from measured lung mechanics and dead space, and on a physical lung model with restrictive, normal and obstructive settings it answered steps in 48 to 81 s with 5.5 to 7.9 % overshoot [7], and a review of 2011 expected closed-loop control on most aspects of ventilatory support [11]. The particular standard for critical care ventilators, current in the United Kingdom as BS EN ISO 80601-2-12:2023, covers their basic safety and essential performance, and it counts a ventilator as a physiologic closed-loop system only when a physiological patient variable adjusts the therapy settings [10].

> ⧖ HISTORY **From the Tank to Positive Pressure (1929 to 1952)**
>
> By 1928 rescue squads in the United States and Canada had discarded the resuscitators that forced air in and out through a valved face mask. The tank respirator described in 1929 sealed the body below the neck in a sheet-metal cylinder, while the head stayed at room pressure. Two blowers and a rotating valve, driven by a variable-speed motor, lowered and raised the pressure around the body, up to about 60 cmH_2O at 10 to 40 breaths a minute. A child paralyzed by poliomyelitis was kept in it almost continuously for 122 hours [12]. In the polio epidemics tanks were built in large numbers. From 1949 a valve that added positive pressure during inspiration raised survival from 21 to 84 %, and in Copenhagen in 1952 positive pressure given by hand with a bag through a tracheostomy substantially reduced deaths from acute polio [4].

The question of which variable to hold became a question of injury. In mechanically ventilated rats, high airway pressure at a normal tidal volume left the lungs like those of controls while a high tidal volume produced permeability edema even at low airway pressure, and 10 cmH_2O of PEEP markedly reduced it [15]. The ARDS Network trial turned the volume into the setting to limit, measuring the plateau after a 0.5 s pause at the end of inspiration [2], the driving pressure then emerged as the variable that best stratified risk [3], and lung-protective strategies built on that understanding of ventilator-induced injury have markedly improved outcomes [1].

The patient's side of the cycle has a literature of its own. In 62 patients who triggered every breath, 15 had an asynchrony index above 10 % of their efforts, ineffective triggering and double triggering were the main patterns, and patients with frequent asynchrony stayed on the ventilator for a median of 25.5 days against 7.5 [9]. A mathematical model of pressure support showed that the flow at the end of the patient's neural inspiration, relative to the peak flow, depends mostly on the ratio of the respiratory time constant to the neural inspiratory time, so that any single fixed flow criterion for ending a breath ends some breaths early and others late [16]. Auto-PEEP, the trapped pressure of an expiration cut short, was brought to clinical attention about 30 years before a review of 2011 listed its effects on hemodynamics, work of breathing, synchrony and the monitoring of mechanics [17]. The mechanics themselves can be fitted by least squares. In paralyzed patients the fit came close to the occlusion values, while under pressure support with a low support level and strong muscle activity it underestimated the resistance by 10.4 $cmH_2O \cdot s/L$ and overestimated the compliance by 35.2 mL/cmH_2O on average [18]. The control-engineering literature has since posed the pressure servo as an adaptive problem, tracking a time-varying target pressure for unknown patient, hose and leak parameters in the presence of effort, with a proof of stability [8]. None of these sources can set the true lung, effort and leak beside each breath the machine reported, and that pairing is what the simulator keeps.

4.3 A Lung Behind a Wye

The simulated ventilator is a blower held at P_b behind a proportional inspiratory valve, a circuit of compliance C_t whose pressure P is read at the wye, and an expiratory valve that vents the circuit to the room, as shown in Figure 4.1. Beyond the wye a leak, an airway resistance R and a single compartment of compliance C make up the patient, whose respiratory muscles add a pressure $P_{mus}(t)$.

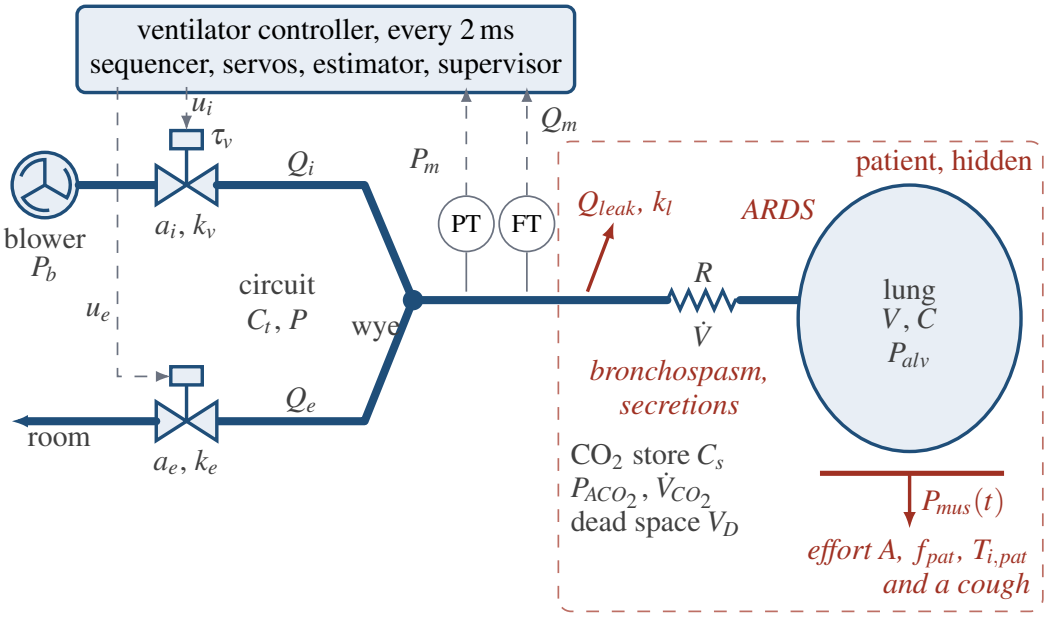


Figure 4.1 The ventilator circuit and the patient behind the wye. The blower holds P_b behind the inspiratory valve of opening a_i and coefficient k_v, the circuit of compliance C_t at pressure P joins the expiratory valve (a_e, k_e) at the wye, and each valve follows its command u_i or u_e through the lag τ_v. The pressure and flow transmitters PT and FT at the wye supply the only measurements P_m and Q_m. Everything in the dashed box is hidden from the controller: the leak Q_{leak} of coefficient k_l past the sensor, the airway resistance R carrying $\dot{V}$, the lung of volume V, compliance C and alveolar pressure P_{alv}, the muscle pressure P_{mus} set by the effort A, rate f_{pat} and neural time $T_{i,pat}$, and the CO_2 store. The disturbances, drawn in red with their text labels in italics, are the leak, bronchospasm and secretions on R, the ARDS stiffening of C, and the effort with its cough.

The five integrated states are the lung volume V above its relaxed volume, the circuit pressure P, the two valve openings a_i and a_e, and the alveolar CO_2 partial pressure P_{ACO_2}. The derivative function stays pure, and everything discrete, the effort schedule, the disease ramps, the breath record and the audit, moves only between integration steps. The signals are classified in Table 4.1.

4.3.1 The Patient

The lung is a single compartment, as in Section 1.4.2, holding gas behind a resistance. Its equation of motion includes the patient's effort, as shown in Equation (4.1).

$$P_{alv} = \frac{V}{C} - P_{mus}(t), \qquad \dot{V} = \frac{P - P_{alv}}{R} \tag{4.1}$$

where P_{alv} is the alveolar pressure and $\dot{V}$ the flow into the lung, so that at the airway $P = R\dot{V} + V/C - P_{mus}$. The shipped patient has $R = 10\,\mathrm{cmH_2O \cdot s/L}$ and $C = 50\,\mathrm{mL/cmH_2O}$, a time constant RC of 0.5 s, and both are the slider values times multipliers that the disease events ramp linearly, the compliance to half over 60 min for ARDS and the resistance to double over 60 s for bronchospasm. Secretions multiply R by $\exp(\sigma_s x_k)$, where x_k is a seeded first-order random sequence with a 3 s time constant updated every 10 ms and σ_s is the slider as a fraction. The effort is one half-sine per breath

Table 4.1 Signals of the Ventilator and the Patient, With Their Defaults

Symbol	Role	Default	Range
V	state, lung volume above relaxation	C·PEEP	—
P	state, circuit pressure	PEEP	—
a_i, a_e	states, valve openings	pass Q_{bias}	0–1
P_{ACO_2}	state, alveolar CO_2	40 mmHg	—
u_i, u_e	MVs, valve commands	—	0–1
P_m	measured, CV of the pressure and PEEP servos	—	—
Q_m	measured, CV of the flow servo	—	—
$\hat{V}_T$	estimated, CV of the volume loop	—	—
ΔP	MV of the volume loop, support	10 cmH_2O	0–40
C	DV, compliance (ARDS ramp)	50 mL/cmH_2O	10–100
R	DV, resistance (bronchospasm)	10 $cmH_2O \cdot$ s/L	3–60
Q_{leak}	DV, leak at 10 cmH_2O	0	0–400 L/min
A	DV, peak effort	0	0–15 cmH_2O
f_{pat}, $T_{i,pat}$	DV, patient's rate, neural time	18 min^{-1}, 1.0 s	6–40, 0.4–2
—	DV, secretions (noise on R)	0 %	0–50
P_{alv}, P_{mus}	hidden truth	—	—

the patient takes, as shown in Equation (4.2).

$$\begin{aligned} P_{mus}(t) &= A \sin\left(\frac{\pi(t-t_k)}{T_{i,pat}}\right), \qquad t_k \le t < t_k + T_{i,pat} \\ t_{k+1} &= t_k + \frac{60}{f_{pat}}\,(1+\varepsilon_k) \end{aligned} \tag{4.2}$$

where A is the peak effort, $T_{i,pat}$ the neural inspiratory time, capped at 80 % of the period, f_{pat} the patient's rate per minute and ε_k a seeded uniform jitter of ± 6 %. A change of effort, rate or neural time applies from the next effort. A cough is an expiratory half-sine of 40 cmH_2O over 0.4 s, subtracted from P_{mus}.

4.3.2 The Circuit and Its Valves

The circuit between the valves and the wye stores gas in its compliance, and the valves and the leak are orifices, as shown in Equation (4.3).

$$\begin{aligned} C_t \dot{P} &= Q_i - Q_e - Q_{leak} - \dot{V} \\ Q_i &= a_i k_v\, \psi(P_b - P), \qquad Q_e = a_e k_e\, \psi(P) \\ Q_{leak} &= k_l\, \psi(P), \qquad \tau_v \dot{a} = u - a \\ \psi(\Delta p) &= \frac{\Delta p}{\sqrt{|\Delta p| + P_L}} \end{aligned} \tag{4.3}$$

where ψ is the square-root orifice law smoothed below $P_L = 0.1$ cmH_2O so that its slope stays finite at zero, Q_i is zero when $P \ge P_b$ and Q_e zero when $P \le 0$, so neither valve passes gas backward, and both openings are clipped to $[0, 1]$ after each step. Each coefficient is set from a stated flow through the same ψ, so the inspiratory valve passes exactly 180 L/min fully open into an open airway and the expiratory valve 120 L/min fully open at 5 cmH_2O, and the leak is entered as its flow at 10 cmH_2O. A disconnect adds an orifice of 400 L/min at 1 cmH_2O. The leak sits on the patient's side of the flow sensor, so the sensor reads $\dot{V} + Q_{leak}$.

Table 4.2 Model Constants. Class V Is a Slider in the Show Model Dialog and C a Constant in the Code. Nothing in This Model Is Flagged Hidden, but the Patient's Values Never Reach the Controller

Symbol	Meaning	Value	Class
P_b	blower pressure	60 cmH_2O (30–80)	V
$Q_{i,max}$	inspiratory valve fully open, open airway	180 L/min	V
$Q_{e,5}$	expiratory valve fully open at 5 cmH_2O	120 L/min	V
C_t	circuit compliance	1.5 mL/cmH_2O (0.3–4)	V
τ_v	valve response time	5 ms (1–30)	V
Q_{bias}	base flow above the flow trigger	1.5 L/min	V
T_{lock}	trigger lockout after inspiration	0.25 s	V
$T_{i,max}$	longest pressure-support breath	2.5 s	V
PBW	predicted body weight	70 kg	V
V_D	dead space	150 mL	V
$\dot{V}_{CO_2}$, C_s	CO_2 production, store	200 mL/min, 45 mL/mmHg	V
σ_P, σ_Q	sensor noise (1σ)	0.05 cmH_2O, 0.3 L/min	V
P_L	orifice smoothing pressure	0.1 cmH_2O	C
K	alveolar CO_2 constant	0.863	C
—	effort jitter, smallest effort	±6 %, 0.5 cmH_2O	C
—	cough	40 cmH_2O, 0.4 s	C
—	disconnect orifice	400 L/min at 1 cmH_2O	C

4.3.3 Carbon Dioxide

Carbon dioxide is one store, a slow outcome signal rather than a model of gas exchange, as shown in Equation (4.4).

$$C_s \dot{P}_{ACO_2} = \dot{V}_{CO_2} - \dot{V}_A \frac{P_{ACO_2}}{K}, \qquad \dot{V}_A = \frac{(V_T - V_D)^+}{T_b} \tag{4.4}$$

where $K = 0.863$ mmHg per (mL/min)/(L/min), V_T is the true tidal volume of the last finished breath, T_b its duration and V_D the dead space, and $\dot{V}_A$ falls to zero after 15 s without a breath. With the defaults of Table 4.2, 420 mL at 16 min^{-1} settles at 39.95 mmHg, and apnea raises the CO_2 by 4.44 mmHg a minute.

4.3.4 Initial State and Numerics

Every case starts with the lung at rest at PEEP, $V = C \cdot \text{PEEP}$ and $P = \text{PEEP}$, the valves at the openings that pass Q_{bias} in automatic mode or at the operator's stations in manual mode, and the CO_2 at 40 mmHg. The engine of Section 1.6.1 carries the five states forward with a fourth-order Runge-Kutta step of 1 ms. The circuit is the stiff part. With both valves shut its time constant $C_t R$ is 15 ms on the shipped patient, and with the expiratory valve wide open at 1 cmH_2O it falls to 2.6 ms, so before each step the integrator computes the node's conductance G and takes $\min(48, \lceil \Delta t\, G / 0.8 C_t \rceil)$ sub-steps, almost always one. The controller and sensors run every 2 ms, a historian row is written every 0.2 s, enough for trends that change once a breath, and the scene draws its waveforms from a 100 Hz buffer of its own. The patient's jitter, the secretions and the sensor noise come from the engine's generators, seeded at 20261007. Halving the step moves the true tidal volume and plateau of the passive cases by at most 0.24 % and the tidal volume of the leak case by 0.50 %, and the application passes its 37 verification checks.

4.3.5 Assumptions

One compartment. The lungs and chest wall are a linear resistance and a linear compliance. Real lungs are many compartments with their own time constants, recruit and derecruit with PEEP and stiffen at high volume, so the passive pressure-volume loop here is straight where a real one bends.

A scripted patient. The effort is a half-sine with jitter and no chemical drive. The patient does not breathe harder when the CO_2 rises or less when the support is generous, and coughs are events.

An idealized machine. The blower holds its pressure at any flow, the valve laws are exact and known to the controller, the circuit has compliance but no resistance, and the only sensors are at the

wye, with Gaussian noise and no drift. There is no sensor at the valves, no compensation for tubing compliance and no trigger filter beyond a 10 ms lag.

An orifice leak. The leak coefficient is fixed, while a real mask leak changes with pressure, posture and fit.

Carbon dioxide only. One store with fixed production and dead space stands in for gas exchange, and oxygenation is not modeled.

A truth audit. Efforts are classified against the hidden P_{mus}, which a ventilator could only infer from its waveforms or from an esophageal balloon, using the classes of the asynchrony literature [9] with simple timing rules.

Teaching values. The 6 mL/kg target and the 30 cmH_2O plateau follow the ARDS Network values [2], the application draws a 15 cmH_2O driving-pressure line that its documentation attributes to the driving-pressure analysis [3], and the 40 cmH_2O high-pressure limit, the alarm settings and the challenge criteria are chosen for teaching. Real settings and alarm limits come from the unit's protocol, the clinician and the device's conformity to its standard [10].

4.4 What the Equations Promise

Four closed forms describe what a mode can hold, and each is checked against the engine.

4.4.1 Holding the Volume or the Pressure

Volume control imposes the flow and leaves the pressure free, so after a pause long enough for the flow to stop the plateau is the recoil of the delivered volume, and pressure control imposes the pressure and leaves the volume to the lung's time constant, as shown in Equation (4.5).

$$P_{plat} = \mathrm{PEEP} + \frac{V_T}{C}, \qquad V_T = C\,\Delta P\left(1 - \mathrm{e}^{-(T_i - T_r/2)/RC}\right) \tag{4.5}$$

where ΔP is the pressure above PEEP, T_i the inspiratory time and T_r the rise time, half of which the half-cosine rise of Section 4.5 costs the lung. At a PEEP of 8 cmH_2O and 420 mL, the plateau is 16.40 cmH_2O at 50 mL/cmH_2O and 31.33 cmH_2O at 18 mL/cmH_2O, and the engine gives 16.4 to 16.5 and 31.3 to 31.5 cmH_2O. At 10 cmH_2O for 1 s with a 0.1 s rise, the volume is 425.2 mL at 50 mL/cmH_2O and 179.1 mL at 18 mL/cmH_2O, against 424 and 179 mL on the engine. A mode is a choice of which side of the first expression the machine guarantees.

🕮 **Pause.** Try Exercise 4.1 before reading on.

4.4.2 Expiration That Does Not Finish

A passive lung empties toward PEEP with the time constant RC. When each expiration lasts T_e, volume control adds V_T per breath and the trapped volume settles where each expiration removes what each breath adds, as shown in Equation (4.6).

$$\mathrm{PEEP}_i = \frac{V_T}{C}\,\frac{\mathrm{e}^{-x}}{1-\mathrm{e}^{-x}}, \qquad x = \frac{T_e}{RC} \tag{4.6}$$

In case 6 after the spasm, $R = 20\,cmH_2O\cdot s/L$ and $C = 50\,mL/cmH_2O$ give $RC = 1.00$ s, and 1.50 s of expiration gives 2.42 cmH_2O for the volume the engine delivers in this case, slightly above the 420 mL setting. The engine measures 2.58, and with the rate of case 6 swept from 34 down to 10 min^{-1}, Figure 4.2(a), it stays 5 to 8 % above the closed form at every x from 0.77 to 5. The difference is the start of each expiration, since a fit of the alveolar decay over the rest of it returns a time constant of 1.000 s, RC itself. Volume control without a pause ends inspiration at the 24.4 cmH_2O peak, and the airway takes about 0.12 s to come down to PEEP while the expiratory servo discharges the circuit, an excess of 0.54 $cmH_2O\cdot s$ that slows the first part of every expiration and leaves 0.16 cmH_2O more trapped at the case's rate. Only x moves the trapped pressure. PEEP does not appear in Equation (4.6).

🕮 **Pause.** Try Exercise 4.2 before reading on.

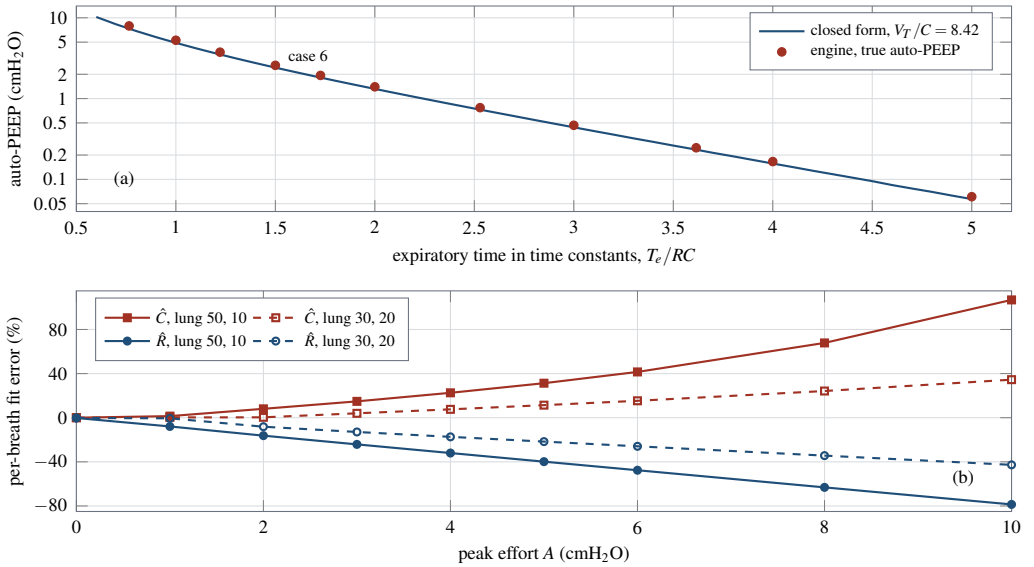


Figure 4.2 Two characteristics of the plant, measured on the engine. (a) True auto-PEEP in volume control in case 6 after the spasm ($R = 20\,\mathrm{cmH_2O \cdot s/L}$, $C = 50\,\mathrm{mL/cmH_2O}$, 1.0 s inspiration, no pause), with the rate set from 10 to 34 $\mathrm{min^{-1}}$, against the expiratory time in time constants, beside Equation (4.6). (b) Error of the per-breath least-squares resistance and compliance against the peak of the patient's effort, over 90 s of assist-controlled volume breaths at 16 $\mathrm{min^{-1}}$ with efforts at 18 $\mathrm{min^{-1}}$, in a model run of case 7 without its scripted events, on the shipped lung (50 $\mathrm{mL/cmH_2O}$, 10 $\mathrm{cmH_2O \cdot s/L}$) and on the lung case 7 ends with (30 $\mathrm{mL/cmH_2O}$, 20 $\mathrm{cmH_2O \cdot s/L}$).

4.4.3 What the Wye Sees

The flow sensor reads $Q_m = \dot{V} + Q_{leak}$ plus noise. A flow trigger fires when this flow, less the leak the estimator believes in, passes its threshold after the lockout, and pressure support ends a breath when the flow falls to a fraction of its peak. A leak that nobody has subtracted starts breaths and keeps them from ending. With 14 L/min of leak at a PEEP of 10 $\mathrm{cmH_2O}$, the filtered wye flow passes a 2 L/min trigger once the lung's own expiratory flow has fallen to 10.4 to 11.1 L/min, which between minutes 31 and 39 of the challenge comes 0.44 to 0.48 s into each expiration, 1.23 to 1.35 time constants of its lung. An effort enters the same flow, so a cycle-off fraction reads the sum of the support and the patient's pull, and the time constant relative to the neural time decides which breaths a fixed fraction ends early [16].

4.4.4 The Edge of the Actuator

Only the inspiratory valve can add gas, a one-sided actuator in the sense of Section 1.4.6, so against a leak the circuit pressure cannot exceed the point where the fully open valve's flow equals the leak's, $k_v \psi(P_b - P) = k_l \psi(P)$. At a PEEP of 8 $\mathrm{cmH_2O}$ the fully open valve passes 168 L/min, while a leak of 300 L/min at 10 $\mathrm{cmH_2O}$ would take 268 L/min, so the pressure falls until the two balance at 3.46 $\mathrm{cmH_2O}$. The same balance at 8 $\mathrm{cmH_2O}$ allows a leak of 187.6 L/min, and the engine holds PEEP to within 0.02 $\mathrm{cmH_2O}$ up to 187 L/min and lets it sag from 188 L/min. Ventilation is lost much earlier. With pressure control at 10 $\mathrm{cmH_2O}$ above PEEP, a 150 L/min leak leaves a breath of 123 mL at a peak airway pressure of 11.1 $\mathrm{cmH_2O}$, and 180 L/min leaves 21 mL.

4.4.5 Effort as an Unmodeled Input

The least-squares fit of Section 4.5 explains the measured pressure with R, $1/C$ and a starting pressure, and the effort is a pressure it cannot see. A patient who triggers each breath puts the effort into the

Table 4.3 Controller Settings, Defaults and Ranges

Symbol	Setting	Default	Range
—	mode	VC	VC, PC, PS
f	rate, backup rate in PS	16 min^{-1}	4–40
T_i, T_p	inspiratory time, pause (VC)	1.0 s, 0.2 s	0.3–3, 0–1
V_T^*	tidal volume, target	420 mL	100–900
ΔP	pressure above PEEP (PC, PS)	10 cmH_2O	0–40
PEEP	end-expiratory pressure	5 cmH_2O	0–20
$\dot{V}_{trig}$	flow trigger	2 L/min	0.5–15
ΔP_{trig}	pressure trigger, below PEEP	2 cmH_2O	0.5–10
f_{cyc}	cycle-off, of peak flow	25 %	5–70
T_r	rise time	0.1 s	0.02–0.5
—	VC profile	square	or decelerating
$K_{c,P}$, $\tau_{I,P}$	pressure servo	4, 60 ms	0.1–20, 2–500
$K_{c,Q}$, $\tau_{I,Q}$	flow servo	2, 30 ms	0.05–6, 2–500
$K_{c,E}$, $\tau_{I,E}$	PEEP servo	10, 30 ms	0.5–50, 2–500
—	leak compensation, adaptive	off, off	on, off
K_V	volume step, of the error	0.6	0.1–1.5
—	trigger, learn from machine breaths only	flow, off	pressure, on
—	protective layer	off	on, off
$P_{plat,max}$	plateau cap	30 cmH_2O	20–40
$V_{T,max}$	volume cap, per kg	8 mL/kg	4–12
P_{high}	high-pressure limit	40 cmH_2O	20–60
α	weight of the newest fit	0.5	0.05–1

Servo gains in % of valve opening per cmH_2O, or per L/min for the flow servo.

same phase of every breath, so the error is systematic rather than random, as shown in Figure 4.2(b), a counterfactual model run of case 7's settings with its scripted events removed. On the shipped lung each 1 cmH_2O of effort costs about 8 percentage points of resistance, from -0.24 to -78.6 % at 10 cmH_2O, while the compliance error grows faster than linearly to $+107$ %. On the lung case 7 ends with, the same 5 cmH_2O effort gives -21.65 and $+11.40$ %, because there a 420 mL breath needs 14.0 cmH_2O of elastic pressure against 8.4 and about twice the resistive pressure, so the effort is a smaller share of what the fit explains. On that lung a 1 cmH_2O effort never triggers, lands at random phases and leaves the fit within -4.6 to $+3.0$ % of the resistance. The same dependence on the effort's share of the pressure appeared in patients, whose fitted mechanics became satisfactory once a higher support level had reduced their muscle activity [18]. The fit faces the problem of dividing credit between inputs that move together, which Volume 1, Chapter 12 treats on recorded plant data.

🕮 **Pause.** Try Exercise 4.3 before reading on.

4.5 The Ventilator's Loops

Every 2 ms the controller reads P_m and Q_m, each with seeded Gaussian noise, and nothing else about the patient. In that period it accumulates the sample into the breath under way, steps the breath sequencer, which at the end of a breath has the estimator fit it and the supervisor update the volume loop and the caps, then runs the valve servos and finally the alarms. The settings are listed in Table 4.3.

4.5.1 The Breath Sequencer

The sequencer is a state machine over expiration, inspiration, the volume-control pause and two hold maneuvers. Expiration ends on a patient trigger once the lockout T_{lock} after inspiration has passed, when the wye flow less the estimated leak, filtered over 10 ms, exceeds $\dot{V}_{trig}$ or the airway pressure falls ΔP_{trig} below PEEP, or on the clock when a breath at the rate f falls due. In volume control the flow follows a square or decelerating profile for $T_i - T_p$ and the pause holds both valves shut. A pressure breath follows the target of Equation (4.7) for T_i in pressure control, and in pressure support it is cycled off once at least 0.15 s and half the rise have passed and the filtered flow falls below f_{cyc} of its peak, or at $T_{i,max}$. Any breath ends at once above P_{high}. The delivery of a breath is fixed when

it starts, so a setting moved mid-breath applies to the next one. When a volume-control breath has a pause, its square profile delivers $Q = V_T/(T_i - T_p + \tau_v + C_t\hat{R})$, because after the flow phase the valve still delivers for its response time and the circuit, charged to the resistive pressure, empties into the lung during the pause, 20 ms of extra flow time on the shipped patient. There is no separate synchronized intermittent mandatory mode, since pressure support with a backup rate covers its lessons.

4.5.2 Three Servos on One PID

All three servos are instances of the velocity-form PID of Section 1.6.2, Equations (1.12) to (1.15), which clamps only its emitted output and lets the integral alone absorb the clamp by back-calculation with $T_t = \tau_I$. They execute every $\Delta t_c = 2\,\text{ms}$, with no derivative action and no rate limit, and both inspiratory servos are reset to zero at the start of every breath. The action $\mathscr{A}$, the a of Equation (1.12), is $+1$ for the two inspiratory servos, whose valve raises the pressure or the flow it controls, and -1 for the PEEP servo, since more expiratory opening lowers the pressure. Each servo adds its output to a feedforward u_{ff} that inverts the valve's own law of Equation (4.3), the opening that would pass a demanded flow $\hat{Q}$ at the measured pressure, and its clamp is $[-u_{ff}, 1-u_{ff}]$ so that the total stays within the valve. The pressure target of a pressure breath and the demand of each servo are given by Equation (4.7).

$$\begin{aligned} P_{sp}(\tau) &= \text{PEEP} + \tfrac{1}{2}\Delta P\big(1-\cos(\pi\tau/T_r)\big), \quad \tau < T_r \\ \hat{Q}_{PC,PS} &= (Q_m + \tau_v\dot{Q}_m)^+ + C_t\dot{P}_{sp} \\ \hat{Q}_{VC} &= Q_{prof}(\tau) + \hat{Q}_{leak} \\ \hat{Q}_E &= \big(u_i k_v \psi(P_b - P_m) - Q_m\big)^+ \end{aligned} \tag{4.7}$$

where $\dot{Q}_m$ is the measured flow's derivative filtered over 6 ms, Q_{prof} the volume-control profile, $\hat{Q}_{leak} = \hat{k}_l\,\psi(P_m)$ the estimated leak when compensation is on and $\hat{Q}_E$ the flow the expiratory valve must shed. The pressure servo holds its integral, $\tau_I = \infty$, until the rise has ended, because integrating the designed lag behind a moving target holds the valve open past it. The PEEP servo tracks the valve whenever it is not in charge, so that it takes over without a bump. In expiration the inspiratory valve passes a base flow of $Q_{bias} + \dot{V}_{trig}$, Q_{bias} alone with a pressure trigger, plus with compensation the estimated leak at the larger of P_m and PEEP. The bias sits above the trigger because a fixed base flow below the threshold could never let a patient trigger, and the expiratory valve, which can only let gas out, holds PEEP against a leak only with this flow from the other side.

On a passive lung, a 15 cmH_2O step from a PEEP of 5 cmH_2O in 0.1 s overshoots by at most 0.77 cmH_2O over R from 5 to 40 $\text{cmH}_2\text{O}\cdot\text{s/L}$ and C from 20 to 80 $\text{mL/cmH}_2\text{O}$, rising from 10 to 90 % in 54 to 67 ms, measured on the second breath of case 3 with the PEEP, the lung and the timing set before Run, so that the lung starts at rest at the new PEEP. Two copies of the controller changed in one place each show what the design buys. A straight-line rise overshoots by at most 0.61 cmH_2O but takes 76 to 82 ms, and letting the integral run during the rise raises the worst overshoot to 1.26 cmH_2O, past the 1.0 cmH_2O that the application's own check allows.

4.5.3 Identification From Every Breath

The estimator identifies the individual lung, the role Section 1.4.4 gives to measurement in a living plant. Over each whole breath it fits the passive equation of motion by least squares, as shown in Equation (4.8).

$$\begin{aligned} &\min_{R,C,P_0} \sum_k \left(P_m(t_k) - R\hat{V}(t_k) - \frac{\hat{V}(t_k)}{C} - P_0\right)^2 \\ &\hat{V} = Q_m - \hat{k}_l\,\psi(P_m), \qquad \widehat{\text{PEEP}}_i = P_0 - P_{ee} \\ &\hat{k}_l = \frac{\int Q_m\,\mathrm{d}t}{\int \psi(P_m)\,\mathrm{d}t} \end{aligned} \tag{4.8}$$

where $\hat{V}$ integrates $\hat{V}$ from the start of the breath, P_0 is the alveolar pressure the breath started from and P_{ee} the measured end-expiratory pressure, filtered through 25 ms during expiration. The normal equations are summed every 2 ms and solved when the next breath starts, and a fit counts only with at least 50 samples, a breath longer than 0.8 s, more than 50 mL, R between 0.2 and 300 $cmH_2O \cdot s/L$ and an elastance between 2 and 500 cmH_2O/L. The smoothed $\hat{R}$, $\hat{C}$ and auto-PEEP weigh the newest fit by $\alpha = 0.5$, and an option lets them learn only from breaths the machine triggered. The leak coefficient $\hat{k}_l$ uses the whole breath, since a lung that ends where it began adds no net flow, and it weighs the newest breath by 0.8. In the shipped challenge it reads 13.8 to 14.1 L/min on every breath between minutes 31 and 39 for the true 14 L/min, auto-triggers and all, but once the patient's efforts fall at different points of each machine breath the premise fails and the per-breath value swings from 9.2 to 20.7 L/min around a mean of 14.2, which is why the monitor reads 11.2 L/min at minute 45. Holds shut both valves, for 0.6 s at the end of inspiration to read the plateau over the last 0.1 s, or for 1.0 s before the next breath to read the total PEEP over the last 0.2 s.

4.5.4 The Volume Loop and the Protective Layer

With adaptive support on, in pressure control or pressure support, the shared PID runs once per breath as a discrete integrator on the measured volume, as shown in Equation (4.9).

$$\begin{aligned} \Delta P_{k+1} &= \Delta P_k + \frac{K_V}{\hat{C}\hat{\eta}} \left(V_T^* - \hat{V}_{T,k}\right) \\ \hat{\eta}_{PC} &= 1 - \mathrm{e}^{-(T_i - T_r/2)/\hat{R}\hat{C}}, \qquad \hat{\eta}_{PS} = 1 - f_{cyc} \end{aligned} \tag{4.9}$$

where $\hat{C}\hat{\eta}$, with $\hat{\eta}$ clipped to $[0.3, 1]$, is the volume one more cmH_2O would buy, so that $K_V = 1$ would remove the whole error in one breath. The step is limited to 3 cmH_2O per breath by the PID's slew limit, the support stays between 2 cmH_2O and the protective cap, or 60 cmH_2O with the layer off, and the loop freezes its output while the leak estimate moves by more than 3 L/min at 10 cmH_2O between breaths, because until a new leak is learned the measured volume carries the part not yet subtracted. With the loop off, the PID tracks the operator's setting, so switching on is bumpless. The protective layer caps each breath before it starts. In volume control the volume is limited to the setting, to $V_{T,max}$ times the predicted body weight and to $\hat{C}(P_{plat,max} - \mathrm{PEEP} - \widehat{\mathrm{PEEP}}_i)$, and in the pressure modes ΔP is limited to $P_{plat,max} - \mathrm{PEEP} - \widehat{\mathrm{PEEP}}_i$ and a breath is cycled off once its volume reaches $V_{T,max}$.

4.5.5 Alarms, Manual Mode and the Three Models

The ventilator's alarms come from its own measurements. DISCONNECT, HIGH PRESSURE and PEEP NOT HELD, the last after 1.5 s more than 2 cmH_2O below PEEP in expiration, are alarms, and LEAK above 10 L/min at 10 cmH_2O, AUTO-PEEP above 2 cmH_2O, LOW TIDAL VOLUME below 75 % of the target and APNEA BACKUP are warnings. ASYNCHRONY comes from the truth audit, which counts an effort that starts in expiration and starts no breath within 0.1 s of its end as missed, a patient trigger outside any effort as an auto-trigger and a second trigger inside one effort as a double, ignoring efforts below 0.5 cmH_2O. Manual mode hands both valves to the operator as a bag with no PEEP valve, a squeeze being an inspiration, while the PEEP servo tracks the hand. The return to automatic mode starts the cycle in expiration: the inspiratory valve takes the base flow at once, and the PEEP servo takes the expiratory valve over from where the hand left it.

The simulation model is Equations (4.1) to (4.4). The control models are the closed forms of Section 4.4, a first-order lung behind a valve whose law is known. The controller's internal models are the inverted valve laws of the feedforwards, the passive equation of motion that the estimator fits, the orifice law it assumes for the leak, and the gain $\hat{C}\hat{\eta}$ of the volume loop. None of them contains the effort.

Table 4.4 The Nine Cases and the Challenge

#	Change from the reference line	Teaching point
1	manual, 1 s squeeze every 5 s, C to 20 at 30 s	the bag
2	VC, PEEP 8, 15 min^{-1}, C 50 to 18 at 30 s	the plateau pays
3	PC as case 2	the volume pays
4	PS 8, effort 6 at 18 min^{-1}, backup 10 min^{-1}	triggering, cycling
5	case 4 with effort 5 at 15, 30 L/min leak at 20 s	a leak
6	VC at 24 min^{-1}, no pause, bronchospasm at 20 s	auto-PEEP
7	holds at 20 and 25, C 30 at 40, spasm at 80, effort 5 at 160 s	identification
8	PS, 15 L/min trigger, then 0.5 L/min and a 2 L/min leak, then VC 0.6 s	asynchrony
9	PC, PEEP 8, compensation, 300 L/min leak at 20 s	honest infeasibility
10	60 min: ARDS ramp, 14 L/min leak at 30, patient at 40	challenge

Table 4.5 Measured Outcomes of the Cases

Case and configuration	Result
1, 50 mL/cmH_2O and 20 mL/cmH_2O	422–423 and 384 mL, peaks 12.6–12.7 and 22.7–22.8 cmH_2O
2, as shipped	419 to 423 mL, plateau from 16.4–16.5 to 31.3–31.5 cmH_2O
2, protective layer	six breaths of 420–423 mL at 31.4–31.5 cmH_2O, then 395–396 mL at 29.9–30.0 cmH_2O
3, as shipped	424 to 179 mL, plateau 17.9 to 18.0 cmH_2O
3, adaptive, then with the layer	420 mL at 31.3 cmH_2O, then 393–394 mL at 29.9 cmH_2O
4, cycle-off 25, 5, 50 %	0.73–0.74, 0.84, 0.60 s against 1.00 s
5, no compensation	11 auto-triggers, 2.50 s breaths, 103 to 414 mL
5, compensation	2 auto-triggers, 0.68 to 0.77 s, 454 to 462 mL
6, 24 min^{-1}, 18, 14	auto-PEEP 2.58, 0.96, 0.35 cmH_2O
7, passive, then effort 5	$\hat{R}$, $\hat{C}$ within 0.5 and 0.2 %, then −21.6 to −21.8 and +10.8 to +12.0 %
8, three parts	8 of 14 missed, 13 of 14 auto, 13 of 14 double
9, 300 L/min leak	3.45 to 3.46 cmH_2O, valve open, no breath
10, shipped, suggested	27.34 and 99.74 % in band

4.6 Nine Cases and an Hour

Table 4.4 lists the cases. Each starts from the reference line of Tables 4.2 and 4.3 on the passive shipped patient and changes it as its row says. Every number below comes from a run of the engine without the interface at the default seed, stepped as the browser steps it (Section 1.6.1). A setting is changed at $t = 0$ before Run unless a Reset is named, which restarts the case with the learner's settings kept. All volumes, plateaus and auto-PEEP values are true values from the hidden record unless they are called measured. The outcomes are collected in Table 4.5.

4.6.1 The Bag

Case 1 runs in manual mode with the inspiratory valve opened to 16 % for 1 s every 5 s and the expiratory valve shut during each squeeze. Into the shipped lung each squeeze delivers 422 to 423 mL at a peak of 12.6 to 12.7 cmH_2O. At 30 s the lung stiffens to 20 mL/cmH_2O, and the same squeeze delivers 384 mL, 9 % less, at 22.7 to 22.8 cmH_2O, 1.8 times the pressure, because the gas comes through a valve from a source of fixed pressure and the flow falls as the airway pressure climbs toward it. With no PEEP valve the lung empties to atmospheric pressure after each release.

4.6.2 Which Variable to Hold

Case 2 delivers 420 mL fifteen times a minute in volume control at a PEEP of 8 cmH_2O while the lung stiffens from 50 to 18 mL/cmH_2O at 30 s. The volume stays at 419 to 423 mL and the pressure pays for it. The plateau read in the pause climbs from 16.4 or 16.5 to between 31.3 and 31.5 cmH_2O, the peak from 21.3 or 21.4 to between 35.8 and 36.2 cmH_2O, and the driving pressure reaches 23.3 to 23.5 cmH_2O. With the protective layer on, the six breaths after the step still deliver 420 to 423 mL at

31.4 to 31.5 cmH_2O, because the layer caps on the identified compliance and the smoothed estimate moves only halfway to each new fit. Trimming starts at 56.5 s, by 72.5 s the layer caps the volume at 396 mL, and from then on breaths of 395 to 396 mL hold the plateau at 29.9 to 30.0 cmH_2O. Ten minutes in, the cost is 44.0 mmHg of CO_2 against 41.7 mmHg unprotected.

> **TRY IT IN THE APP** **2 · Volume Control and a Stiffening Lung**
>
> Load scenario 2 and press Run on the Loops view. At 30 s the lung stiffens from 50 to 18 mL/cmH_2O. Watch the plateau badge: the volume stays at 419 to 423 mL while the plateau climbs from 16.4 or 16.5 to between 31.3 and 31.5 cmH_2O, past the dashed 30 cmH_2O limit.

Case 3 holds 10 cmH_2O above the same PEEP. The plateau stays at 17.9 to 18.0 cmH_2O and the volume falls from 424 to 179 mL. With adaptive support the first breath after the step is still 177 mL, the next five are 231, 285, 339, 380 and 402 mL, the first three each limited to 3 cmH_2O of new support and so about 54 mL apart at 18 mL/cmH_2O, and by the tenth the volume is 420 mL at 23.4 cmH_2O of support and a plateau of 31.3 cmH_2O, as in volume control. With the protective layer as well, the support stops at 22.0 cmH_2O, the plateau limit less PEEP, and the breath settles at 393 to 394 mL under a plateau of 29.9 cmH_2O. Adaptive pressure control and volume control end in the same place, and the layer is the ceiling over both.

4.6.3 A Patient Who Breathes

In case 4 the patient pulls 6 cmH_2O for about 1 s eighteen times a minute on pressure support of 8 cmH_2O, every effort triggers a breath, and the volume is 482 to 486 mL. The machine ends each breath when the flow falls to 25 % of its peak, after 0.73 to 0.74 s, while the patient pulls for 1.00 s. A cycle-off of 5 % stretches the breath to 0.84 s and 50 % cuts it to 0.60 to 0.61 s, and neither matches the patient, because the flow the machine reads carries the effort and the support together. A low ratio of time constant to neural time, 0.5 here, is the condition under which a fixed fraction ends breaths early [16]. Support levels of 4 and 14 cmH_2O give 333 to 335 and 705 to 712 mL. The estimator, which assumes a passive lung, reads this patient's compliance as 89.9 against a true 50 mL/cmH_2O.

> **HISTORY** **The Synchrony Problem of 1929**
>
> The tank respirator's authors set its rate and pressures by hand, watching a water manometer, and found that a breathing subject does not simply follow the machine. As they raised the pressure, a subject fell into rhythm with the pump at about 5 to 10 cmH_2O and stayed in rhythm at any higher pressure. One man of normal physique needed 18 to 20 cmH_2O before the pump completely controlled his breathing. Refractory subjects were brought into step by starting at 18 to 20 cmH_2O and coming down quickly to 5 to 10 cmH_2O [12]. The tank kept a rhythm set by hand, while the ventilator of Section 4.5 starts a breath when the patient's own pull crosses its trigger.

4.6.4 A Leak at the Mask

Case 5 opens a 30 L/min leak at 20 s under a patient on pressure support who had triggered every breath at 454 to 455 mL. Without compensation three things break together. The leak's flow passes the trigger as each lockout ends, giving 11 auto-triggers in the 40 s to the end of the first minute, the first of them 2 ms after the leak opens, while only 3 of the patient's 10 efforts trigger anything. Every breath runs to the 2.50 s limit, because the flow never falls to a quarter of its peak. The measured volume reads 1521 to 1830 mL while the lung receives 103 to 414 mL and starts breaths from up to 5.99 cmH_2O of trapped pressure. With compensation the estimator learns 30.0 L/min, the auto-triggers stop after the first two, 9 of 10 efforts trigger, the breaths that follow last 0.68 to 0.77 s, and after one breath of 370 mL from the trapped volume the lung receives 454 to 462 mL, as it did before the leak.

> **TRY IT IN THE APP** **5 · A Leak**
>
> Load scenario 5 and press Run with Leak compensation off. After the leak opens at 20 s, the red rings on the pressure trace mark breaths the patient did not ask for, 11 of them before the end of the first minute. Press Reset, turn on Leak compensation on the expiratory valve card and run again: the auto-triggers stop after the first two.

4.6.5 Bronchospasm

Case 6 runs volume control at 24 min^{-1} with 1 s in and 1.5 s out while the resistance doubles to 20 $cmH_2O \cdot s/L$ from 20 s. The auto-PEEP rises from at most 0.48 to 2.58 cmH_2O, the per-breath fit reads 2.57 to 2.60 cmH_2O, and an expiratory hold requested at 90 s reads a total PEEP of 7.48 cmH_2O, an auto-PEEP of 2.48 against a true 2.49, while the airway pressure at the end of each expiration still reads the set PEEP. Only the expiratory time helps. With one setting changed at a time, the case reloaded before each, as its briefing directs, the auto-PEEP is 0.96 cmH_2O at 18 min^{-1} and 0.35 cmH_2O at 14 min^{-1}, and a 0.7 s inspiration, which leaves 1.80 s out, gives 1.79 to 1.80 cmH_2O. A PEEP of 10 cmH_2O leaves the trapping at 2.58 cmH_2O and lifts the plateau from 16.0 to 21.0 cmH_2O.

4.6.6 Identification Breath by Breath

Case 7 fits every breath of volume control while the lung stiffens to 30 mL/cmH_2O at 40 s, the resistance doubles from 80 s and the patient starts 5 cmH_2O efforts at 160 s. On the passive lung each fit recovers R within 0.5 %, C within 0.2 % and the auto-PEEP within 0.04 cmH_2O at both compliances, except the breath during which the lung stiffens, whose resistance comes out 42 % low, and while the resistance ramps the fit runs 0.5 to 1.0 % high, since R changes within the breath. The inspiratory hold at 20 s reads 13.44 cmH_2O, exactly the true plateau, which gives a static compliance of 49.8 mL/cmH_2O, and the expiratory hold reads 5.03 cmH_2O, an auto-PEEP of 0.03 cmH_2O. Once the patient breathes, every breath the patient triggers fits the resistance 21.6 to 21.8 % low and the compliance 10.8 to 12.0 % high, and by 220 s the smoothed resistance has drifted to 15.65 $cmH_2O \cdot s/L$ against 20.00. Learning only from machine-triggered breaths holds it at 17.80, since an effort can begin inside a breath the machine started.

4.6.7 Three Kinds of Asynchrony

Case 8 runs 120 s in three parts against a weak patient of 3 cmH_2O at 20 min^{-1}. With the flow trigger at 15 L/min, 8 of 14 efforts are missed and the patient gets backup breaths at 8 min^{-1}. At 40 s the trigger is softened to 0.5 L/min as a 2 L/min leak opens, and 13 of the 14 breaths that follow are auto-triggers. At 80 s the cuff is sealed, the mode becomes volume control with a 0.6 s inspiration and the patient pulls 12 cmH_2O for 1.5 s, and 13 of 14 efforts trigger twice. The second breath stacks on the first, so the plateau reaches 23.7 cmH_2O where the first breath of each pair stays at 14.4 to 14.8 cmH_2O, from up to 10.31 cmH_2O of trapped pressure. A 2 L/min trigger set before Run catches all 14 efforts of the first part, and set back to 2 L/min after the scripted change at 40 s it catches all 13 of the second without firing once on the 2 L/min leak.

4.6.8 A Leak No Blower Can Hold

Case 9 runs pressure control at 10 cmH_2O above a PEEP of 8 cmH_2O with compensation on, and at 20 s the mask comes away, 300 L/min at 10 cmH_2O. The estimator measures the leak as 300 L/min, the inspiratory valve is fully open from 21.8 s on, and the airway pressure sits at 3.45 to 3.46 cmH_2O in inspiration and expiration alike, the balance of Section 4.4, so no breath reaches the lung. The ventilator raises PEEP NOT HELD, LEAK and LOW TIDAL VOLUME, the status strip shows the first, and the CO_2 climbs from 40.0 to 42.9 mmHg between 20 and 60 s. No controller makes the blower bigger, and the alarm and a person at the mask are the answer.

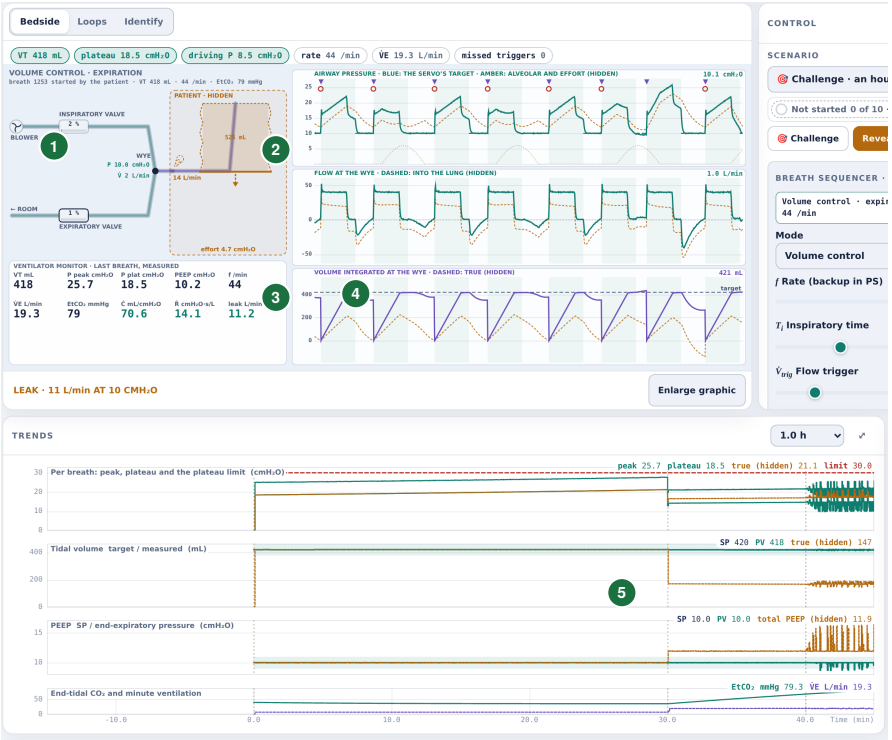


Figure 4.3 The challenge as shipped, volume control at 420 mL, 18 min^{-1} and a PEEP of 10 cmH_2O, at minute 45, Bedside view with a one-hour trend window, cropped to the plant view and the trends: (1) the circuit with its two valves and the wye, (2) the hidden patient, (3) the ventilator monitor, (4) the waveforms and (5) the tidal volume trend. The clock in the header and the scorecard of the Learn panel are cut. The header of the plant view reads volume control, expiration, breath 1253 started by the patient, and the monitor reads a tidal volume of 418 mL, a peak of 25.7 cmH_2O, a plateau of 18.5 cmH_2O, PEEP 10.2 cmH_2O, 44 min^{-1}, 19.3 L/min of minute ventilation, an end-tidal CO_2 of 79 mmHg, $\hat{C}$ 70.6 mL/cmH_2O, $\hat{R}$ 14.1 $\text{cmH}_2\text{O}\cdot$s/L and a leak of 11.2 L/min, while the hidden panel shows the true 14 L/min leak and an effort of 4.7 cmH_2O. In the 10 s of waveforms all eight breaths are marked as patient-triggered, seven carry the red ring of an auto-trigger, and the volume integrated at the wye reaches the 420 mL target each time while the dashed true volume peaks at 200 to 225 mL and falls back as the leak drains the lung in each pause. The status strip reads LEAK, 11 L/min at 10 cmH_2O. Over the hour the trends show the true tidal volume leaving 420 mL at minute 30 and ending the window at 147 mL, the true plateau at 21.1 cmH_2O against a measured 18.5, the total PEEP at 11.9 cmH_2O against a measured 10.0, and the CO_2 at 79.3 mmHg.

4.6.9 An Hour on the Ventilator

The challenge ventilates for an hour while the lung stiffens steadily from 50 to 25 mL/cmH_2O, a 14 L/min leak opens past the cuff at minute 30, and from minute 40 the patient breathes at 22 min^{-1}, each effort growing to 6 cmH_2O over a minute. The target, scored on the hidden record, asks for at least 98 % of breaths within 10 % of 420 mL, no plateau above 30 cmH_2O, no auto-PEEP above 3 cmH_2O, and at most 10 missed efforts and false triggers together, and it allows 2 % of breaths out of band because the first few breaths of a new leak come before it is learned. Stiffening alone cannot fail volume control here, since at a PEEP of 10 cmH_2O and 420 mL the plateau reaches 30 cmH_2O only below 21 mL/cmH_2O and the lung stops at 25, so the hour turns on the leak. The shipped configuration, volume control at 18 min^{-1}, is faultless for half an hour, with 420 to 422 mL per breath and a plateau that rises to 21.2 cmH_2O. The leak then auto-triggers it 1211 times, at 47.7 min^{-1}

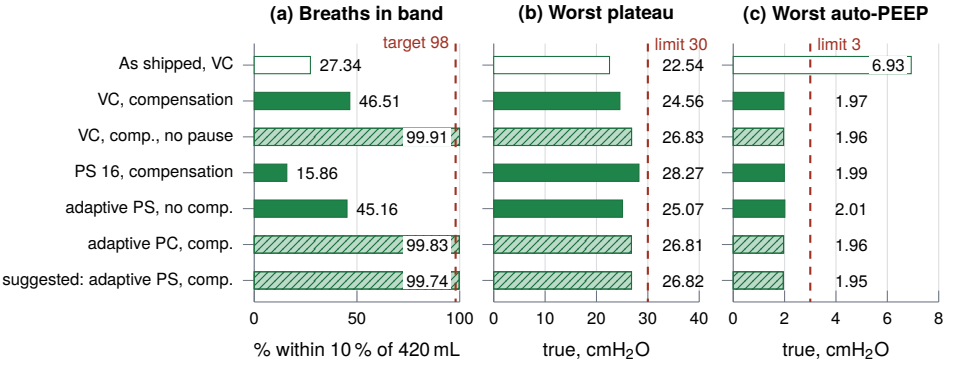


Figure 4.4 The seven configurations of Table 4.6 over the hour of the challenge: (a) the share of breaths within 10 % of 420 mL against the 98 % target, (b) the worst true plateau against 30 cmH_2O and (c) the worst true auto-PEEP against 3 cmH_2O, limits dashed. White: the shipped configuration. Solid: the other rows that fail. Hatched: the rows that pass. Every configuration keeps the plateau under its limit, every one but the shipped one keeps the auto-PEEP under its limit, and every row that fails falls short of the share of breaths in band.

between minutes 30 and 40, and each breath sends part of its volume out of the leak, so the lung receives 139 to 193 mL and starts breaths from up to 6.93 cmH_2O. Only 27.34 % of the hour's breaths are in band, and the CO_2 ends at 111.0 mmHg, as shown at minute 45 in Figure 4.3.

Table 4.6 compares the configurations, and Figure 4.4 draws three of its columns as bars. Adaptive pressure support with leak compensation, the configuration the challenge suggests, meets the whole target, the same whether it is revealed before Run or carried through a Reset, with 99.74 % in band, 1.74 points above the 98 % required, a plateau of at most 26.82 cmH_2O, auto-PEEP of at most 1.95 cmH_2O, one missed effort and two auto-triggers, and it ends with 14.2 cmH_2O of support and a CO_2 of 29.3 mmHg, on which the target is silent. Its three misses, 517 mL, 371 mL and 516 mL, are the two breaths before the new leak is learned and the first after it, all within 5.4 s of the leak opening. On seeds 1 to 5, 7 and 99 it passes with 99.65 to 99.74 % in band and auto-PEEP of 1.92 to 1.96 cmH_2O, and the shipped configuration fails on all of them with 27.25 to 27.30 %. Volume control with compensation alone reaches 46.51 %, because the lung drains through the leak during the 0.2 s pause with both valves shut, and without the pause it passes. Adaptive support without compensation still takes 660 auto-triggers, and a fixed support of 16 cmH_2O with compensation leaves 84 % of breaths out of band as the compliance halves. The target is a worst case over breaths. With the leak slider set to another size when the leak opens, an exploration the challenge itself does not credit, the shipped volume control passes with no leak or 1 L/min, fails at 2 L/min on 611 auto-triggers with 99.16 % still in band, and collapses to 39.07 % at 3 L/min. The suggested configuration passes at 20 L/min and fails at 30 L/min on a single breath, the second after the leak opens, which starts 4.58 cmH_2O above PEEP, while from minute 31 no breath exceeds 0.21 cmH_2O.

Table 4.6 The Challenge Over 60 min at the Default Seed. The Volume Is the Share of Breaths Within 10 % of 420 mL, and Triggers Are Missed Efforts and False Triggers Together

Configuration	Vol. (%)	P_{plat}	PEEP$_i$	Trig.	Pass
as shipped, VC	27.34	22.54	6.93	1211	no
VC, compensation	46.51	24.56	1.97	4	no
VC, compensation, no pause	99.91	26.83	1.96	4	yes
PS 16 cmH_2O, compensation	15.86	28.27	1.99	3	no
adaptive PS, no compensation	45.16	25.07	2.01	660	no
adaptive PC, compensation	99.83	26.81	1.96	3	yes
suggested, adaptive PS, comp.	99.74	26.82	1.95	3	yes

Plateau and auto-PEEP are the worst true values, in cmH_2O.

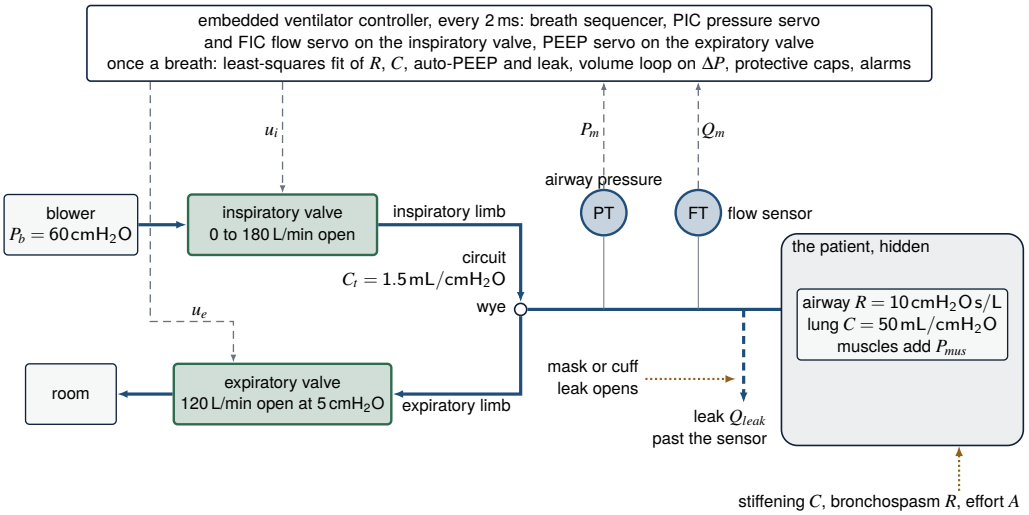


Figure 4.5 The loops of Section 4.5 with the hardware of an intensive care ventilator. The blower at P_b feeds the inspiratory valve, the final element of the pressure servo (PIC) and the flow servo (FIC). The expiratory valve vents the circuit to the room and holds PEEP. The only sensors are the airway pressure transmitter (PT) and the flow sensor (FT) at the wye, read by the embedded controller every 2 ms. Everything in the shaded box, and the leak between the sensor and the patient, is hidden from it. Thick solid arrows carry gas, thin dashed arrows signals, dotted arrows disturbances. Ranges are those of Table 4.2.

✦ **DID YOU KNOW? A Fifth of a Second Decides the Hour**

From the simulator: with leak compensation on, volume control keeps 46.51 % of the challenge's breaths within 10 % of the target. The same configuration without its end-inspiratory pause keeps 99.91 %, the highest share of any configuration in Table 4.6.

4.7 Technology in Practice

The blower, the two valves and the two transmitters of Figure 4.1 stand for the hardware of an intensive care ventilator. Figure 4.5 draws the loops of Section 4.5 with that hardware in place of the schematic's symbols.

What Measures It

Airway pressure is read by a pressure transducer and flow by a flow sensor, which in ventilators is most often a differential pressure sensor, a hot-wire anemometer or an ultrasonic meter. Where the flow sensor sits matters. A sensor near the expiratory valve also counts the gas compressed in the circuit, which grows with the peak pressure and changes as water condenses, while a sensor at the wye does not [19]. The chapter's sensors sit at the wye, read the circuit pressure P and the flow $\dot{V} + Q_{leak}$ of Equation (4.3), and carry Gaussian noise of 0.05 cmH_2O and 0.3 L/min with no drift.

> **TECHNOLOGY** **Sensor: The Flow Sensor at the Wye**
>
> United States regulation describes a pneumotachometer as a device that finds gas flow from the pressure drop across a known resistance, a set of capillaries or a metal screen [20]. Ventilators most often use differential pressure sensors, hot-wire anemometers or ultrasonic meters. In a variable-orifice sensor a slit flap opens wider as the flow rises, and a jet from a narrow fitting can set the flap fluttering. A heated wire is cooled by the passing gas, reads most finely at low flow and can drift at zero as the gas temperature changes. The standard for critical care ventilators allows a measured volume above 50 mL to be off by 4.0 mL plus 15 % [19]. Water condensing in the circuit can show as a rhythmic ripple on the expiratory flow trace and trigger breaths the patient did not start [21].

What Moves It

The final elements are the inspiratory valve, which meters gas from the supply into the circuit, and the expiratory valve, which lets it out and holds PEEP. In the chapter each is an orifice whose opening follows its command through a lag of 5 ms, with the inspiratory valve passing 180 L/min fully open into an open airway. Only the inspiratory valve can add gas, which is the one-sided authority that sets the largest leak of Section 4.4.

> **TECHNOLOGY** **Final Element: Inspiratory and Expiratory Valves**
>
> A continuous ventilator, a Class II device in United States regulation, mechanically controls or assists the patient's breathing [22]. A system of solenoid-driven valves has been the design of choice for modern ventilators, which must deliver precise volumes, rates and PEEP and allow rapid adjustment. One open low-cost design switches the inspiratory gas with a solenoid valve and holds PEEP with a pneumatically driven exhalation valve near the patient. Such valves fail safe to a defined open or closed position, last many cycles and act fast. The standard for critical care ventilators asks for an anti-asphyxia valve, a mechanical relief valve against overpressure, and alarms for low pressure, which means a disconnected circuit, for high pressure and for insufficient flow [23].

Where the Controller Runs

Open designs show where such loops run: one emergency ventilator's monitor sampled its pressure sensor 100 times a second on a microcontroller [24], and a low-cost solenoid-valve ventilator was controlled through a microcontroller board [23]. A modern intensive care ventilator answers a patient's trigger in under 100 ms [21]. The chapter's controller runs every 2 ms behind a 10 ms trigger filter, and the fit and the volume loop run once a breath, a layer above the servos. The essential performance of the whole machine is set by the standard for critical care ventilators [10].

What Is Modernizing It

Three developments bear on these loops. Ventilation can now be coupled to the electrical activity of the diaphragm, measured by an esophageal catheter, so that the support follows the patient's neural output. Synchrony improved, though the clinical impact still had to be determined. Transpulmonary pressure, the airway pressure less a measured esophageal pressure, and electrical impedance tomography are being studied for setting PEEP [25]. Both would give the controller a view of the effort and the lung that the wye of this chapter cannot. Home ventilators, finally, send data through the internet to

cloud systems, where clinicians review usage, estimated unintentional leak, tidal volume and minute ventilation [4].

> **IN PRACTICE** **Reading the Waveforms at the Bedside**
>
> A basic anesthetic monitoring standard requires that, whenever a mechanical ventilator controls the breathing during anesthesia, a device in continuous use detects a disconnected breathing system and sounds an audible alarm [26]. Beyond the alarms, the clinician reads the pressure and flow waveforms. A total rate above the set rate in a controlled mode suggests auto-triggering, whose causes include air leaks, the trigger setting and condensation in the circuit. A modern ventilator answers a trigger in under 100 ms when its sensitivity is set correctly. Visual inspection is the accepted method, yet in one study intensive care physicians recognized fewer than a third of the asynchronies [21].

4.8 Limitations

The application teaches the control structure of a ventilator and is not a medical device. Its lung, its patient and every limit it draws were chosen to make the control lessons visible, nothing in it has been validated against patient data, and none of its settings or results may be carried to a ventilator that serves a person. The lung-protective targets are those of a trial [2], the requirements a real ventilator meets are those of its standard [10], and the asynchrony classes come from the clinical literature [9].

> **OPEN CHALLENGE** **A Closed Loop With a Patient Inside**
>
> A physiological closed loop feeds a measurement of the patient back to the ventilator's settings, and an international collateral standard now covers such controllers. Their sensors limit them. Intravascular sensors of arterial carbon dioxide were too slow for some loops and are no longer sold, so control fell back on end-tidal carbon dioxide, whose gap to the arterial value grows in a diseased lung. Sensor failure, unpredictable disturbances and software errors remain open problems [25]. The volume loop of Section 4.5 meets a mild form of this when it freezes while a new leak is learned. Progress would be safety concepts that keep an automated loop harmless, transparent and lighter on the clinician's workload [25].

The modeling limits are those of the assumptions in Section 4.3.5. One linear compartment cannot recruit, derecruit or overdistend, so the application cannot show why PEEP is set in ARDS or why a pressure-volume curve bends. The scripted effort ignores CO_2 and support, so the patient never fights back or tires. The machine's laws are known exactly to its controller, its sensors do not drift, its leak has a fixed law, oxygenation is absent, and the audit uses an effort that a bedside cannot see. Every computer replays the same hour from the same seed, which no patient does.

> **OPEN CHALLENGE** **Seeing the Effort the Wye Cannot**
>
> A review of 19 studies of automatic asynchrony detection by rules, machine learning and deep learning reported an average sensitivity of 0.80 and specificity of 0.93. Yet most detectors were available only offline, found a small subset of asynchronies, mostly ineffective efforts and double triggers, or were still in development or validation, as 16 of the 19 were. Only 11 studies had a reference measurement of the patient's effort, such as esophageal pressure or the electrical activity of the diaphragm [27]. The truth audit of Section 4.3.5 uses that reference, the hidden P_{mus}. Progress would be detectors of several kinds of asynchrony, validated on new data against a measured effort and run in real time at the bedside [27].

4.9 Conclusion

A ventilator holds what its two sensors at the wye report, and this simulator measures how far that can be from the breath the lung receives. On a lung stiffened to 18 mL/cmH_2O, volume control keeps 420 mL by letting the plateau reach 31.3 to 31.5 cmH_2O, pressure control keeps its plateau by letting the volume fall to 179 mL, and the protective layer, working from the identified compliance, holds both at 393 to 396 mL under 30 cmH_2O. Trapped pressure follows the expiratory time alone, 5 to 8 %

above its closed form. A leak past the wye shows up as auto-triggers, hung breaths and a measured volume of 1521 to 1830 mL for 103 to 414 mL delivered, and a 300 L/min leak holds the airway at 3.46 cmH_2O whatever the controller does. The per-breath fit identifies a passive lung within 0.5 %, and an effort that triggers biases it systematically. Over the hour, volume control with a leak keeps 27.34 % of its breaths in band and adaptive pressure support with compensation 99.74 %. The lung is one linear compartment and the patient a script, and a fuller model would add recruitment, a chemoreflex, oxygenation and a changing mask leak. In this browser plant the measured and the delivered breath can be compared on every cycle.

KEY POINTS

- A ventilator holds what its two sensors at the wye report, and that can be far from the breath the lung receives.
- On a lung stiffened to 18 mL/cmH_2O, volume control keeps 420 mL by letting the plateau reach 31.3 to 31.5 cmH_2O, and pressure control keeps its plateau by letting the volume fall to 179 mL.
- A leak past the wye shows up as auto-triggers, hung breaths and a measured volume of 1521 to 1830 mL for 103 to 414 mL delivered.
- The per-breath fit identifies a passive lung within 0.5 %, and an effort that triggers biases it systematically.

Where else this applies. Cycles whose events the plant partly decides return in Volume 8, Chapter 7 and Chapter 2.

References

[1] A. S. Slutsky, "History of mechanical ventilation. From Vesalius to ventilator-induced lung injury," *American Journal of Respiratory and Critical Care Medicine*, vol. 191, no. 10, pp. 1106–1115, 2015.

[2] The Acute Respiratory Distress Syndrome Network, "Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome," *New England Journal of Medicine*, vol. 342, no. 18, pp. 1301–1308, 2000.

[3] M. B. P. Amato, M. O. Meade, A. S. Slutsky, L. Brochard, E. L. V. Costa, D. A. Schoenfeld, T. E. Stewart, M. Briel, D. Talmor, A. Mercat, J.-C. M. Richard, C. R. R. Carvalho, and R. G. Brower, "Driving pressure and survival in the acute respiratory distress syndrome," *New England Journal of Medicine*, vol. 372, no. 8, pp. 747–755, 2015.

[4] R. F. D'Cruz and N. Hart, "A history of home mechanical ventilation: The past, present and future," *Chronic Respiratory Disease*, vol. 21, p. 14799731241240776, 2024.

[5] R. Goebel, R. G. Sanfelice, and A. R. Teel, "Hybrid dynamical systems," *IEEE Control Systems*, vol. 29, no. 2, pp. 28–93, 2009.

[6] R. L. Chatburn, M. El-Khatib, and E. Mireles-Cabodevila, "A taxonomy for mechanical ventilation: 10 fundamental maxims," *Respiratory Care*, vol. 59, no. 11, pp. 1747–1763, 2014.

[7] T. P. Laubscher, W. Heinrichs, N. Weiler, G. Hartmann, and J. X. Brunner, "An adaptive lung ventilation controller," *IEEE Transactions on Biomedical Engineering*, vol. 41, no. 1, pp. 51–59, 1994.

[8] J. Reinders, B. Hunnekens, F. Heck, T. Oomen, and N. van de Wouw, "Adaptive control for mechanical ventilation for improved pressure support," *IEEE Transactions on Control Systems Technology*, vol. 29, no. 1, pp. 180–193, 2021.

[9] A. W. Thille, P. Rodriguez, B. Cabello, F. Lellouche, and L. Brochard, "Patient-ventilator asynchrony during assisted mechanical ventilation," *Intensive Care Medicine*, vol. 32, no. 10, pp. 1515–1522, 2006.

[10] *Medical electrical equipment. Particular requirements for basic safety and essential performance of critical care ventilators*, BSI British Standards Standard BS EN ISO 80601-2-12:2023, Nov. 2023.

[11] R. M. Kacmarek, "The mechanical ventilator: Past, present, and future," *Respiratory Care*, vol. 56, no. 8, pp. 1170–1180, 2011.

[12] P. Drinker and L. A. Shaw, "An apparatus for the prolonged administration of artificial respiration: I. A design for adults and children," *Journal of Clinical Investigation*, vol. 7, no. 2, pp. 229–247, 1929.

[13] N. R. MacIntyre, "Respiratory function during pressure support ventilation," *Chest*, vol. 89, no. 5, pp. 677–683, 1986.

[14] M. Younes, "Proportional assist ventilation, a new approach to ventilatory support: Theory," *American Review of Respiratory Disease*, vol. 145, no. 1, pp. 114–120, 1992.

[15] D. Dreyfuss, P. Soler, G. Basset, and G. Saumon, "High inflation pressure pulmonary edema: Respective effects of high airway pressure, high tidal volume, and positive end-expiratory pressure," *American Review of Respiratory Disease*, vol. 137, no. 5, pp. 1159–1164, 1988.

[16] Y. Yamada and H.-L. Du, "Analysis of the mechanisms of expiratory asynchrony in pressure support ventilation: a mathematical approach," *Journal of Applied Physiology*, vol. 88, no. 6, pp. 2143–2150, 2000.

[17] J. J. Marini, "Dynamic hyperinflation and auto–positive end-expiratory pressure: Lessons learned over 30 years," *American Journal of Respiratory and Critical Care Medicine*, vol. 184, no. 7, pp. 756–762, 2011.

[18] G. A. Iotti, A. Braschi, J. X. Brunner, T. Smits, M. Olivei, A. Palo, and R. Veronesi, "Respiratory mechanics by least squares fitting in mechanically ventilated patients: Applications during paralysis and during pressure support ventilation," *Intensive Care Medicine*, vol. 21, no. 5, pp. 406–413, 1995.

[19] Z.-L. Chen, Y.-Z. Yan, H.-Y. Yu, Q.-B. Wang, W. Wang, and M. Zhong, "Influence of compliance and resistance of the test lung on the accuracy of the tidal volume delivered by the ventilator," *BMC Pulmonary Medicine*, vol. 24, no. 1, p. 498, 2024.

[20] U.S. Food and Drug Administration, "Pneumotachometer," Code of Federal Regulations, Title 21, Section 868.2550, revised as of April 1, 2024, U.S. Government Publishing Office, govinfo.gov, 2024.

[21] S. N. Saavedra, P. V. Sepúlveda Barisich, J. B. Parra Maldonado, R. B. Lumini, A. Gómez-González, and A. Gallardo, "Asynchronies during invasive mechanical ventilation: Narrative review and update," *Acute and Critical Care*, vol. 37, no. 4, pp. 491–501, 2022.

[22] U.S. Food and Drug Administration, "Continuous ventilator," Code of Federal Regulations, Title 21, Section 868.5895, revised as of April 1, 2024, U.S. Government Publishing Office, govinfo.gov, 2024.

[23] D. Shyu, P. Bliss, A. Adams, and R. J. Cho, "Development and performance evaluation of a solenoid valve assisted low-cost ventilator on gas exchange and respiratory mechanics in a porcine model," *PLOS ONE*, vol. 19, no. 5, p. e0303443, 2024.

[24] W. P. King, J. Amos, M. Azer *et al.*, "Emergency ventilator for COVID-19," *PLOS ONE*, vol. 15, no. 12, p. e0244963, 2020.

[25] P. von Platen, A. Pomprapa, B. Lachmann, and S. Leonhardt, "The dawn of physiological closed-loop ventilation: A review," *Critical Care*, vol. 24, no. 1, p. 121, 2020.

[26] American Society of Anesthesiologists, "Standards for basic anesthetic monitoring," Standard of the American Society of Anesthesiologists, approved 1986, last amended October 15, 2025, asahq.org, 2025.

[27] T. P. Rietveld, B. J. P. van der Ster, A. Schoe, H. Endeman, A. Balakirev, D. Kozlova, D. A. M. P. J. Gommers, and A. H. Jonkman, "Let's get in sync: Current standing and future of AI-based detection of patient-ventilator asynchrony," *Intensive Care Medicine Experimental*, vol. 13, no. 1, p. 39, 2025.

4.10 Exercises

4.1 ⌕ Concept check · 1 min

In case 2 the ventilator delivers 420 mL in volume control at a PEEP of 8 cmH_2O, and at 30 s the lung stiffens from 50 to 18 mL/cmH_2O. What happens with the protective layer off?

(a) The volume falls to about 179 mL and the plateau stays near 18 cmH_2O

(b) The volume stays near 420 mL and the plateau climbs past the 30 cmH_2O limit

(c) The flow servo saturates, so the volume and the plateau both fall

(d) The volume falls and the plateau climbs, since the servo splits the error

4.2 ⌕ Concept check · 2 min

In case 6, after the bronchospasm, the lung has $R = 20\,cmH_2O\,s/L$ and $C = 50\,mL/cmH_2O$, and each expiration lasts 1.50 s. Which single change of setting lowers the true auto-PEEP?

(a) Raising PEEP from 5 to 10 cmH_2O, which holds the airways open in expiration

(b) Turning on the protective layer, which caps the plateau at 30 cmH_2O

(c) Lowering the rate, which gives each expiration more time constants to empty

(d) Raising the tidal volume, so that each breath clears the lung more fully

4.3 Concept check · 2 min

In case 7, from 160 s, a patient with 5 cmH_2O efforts triggers the volume-controlled breaths. How does the per-breath least-squares fit err on the breaths the patient triggers?

(a) It scatters at random about the truth, so the smoothed estimate stays unbiased

(b) It reads the resistance high and the compliance low, as a narrowed airway would

(c) It does not err, because each effort ends before the flow of the breath begins

(d) It reads the resistance low and the compliance high, by about the same amount on every breath

4.4 Theory and calculation · 6 min

Use Equation (4.5) for a passive lung with $R = 10\,cmH_2O\,s/L$.

(a) Volume control delivers 420 mL at a PEEP of 8 cmH_2O. Find the plateau at $C = 50$ and at $C = 18\,mL/cmH_2O$. *Checkpoint:* 16.40 cmH_2O and 31.33 cmH_2O.

(b) Pressure control holds 10 cmH_2O above PEEP for $T_i = 1.0$ s with a rise time $T_r = 0.1$ s. Find the tidal volume at the same two compliances. *Checkpoint:* 425.2 mL and 179.1 mL.

(c) The challenge runs volume control at 420 mL and a PEEP of 10 cmH_2O. Find the compliance below which the plateau exceeds 30 cmH_2O, and say whether the stiffening to 25 mL/cmH_2O alone can fail the plateau target. *Checkpoint:* 21 mL/cmH_2O, so it cannot.

4.5 Theory and calculation · 5 min

Case 6 after the spasm has $R = 20\,cmH_2O\,s/L$, $C = 50\,mL/cmH_2O$, volume control at $V_T = 420$ mL and $T_i = 1.0$ s with no pause.

(a) Find RC, the expiratory time at 24 min^{-1} and the trapped pressure from Equation (4.6). *Checkpoint:* 1.00 s, 1.50 s and 2.41 cmH_2O.

(b) Repeat the trapped pressure at 18 min^{-1}. *Checkpoint:* 0.90 cmH_2O.

(c) The engine measures 2.58 and 0.96 cmH_2O. Find by how much it exceeds the closed form in each case, and give the reason. *Checkpoint:* About 7 and 6 %.

4.6 Theory and calculation · 7 min

The orifice law of Equation (4.3) is $\psi(\Delta p) = \Delta p/\sqrt{|\Delta p| + 0.1}$ with pressures in cmH_2O. The blower holds 60 cmH_2O, the fully open inspiratory valve passes 180 L/min into an open airway, and a leak is entered as its flow at 10 cmH_2O.

(a) Find the flow of the fully open inspiratory valve into a circuit held at 8 cmH_2O. *Checkpoint:* 168 L/min.

(b) Find the flow that a leak entered as 300 L/min takes at 8 cmH_2O. *Checkpoint:* 268 L/min.

(c) Find the largest leak, entered at 10 cmH_2O, against which a PEEP of 8 cmH_2O can still be held, and say what leak compensation adds above it. *Checkpoint:* 187.6 L/min. Nothing.

4.7 App challenge · 8 min

Load scenario 2, Volume control and a stiffening lung. Use the default seed and the speed 5×, and run each case to 120 s.

(a) Run the scenario as shipped, with the Lung-protective layer off. Read the true plateau over the last 10 s in the exported CSV column `pplat_true_cmH2O`, and the tile Highest plateau (true). *Checkpoint:* 31.3 cmH_2O, within 0.1 cmH_2O. Tile 31.5 cmH_2O.

(b) Turn on the Lung-protective layer, press Reset and run again. Read `vt_true_mL` and `pplat_true_cmH2O` over the last 10 s, and the same tile. *Checkpoint:* 395 to 396 mL at 29.9 to 30.0 cmH_2O. Tile 31.5 cmH_2O.

Explain why the tile reads the same in both runs.

4.8 App challenge · 7 min

Load scenario 6, Bronchospasm and auto-PEEP. Use the default seed and the speed 5×, run each case to 120 s and read the tiles Highest auto-PEEP (true) and Highest plateau (true). Reload the case with the reload button before each change, and make the change before Run.

(a) Run the scenario as shipped, at 24 min^{-1}. *Checkpoint:* 2.58 cmH_2O and 16.0 cmH_2O.

(b) Set Rate to 18, then to 14 min^{-1}. *Checkpoint:* 0.96 and 0.35 cmH_2O.

(c) Leave the rate at 24 and set PEEP to 10 cmH_2O. *Checkpoint:* 2.58 cmH_2O and 21.0 cmH_2O.

Accept the auto-PEEP within 0.02 cmH_2O, and compare the results with the closed form.

4.9 App challenge · 8 min

Load the challenge, an hour on the ventilator. Use the default seed and the speed 60×, and run the full hour each time. Read the tiles Breaths within ±10 %, Highest auto-PEEP (true), the Audit and End-tidal CO_2.

(a) Run the challenge as shipped, in volume control. *Checkpoint:* 540 of 1975, 27.3 %. 6.93 cmH_2O. Audit 0, 1211, 0. 111.0 mmHg.

(b) Reload, and before Run set Mode to pressure support, Leak compensation on and Adaptive support on. Run the hour again. *Checkpoint:* 1157 of 1160, 99.7 %. 1.95 cmH_2O. Audit 1, 2, 0. 29.3 mmHg.

Say which part of the target the first run fails, and what each of the two switches in the second run contributes.

4.10 Application area · 4 min

The app's lung-protective values come from the ARDS Network trial. Take a predicted body weight of 70 kg.

(a) Find the tidal volume at 6 mL/kg and the volume cap at 8 mL/kg. *Checkpoint:* 420 mL and 560 mL.

(b) The trial reported a mortality of 31.0 % against 39.8 %. Find the absolute difference and the number of patients treated for each death avoided. *Checkpoint:* 8.8 points, about 11 patients.

(c) In case 2 after the stiffening, the plateau is 31.3 at a PEEP of 8 cmH_2O. Find the driving pressure, and compare it with the 15 cmH_2O line the app draws. *Checkpoint:* 23.3 cmH_2O, 8.3 cmH_2O above the line.

4.11 Application area · 4 min

The carbon dioxide of the app is one store with a production of 200 mL/min, a capacity of 45 mL/mmHg, a dead space of 150 mL and $K = 0.863$ in Equation (4.4).

(a) Find the alveolar ventilation and the steady CO_2 for 420 mL at 16 min^{-1}. *Checkpoint:* 4.32 L/min and 39.95 mmHg.

(b) Find how fast the CO_2 rises in apnea. *Checkpoint:* 4.44 mmHg a minute.

(c) In the challenge as shipped the CO_2 reaches 111.0 mmHg. Name two responses a real patient would show that the app's patient cannot.

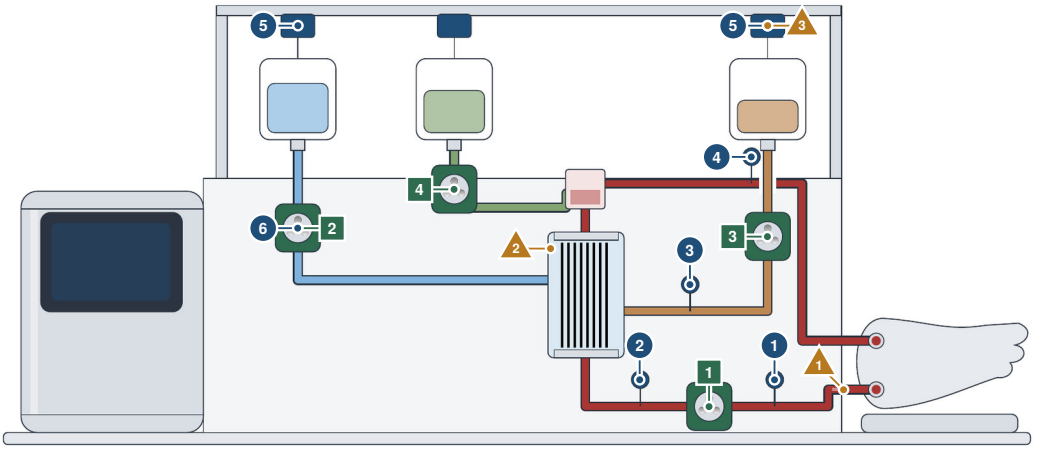
4.12 Review the agent · 5 min

An AI assistant is shown the challenge at minute 45, run as shipped, and writes: "The monitor reads a tidal volume of 418 mL against the 420 mL target and a plateau of 18.5 cmH_2O, far below 30. The identified compliance is 70.6 mL/cmH_2O, up from 50 at the start, so the lung is improving. The leak

reads only 11.2 L/min. Keep volume control, and consider lowering PEEP as the lung recovers.” Find the flaw and say how to check it.

CHAPTER 5

Continuous Kidney Replacement Simulator for Cumulative Balance and Measurement Redundancy



The controller moves 1 blood pump, draws blood through the filter 2 dialysate pump, feeds the filter jacket 3 effluent pump, takes the used fluid and the water 4 replacement pump, returns fluid to the blood

The controller reads 1 pressure in the access line 2 pressure before the filter 3 pressure in the effluent line 4 pressure in the return line 5 a scale holding each bag 6 the turns counted on every pump

What disturbs it 1 a narrowed catheter 2 a filter that clogs 3 a scale that drifts

The water the patient loses is what the effluent bag gains less what the two supply bags lose, and the scales and the pump counts measure the same flows twice. Illustration, not to scale.

IN PLAIN TERMS Weighing the Water a Machine Removes

The main idea. When the kidneys fail, a machine can clean the blood around the clock and take off extra water. It pumps liters of fluid in and out every hour. The water the patient loses is the small difference between the two.

The challenge. A tiny error in any pump grows large in that difference, so the machine weighs its fluid bags to check. If a scale slowly drifts, the machine trusts it and moves the patient off course while its own reading looks fine.

Why it matters. Over a day, those small errors decide how much water a very sick patient really loses.

THREE WAYS TO READ THIS CHAPTER

Explore	11 min	The first page, the fact box of Section 5.1, the history box of Section 5.2, the try-it boxes of Sections 5.4 and 5.6, the second fact box and the key points
Study	50 min	The whole chapter with its pause points and the twelve exercises
Practice	15 min	The first page, Section 5.6, Section 5.7, Section 5.8 and the open challenges

The control question The water a patient on continuous kidney replacement should lose is the small difference of liters an hour pumped in and out. What does a balance loop on a weighed running total remove, what does it pass to the patient while its own error stays small, and which second measurement shows the difference?

Learning objectives

- Compute how each pump's relative error is amplified in the net removal, and predict the removal error of a balance kept on pump counts.
- Derive the poles and the step response of a PI balance loop on an integrating running total.
- Explain why the balance loop follows a drifting scale, and size the smallest drift a plausibility check against the pump counts flags.
- Analyze how a shared, saturated effluent pump changes the sign of the dose loop's gain, and choose the structural remedy.
- Evaluate a circuit configuration over a day against removal, dose, pressure and ledger criteria.

Prerequisites Sections 1.6.2 (velocity form and back-calculation), 1.4.5 (redundancy), 1.4.6 (overrides on limits) and 1.6.4 (scenarios and the challenge).

The app https://apmonitor.com/apps/crrt-fluid-control/

See also Volume 4, Chapter 2 for a protective override that rests on its clamp and Volume 4, Chapter 12 for allocation of one limited resource. Volume 5, Chapter 13 for a running estimate that drifts until an independent fix arrives and Volume 5, Chapter 10 for a fast measurement corrected by a slower second one.

5.1 Introduction

When the kidneys fail in intensive care, continuous kidney replacement therapy removes waste solutes and water from the blood around the clock, often for days. The water matters as much as the solutes. In a multicenter observational study of critically ill adults with acute kidney injury, fluid overload of more than 10 % of baseline body weight was independently associated with mortality, although the study left open whether the overload caused the harm or marked a sicker patient [1]. The machine that delivers the therapy never weighs the patient. It knows the removal only through its own pumps and bags, and the removal it reports is a calculation on them.

DID YOU KNOW? Prescribed Is Not Achieved

In 500 critically ill adults treated over 3630 patient-days, with data from the health record and from the machines themselves, the therapy stood paused for a median 8.1 % of its time, and the achieved fluid balance missed the prescribed one by a median 17.4 mL per kilogram a day [2].

That calculation is a small difference between large numbers. In the first case of the simulator the blood pump counts 150 mL/min, or 9 L/h, through the filter, the dialysate and replacement pumps count 930 mL/h and 1000 mL/h into the circuit, the effluent pump counts 2030 mL/h out of it, and the patient is meant to lose 100 mL/h. An error in any of these flows reaches the removal whole, so each pump's relative error is amplified by the ratio of its flow to the removal, 20.3 for the effluent pump, 10.0 for the replacement pump and 9.3 for the dialysate pump. Roller pumps add a bias of their own, since they count revolutions and deliver less than the count when their inlet is under suction [3]. The tubing tolerances of the simulated circuit make the true removal 88.8 mL/h at reset while the pumps count 100 mL/h, and a balance kept on the pump counts leaves the patient 466 mL behind after a day.

Weighing the bags removes that bias and raises a second question. The simulator addresses the belief that a balance loop with a small error means a balanced patient. A loop that holds the weighed running total on its target rejects every pump error the scales can see, and it rejects a drifting scale just as thoroughly, by moving the patient until the reading is right. With the effluent scale drifting 60 g/h, 3 % of the effluent flow, the loop keeps its own measured error within 28.4 mL while the

true removal falls 304 mL behind the prescription in 6 h and 648 mL in 12 h. A measurement cannot check itself, and only an independent measurement of the same flow, here each pump's own count, reveals the drift, the redundancy of Section 1.4.5. A survey vehicle meets the same limit when its dead-reckoned position drifts until an acoustic fix arrives, as Volume 5, Chapter 13 shows.

Many plants share this shape. When a small net quantity is computed from large measured streams, the closure of the balance and the gross errors of individual meters decide what the result is worth, and adjusting redundant measurements to close material balances is the subject of data reconciliation in process plants [4]. A loss-in-weight feeder that holds its weighed flow but not the concentration of its product, in Volume 12, Chapter 12, is another case. A sensor fault is detected by comparing the sensor with a model or a second measurement of the same quantity, the basis of model-based fault detection [5]. Equipment limits that take a manipulated variable away from the loop that normally owns it are handled with override and selector control [6], and dialysis already uses feedback on a patient measurement to shape the removal, as when blood-volume tracking adjusts the weight-loss rate of intermittent hemodialysis [7]. However, a transfer function from pump speed to removal rate hides the running total, the subtraction and the instrument, and those three decide what a small error becomes by the end of a day.

This chapter uses a simulated continuous venovenous hemodiafiltration circuit, whose fluid balance is a running total assembled from four imperfect roller pumps and three scales, to show how the error each instrument makes reaches the patient while the loops see only what the instruments report. The therapy, gravimetric balancing, override control and plausibility checks are all in clinical or industrial use. The circuit model keeps its truth apart from its instruments, and its fluid ledger closes within 3×10^{-9} mL through flow, pauses and bag changes. Eight cases and a day-long challenge price a measurement, an override, a shared actuator and a pause in milliliters of removal error. The application is an educational equipment-control simulation. Its patient model, its filter coefficients and every clinical threshold it displays are illustrative and have not had clinical review, and nothing in it or in this chapter is clinical guidance.

Section 5.2 reviews how continuous therapy and its automatic balancing developed. Section 5.3 gives the circuit model, from the roller pumps to the scales and the ledger. Section 5.4 analyzes the balance as a small difference of large flows: how pump errors reach it, why the balance loop sees an integrator, why it follows a drifting scale and why the effluent pump is shared by three loops. Section 5.5 states the controller as implemented, and Section 5.6 reports the eight cases and the challenge. Section 5.7 describes the real scales, pumps and controllers, Section 5.8 states the limitations and Section 5.9 draws the results together.

5.2 Continuous Therapy and Automatic Balancing

Continuous therapy began as a way to remove water. In 1977 an arteriovenous hemofiltration circuit, perfused at about 100 mL/min by the patient's own arteriovenous pressure gradient with no blood pump, was reported to withdraw 200 to 600 mL/h from over-hydrated patients who were resistant to diuretics [8]. How the machines of that period came to balance the fluid themselves, first by the height of a bag and then by scales and pumps, is told in the history box of this section.

> ⧖ HISTORY **From the Height of a Bag to Scales (1977 to 1978)**
>
> The first continuous arteriovenous hemofiltration was given in 1977, in an intensive care unit in Göttingen, at a time when acute kidney failure was treated mostly with peritoneal dialysis or hemodialysis [9]. The removal was first controlled by hand. Hanging the filtrate bag higher or lower changed the suction of the fluid column below the filter, and the replacement fluid was also set by hand. Later systems balanced the fluid automatically, by gravity, with scales and electronic clamps, and peristaltic pumps soon replaced those simple mechanisms [9]. A clinical report of 1978 found automatic balancing accurate enough to replace the bed scales that conventional hemodialysis needed for patients confined to bed, and the constant weight loss seemed well tolerated by the patients [10].

Diffusion joined the continuous circuit a decade later. Continuous hemodialysis passed dialysate at 16.6 mL/min through a small flat-plate dialyzer in a pumpless circuit, and over blood flows from 70 to 190 mL/min the dialysate left in equilibrium with the blood, so the total clearance equaled the dialysate outflow of about 25 mL/min [11]. That equilibrium is why the clearance of a continuous therapy can be set by a fluid flow, and the diffusive term of the simulator reproduces it on a new filter. The limitations of the arteriovenous circuit led to pumped venovenous hemofiltration and hemodialysis, and collaboration with industry produced dedicated machines of greater accuracy [9]. Pumps brought an error of their own. In hemodialysis the blood flow measured by an ultrasonic sensor was 11.6 to 14.8 % below what the roller pump indicated at settings from 300 to 400 mL/min, and the shortfall followed the negative pressure in the arterial line [3].

The prescription became a rate per kilogram. A trial published in 2000 randomized 425 patients among ultrafiltration doses of 20, 35 and 45 mL/h per kilogram [12], and a multicenter trial in 2009 compared effluent flows of 40 and 25 mL/h per kilogram in postdilution continuous venovenous hemodiafiltration [13], the configuration simulated here. Both defined the dose by a fluid flow per kilogram. A machine delivers such a dose only while it runs and only as far as its filter still clears, which is why the simulator prescribes an effective dose as urea clearance per kilogram and scores its time average with the downtime included.

The patient side has dynamics of its own. Ultrafiltration takes water from the plasma, and the interstitium refills it at a rate described by Starling's hypothesis, with a filtration coefficient fitted to online blood-volume measurements during hemodialysis [14]. The circuit side has limits. Premature clotting shortens the life of the circuit, and the settings of the therapy, partial predilution and an individually controlled filtration fraction among them, are part of keeping it open [15]. A consensus on precision fluid management now frames fluid management during the therapy around three linked goals, the patency of the circuit, electrolyte and acid-base homeostasis and the regulation of the patient's fluid balance [16]. The simulator covers the first and the third and leaves electrolytes and acid-base out.

5.3 Circuit Model

The simulated circuit is a continuous venovenous hemodiafiltration with postdilution replacement on an illustrative 80 kg patient. Blood is drawn through the access lumen of a catheter by a roller pump, rises through the fibers of a hollow-fiber filter and returns through the return lumen, and the replacement fluid joins it in the venous chamber. Dialysate enters the filter jacket at the blood outlet end and flows counter-current to the blood, and the effluent pump draws the jacket out into a bag on a scale. Every flow, pressure, instrument and disturbance in the equations below is placed in Figure 5.1, and Table 5.1 classifies them.

5.3.1 Roller Pumps and Their Counts

Each roller pump delivers its stroke volume per revolution, reduced by its tubing's loading error, its wear and any suction at its inlet, as shown in Equation (5.1).

$$\begin{aligned} s_k &= V_{s,k}\,(1+e_{0,k})\max(0.5,\ 1-w_k N_k), \qquad \frac{\mathrm{d}N_k}{\mathrm{d}t} = n_k \\ Q_b &= s_b n_b (1+k_{in,b} P_{acc}), \qquad Q_d = s_d n_d, \qquad Q_r = s_r n_r \\ Q_e &= s_e n_e \max\bigl(0,\ 1+k_{in,f}\min(0, P_{eff})\bigr) \end{aligned} \tag{5.1}$$

where k is the blood (b), dialysate (d), replacement (r) or effluent (e) pump, $V_{s,k}$ the nominal stroke, $e_{0,k}$ the stroke error of the segment as loaded, w_k the stroke lost per revolution, N_k the revolutions turned, n_k the speed and k_{in} the under-filling per unit of inlet suction. The supply bags hang above their pumps, so only the blood and effluent pumps draw against suction. A supply pump delivers nothing once its bag is empty, and the effluent pump nothing once its bag reaches its physical limit.

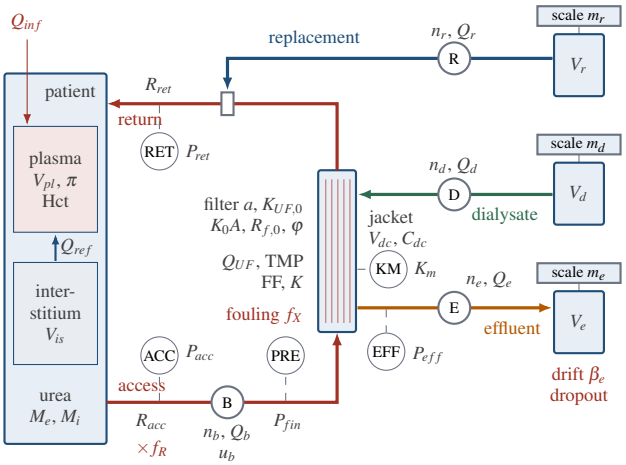


Figure 5.1 Continuous venovenous hemodiafiltration circuit as modeled. Blood leaves the patient through the access lumen of resistance R_{acc} (scaled by the restriction f_R), passes the access transducer ACC, the blood pump B and the pre-filter transducer PRE, rises through the fibers of a filter with open area a, and returns past the venous chamber and the return transducer RET. The dialysate pump D feeds the jacket counter-current to the blood, the effluent pump E draws the jacket (stored volume V_{dc}, pressure P_{eff} read by EFF) into the effluent bag, and the replacement pump R returns fluid to the venous chamber, or before the filter when predilution is chosen (not drawn). Each pump's encoder counts its revolutions N_k, each bag hangs on a scale reading m_k, and the idealized monitor KM reports the urea clearance. The patient is an illustrative plasma and interstitium with urea pools.

Table 5.1 Roles of the Circuit Variables at the Reference Configuration

Variable	Role	Reference	Range
$V_{net} - V_{Rx}$	CV, removal error	0 mL	±60 band
Q_b	CV, by pump count	150 mL/min	50–300
P_{acc}	constraint, override	−150 mmHg	−240 to −60
TMP	constraint, override	200 mmHg	100–300
FF	constraint, ceiling	25 %	10–35
D	CV, clearance per kg	25 mL/(kg h)	10–50
u_b, u_d, u_r, u_e	MVs, pump drives	%	0–100
$Q_{e,cmd}$	MV, balance	2030 mL/h	0–4500
Q_d, Q_r	MVs, dose, allocation	930, 1000 mL/h	0–4000
$Q_{net,Rx}$	prescription	100 mL/h	0–500
f_R, f_X	DV, restriction, fouling	1, 1	1–20, 0–20
β_d, β_r, β_e	DV, scale drift	0 g/h	±200
Q_{inf}	DV, other infusion	0 mL/h	0–200
n_k, N_k	states, pumps	4 each	—
V_{dc}, a	states, jacket, area	−0.21 mL, 1	—
V_d, V_r, V_e	states, bags	3.2, 4.1, 1.3 L	—
V_{pl}, V_{is}	states, patient water	3.2, 13 L	—
M_e, M_i	states, urea	30 mmol/L each	—
ledgers	states, integrals	9	—

Each drive follows its command through a lag and droops against the pressure it pumps into, as shown in Equation (5.2).

$$\tau_m \frac{\mathrm{d}n_k}{\mathrm{d}t} = n_{max,k} \frac{u_k}{100} \max\left(0.5,\ 1 - \delta \max(0, \Delta P_k)\right) - n_k \tag{5.2}$$

where the drive command u_k is in percent, the full speed is $n_{max,k} = Q_{max,k}/V_{s,k}$, δ is the droop and ΔP_k the pressure rise across the pump, $P_{fin} - P_{acc}$ for the blood pump, $P_{eff} - P_{head}$ for the dialysate

pump, $P_{ret} - P_{head}$ for the replacement pump and $-P_{eff}$ for the effluent pump, with P_{head} the head of the hanging bags. The machine's pump count is $V_{s,k}\hat{n}_k$, the nominal stroke times the speed its encoder measures over the last control interval, and it knows nothing of $e_{0,k}$, w_k or k_{in}. The loading errors of +0.4 % on the dialysate segment, −0.3 % on the replacement segment and −0.5 % on the effluent segment account for 10.9 mL/h of the 11.2 mL/h by which the true removal falls short of the count at reset, and the slight suction at the effluent pump for the rest.

5.3.2 Access, Filter and Return Pressures

The access lumen is a resistance in series with the blood pump, and Equation (5.3) solves the pump's inlet pressure and its delivery together.

$$P_{acc} = P_{cv} - f_R R_{acc} Q_b, \qquad Q_b = \frac{s_b n_b (1 + k_{in,b} P_{cv})}{1 + s_b n_b k_{in,b} f_R R_{acc}} \tag{5.3}$$

where P_{cv} is the central venous pressure and f_R the access restriction. Below the collapse pressure P_{floor} the line collapses, P_{acc} is held at P_{floor} and $Q_b = (P_{cv} - P_{floor})/(f_R R_{acc})$. At reset the blood pump counts 150 mL/min, delivers 144.7 mL/min and holds the access at −76.9 mmHg.

Filtration across the membrane follows a Starling law whose conductance falls as the filtered plasma concentrates [17], and the pressures that drive it follow from the flows, as shown in Equation (5.4).

$$\begin{aligned} Q_{UF} &= K_{UF,0}\, a \Big(1 - \frac{Q_{UF}}{\varphi Q_p}\Big)(\mathrm{TMP} - \pi) \\ \mathrm{TMP} &= \tfrac{1}{2}(P_{fin} + P_{ret}) - P_{eff}, \quad \pi = \pi_0 V_{pl,0}/V_{pl} \\ P_{ret} &= P_{cv} + R_{ret}(Q_b + Q_r - Q_{UF}) \\ P_{fin} &= P_{ret} + \tfrac{1}{2}(R_{f,0}/a)(2Q_{fin} - Q_{UF}) \\ Q_p &= Q_b(1 - \mathrm{Hct}) + Q_{r,pre}, \quad \mathrm{Hct} = V_{rbc}/(V_{rbc} + V_{pl}) \end{aligned} \tag{5.4}$$

In (5.4), a is the open share of the fibers, $K_{UF,0}$ and $R_{f,0}$ the ultrafiltration coefficient and blood-side resistance of the new filter, φ the filtration fraction at which the concentrated plasma stops filtering, π the oncotic pressure, $Q_{fin} = Q_b + Q_{r,pre}$ the blood entering the filter, $Q_{r,pre}$ the replacement flow when it enters before the filter and zero otherwise, and $V_{rbc} = V_{pl,0}\mathrm{Hct}_0/(1 - \mathrm{Hct}_0) = 1.37\,\mathrm{L}$ the fixed red-cell volume. Substituting the pressures gives $\mathrm{TMP} - \pi = A - R_p Q_{UF}$, with $A = P_{cv} + R_{ret}(Q_b + Q_r) + R_{f,0} Q_{fin}/(2a) - P_{eff} - \pi$ and $R_p = R_{ret} + R_{f,0}/(4a)$, so the first line of (5.4) is a quadratic in Q_{UF} whose smaller root is taken in closed form. For $A \le 0$ the membrane back-filters without throttling, $Q_{UF} = K_{UF,0} a A/(1 + K_{UF,0} a R_p)$, and with no plasma flow nothing filters. At reset TMP is 77.1 mmHg, the pre-filter and return pressures are 94.3 and 58.1 mmHg, 1086 mL/h crosses the membrane and the filtration fraction $\mathrm{FF} = Q_{UF}/Q_p$ is 17.9 %.

5.3.3 The Dialysate Compartment

The pumps impose the ultrafiltration, and the membrane settles at whatever pressure passes it. The jacket and the effluent line store fluid in a compliance, as shown in Equation (5.5).

$$\frac{\mathrm{d}V_{dc}}{\mathrm{d}t} = Q_d + Q_{UF} - Q_e, \qquad P_{eff} = \frac{V_{dc}}{C_{dc}} \tag{5.5}$$

where V_{dc} is the volume stored above the unpressurized jacket and C_{dc} the compliance. Linearizing (5.5) about the reset point gives the time constant $\tau_{dc} = C_{dc}/(\partial Q_{UF}/\partial(-P_{eff}) + \partial Q_e/\partial P_{eff})$. On the new filter the membrane passes 14.0 mL/h more for each millimeter of mercury drawn from the jacket, half the nominal coefficient because of the throttling and the return-pressure coupling in (5.4), the effluent pump's under-filling adds 0.4 mL/h per millimeter of mercury, and $\tau_{dc} = 62.4\,\mathrm{s}$. With the same pump flows at 15 % open area it is 326.6 s. Both are short against the hours over which the balance is judged.

5.3.4 Clearance and Fouling

Urea leaves the blood by diffusion into the dialysate and by convection with the ultrafiltrate, as shown in Equation (5.6).

$$\begin{aligned} K &= \min\{\lambda\,[Q_d(1-\mathrm{e}^{-K_0 A\,a/Q_d}) + S\,Q_{UF}^{+}],\ 0.9\,Q_{pl}\} \\ \lambda &= Q_{pl}/(Q_{pl}+Q_{r,pre}) \end{aligned} \tag{5.6}$$

where λ is the dilution by predilution fluid, one in postdilution, $Q_{pl} = Q_b(1-\mathrm{Hct})$ the plasma flow, K_0A the urea mass-transfer coefficient times area of the new filter, S the sieving coefficient and Q_{UF}^{+} the forward ultrafiltration. The effective dose is $D = K/W$ with W the weight. The diffusive term is the counter-current exchanger in the limit of a blood flow much larger than the dialysate flow, and it approaches Q_d when $K_0A\,a \gg Q_d$, the equilibrium seen in the first continuous hemodialysis [11]. At reset the clearance is 33.3 mL/min, of which 15.2 mL/min is diffusive and 18.1 mL/min convective, a dose of 25.0 mL/(kg h). The open area falls at a baseline fouling rate, faster at a high filtration fraction and while the blood stands still, as shown in Equation (5.7).

$$\begin{aligned} \frac{\mathrm{d}a}{\mathrm{d}t} = -a\big[&k_0 f_X + k_{FF}\max(0,\ \mathrm{FF}-0.2) \\ &+ k_{st}\max(0,\ 1-Q_b/Q_{st})\big] \end{aligned} \tag{5.7}$$

while $a > a_{min}$, where f_X is the fouling disturbance and $Q_{st} = 30\,\mathrm{mL/min}$. The coefficients K_{UF}, K_0A and $1/R_f$ of the filter all scale with a.

5.3.5 The Illustrative Patient

The patient is two fluid compartments and two urea pools, as shown in Equation (5.8), and it is illustrative.

$$\begin{aligned} \frac{\mathrm{d}V_{pl}}{\mathrm{d}t} &= Q_r - Q_{UF} + Q_{ref} + Q_{inf} \\ \frac{\mathrm{d}V_{is}}{\mathrm{d}t} &= -Q_{ref}, \qquad Q_{ref} = \frac{V_{is} - rV_{pl}}{\tau_{ref}(1+r)} \\ \frac{\mathrm{d}M_e}{\mathrm{d}t} &= G + K_c(C_i - C_e) - KC_e \\ \frac{\mathrm{d}M_i}{\mathrm{d}t} &= -K_c(C_i - C_e) \end{aligned} \tag{5.8}$$

where V_{pl} and V_{is} are the plasma and interstitial volumes, $r = V_{is,0}/V_{pl,0} = 4.06$ the ratio toward which the refill works, τ_{ref} the refill time constant, Q_{inf} any other infusion, M_e and M_i the urea in the extracellular and intracellular pools, $C_e = M_e/(V_{pl}+V_{is})$ and $C_i = M_i/V_{ic}$ their concentrations, G the urea generation and K_c the intercompartment clearance. The first line makes any departure from the starting ratio decay with time constant τ_{ref}, a first-order stand-in for the Starling refill that has been fitted to blood-volume data [14]. The patient exchanges fluid with the circuit directly, and the extracorporeal blood volume is not a state.

5.3.6 Bags, Scales and the Ledger

The bags integrate the pump flows, $\dot{V}_d = -Q_d$, $\dot{V}_r = -Q_r$ and $\dot{V}_e = Q_e$, and a bag change replaces a supply bag with a full one of volume V_{bag} or empties the effluent bag, recording what was taken down and what was hung. The scales read as shown in Equation (5.9),

$$m_k = \rho V_k + m_{tare} + b_k + \nu_k, \qquad \frac{\mathrm{d}b_k}{\mathrm{d}t} = \beta_k \tag{5.9}$$

rounded to 1 g, where $\rho = 1.00\,\mathrm{kg/L}$ for every fluid, m_{tare} is the bag and hook, b_k a bias that grows at the drift rate β_k and ν_k colored noise. A dropout event removes the reading altogether. The machine's net removal by the scales is the weight gained by the effluent bag less the weight lost by the two

Table 5.2 Model Parameters. Visible (vis.) Parameters Are Shown and Adjustable in the Model Panel, Hidden (hid.) Ones Are Not

Symbol	Meaning	Value	Class
W	weight, dose basis	80 kg	vis.
$V_{pl,0}$, $V_{is,0}$	plasma, interstitium	3.2, 13 L	vis.
Hct_0, τ_{ref}	hematocrit, refill	30 %, 40 min	vis.
P_{cv}, π_0	venous, oncotic	8, 25 mmHg	hid.
C_0, G	urea	30 mmol/L, 400 mmol/d	vis.
V_{ic}, K_c	urea pool, exchange	30 L, 600 mL/min	hid.
R_{acc}, R_{ret}	lumens	0.587, 0.35 mmHg min/mL	vis.
$R_{f,0}$	filter blood side, new	0.267 mmHg min/mL	vis.
P_{floor}	line collapse	−450 mmHg	hid.
$K_{UF,0}$	ultrafiltration, new	28 mL/(h mmHg)	vis.
K_0A	urea transfer, new	60 mL/min	vis.
φ, S, a_{min}	throttle, sieving, floor	0.70, 1.0, 0.03	hid.
k_0, k_{st}	fouling, stasis	2, 20 %/h	vis.
k_{FF}	per % of FF over 20	0.4 %/h	vis.
C_{dc}	jacket compliance	0.25 mL/mmHg	vis.
$V_{s,b}$, $V_{s,f}$	nominal strokes	7.5, 1.0 mL	vis.
$Q_{b,max}$	blood pump	450 mL/min	vis.
$Q_{d,max}$, $Q_{r,max}$	fluid pumps	4000, 4000 mL/h	vis.
$Q_{e,max}$	effluent pump	4500 mL/h	vis.
τ_m, δ	drive lag, droop	2 s, 3 % per 100 mmHg	vis.
$k_{in,b}$, $k_{in,f}$	under-filling	5, 2 % per 100 mmHg	vis.
w_b, w_f	wear	0.2, 0.15 % per 10^4 rev	vis.
$e_{0,d}$, $e_{0,r}$, $e_{0,e}$	loading	+0.4, −0.3, −0.5 %	vis.
$e_{0,b}$, P_{head}	blood error, bag head	+0.3 %, 20 mmHg	hid.
V_{bag}	supply and effluent bag	5 L	vis.
—	change levels	150, 4800 mL	vis.
t_{swap}	bag change	180 s	vis.
—	effluent limit, ramp	5.5 L, 30 s	hid.
m_{tare}, —	tare, resolution	250, 1 g	hid.
—	scale, pressure noise	2 g, 1.5 mmHg	vis.
—	transducer lag	1 s	vis.
—	monitor interval, noise	10 min, 3 %	vis.
W_c	check window	20 min	vis.
q_{abs}, q_{rel}	check tolerance	20 mL/h, 1 %	vis.
Hct_{ent}	entered hematocrit	30 %	vis.
—	TMP limit, stop	250, 350 mmHg	vis.
—	pre-filter warning, stop	300, 450 mmHg	vis.
—	access stop	−250 mmHg	vis.
ρ	every fluid	1.00 kg/L	fixed

supply bags, carried across each bag change by recording the final weight of the bag coming off and re-taring the new one. The true net removal obeys Equation (5.10).

$$V_{net} = \int_0^t (Q_e - Q_d - Q_r)\,\mathrm{d}t = -\Delta V_{pat} + V_{inf} - \Delta V_{circ} \tag{5.10}$$

where ΔV_{pat} is the change in $V_{pl} + V_{is}$, V_{inf} the other infusions and ΔV_{circ} the change in the circuit's stored volume, including what a discarded filter held. The prescribed trajectory $V_{Rx} = \int Q_{net,Rx}\,\mathrm{d}t$ is itself a state, and the removal error $V_{net} - V_{Rx}$ is what the scorecard judges. The ledger of patient, circuit, hanging bags, discarded bags and discarded filters, less the bags hung and the other infusions, stays constant within 3×10^{-9} mL in every run reported here.

5.3.7 Initialization and Numerics

Table 5.2 gives the parameters, and every clinical limit among them is illustrative. Every case starts with the treatment under way. Each pump turns at the speed that gives its prescribed flow by pump count, the jacket volume is found by bisection where $\dot{V}_{dc} = 0$, the filter is new, the bags hold 3.2, 4.1 and 1.3 L, and the patient sits at refill equilibrium with urea at C_0 in both pools. The pump strokes are

Table 5.3 Removal Error of Case 1 Over 8 h Against the Stroke Error of the Effluent Segment

Stroke error (%)	−1.0	−0.5[a]	0	+0.5	+1.0
Pump count, at 8 h (mL)	−184.6	−108.4	−33.1	+42.9	+117.6
Scales, worst (mL)	−9.7	−13.8	−10.2	−10.6	−9.5

[a] the loaded segment of the application

taken as already learned against the scales on earlier bags. Twenty-six states are integrated together by classical fourth-order Runge-Kutta at 0.5 s. They are the four speeds and four revolution counts, V_{dc}, a, the three bags, V_{pl}, V_{is}, M_e, M_i, five flow ledgers, V_{Rx} and the integrals of clearance, urea removed and urea generated. Flows and pressures are algebraic in the states, and bag swaps, filter changes and the hard limits of a full or dry bag are discrete events after each step. With the instruments noise-free, halving the step changes the removal error of case 1 at 4 h by 0.002 mL and its TMP by 0.0001 mmHg, and the TMP of case 3 at 6 h by 0.0002 mmHg, far inside the application's own check of 1 mL, 0.5 and 1 mmHg.

5.3.8 Assumptions

The patient is illustrative. It has two fluid compartments with a first-order refill and two urea pools with constant generation, and no blood pressure, tolerance of fluid removal, electrolytes, acid-base or drug clearance, so the model is silent wherever the patient's response to removal matters.

The membrane is lumped. Its filtration coefficient, urea transfer and blood-side conductance are proportional to one open-area fraction whose loss is prescribed. Real fouling is a protein layer and clot that also changes the sieving, and only urea is followed, with a sieving coefficient of 1. The concentration of the plasma along the fibers is a throttling factor rather than a distributed model, and the oncotic pressure follows the plasma protein alone.

The pumps are linear. Delivery is proportional to speed, wear to revolutions and under-filling to suction, where real tubing need not behave linearly at high suction. All fluids weigh 1.00 kg/L, so a scale reading in grams is a volume in milliliters. The clearance monitor reports the true clearance with noise, whereas real circuits more often infer clearance from blood and effluent samples. Anti-coagulation, the extracorporeal blood volume and the saline that primes a new filter are outside the ledger, and the blood is taken as returned before every filter change.

5.4 A Small Difference of Large Flows

Four properties of the balance follow from the model, and each shapes the controller.

5.4.1 How Pump Errors Reach the Removal

At the rates of case 1 the removal responds to a relative error ε_k in pump k as $\partial Q_{net}/\partial \varepsilon_k = \pm Q_k$, and relative to the prescription the amplification is $Q_k/Q_{net,Rx}$, 20.3 for the effluent pump, 10.0 for the replacement pump and 9.3 for the dialysate pump. A balance on the pump counts inherits every such error. Table 5.3 runs case 1 for 8 h with the effluent segment's stroke error set from −1 % to +1 %. On the pump count the error at 8 h moves 151.1 mL per percent, close to 1 % of the 15.6 L the effluent pump moves in that time, and it is −33.1 mL even with a perfect effluent segment, because the other segments, the wear and the suction still bias the count. On the scales the worst error of the same runs stays between 9.5 and 13.8 mL, whatever the segment does.

🕮 **Pause.** Try Exercise 5.2 before reading on.

5.4.2 The Balance Loop Sees an Integrator

The balance loop trims the effluent pump, and the weighed running total integrates the trim. Because the scales weigh the bags, the measured total answers the pumps within the 2 s drive lag, and the jacket lag of Section 5.3 acts on the patient and on TMP, not on the measurement. The control

Table 5.4 Scale Supervision Against a Constant Drift of the Effluent Scale Starting at 1 h, Case 5

Drift (g/h)	Flagged after onset	Error at 12 h (mL)
20	never	−223.3
30	never	−322.7
35	never	−379.3
40	25.0 min	−44.7
60	16.0 min	−33.2
80	12.0 min	−46.4
60, supervision off	—	−647.8

model is therefore an integrator, $V_{net,m}(s) = \Delta(s)/s$ in the trim Δ, and a PI controller with gain K_c and integral time τ_I gives the closed-loop characteristic polynomial $s^2 + K_c s + K_c/\tau_I$. The defaults, $K_c = 2/\text{h}$ and $\tau_I = 1\,\text{h}$, place the poles at $-1 \pm \text{j}$ per hour, a damping ratio of 0.707 and a natural frequency of 1.41 rad/h. A step ΔQ in an unmeasured rate then leaves the error $\Delta Q\, e^{-t} \sin t$, with t in hours, which peaks at $0.322\,\Delta Q$ after 0.785 h.

To test that model, the 150 mL/h step of case 2 was applied at 0.25 h with the feedforward off, the trim bound raised to 1000 mL/h and the instruments noise-free. The linear prediction is a peak of 48.4 mL at 0.785 h after the step. The engine's error is already 2.2 mL behind when the step comes, because the loop is still absorbing the 11.2 mL/h shortfall it started with, and it peaks 53.4 mL behind 0.79 h after the step. Taken as the stepped run less an otherwise identical unstepped run, the step response peaks at 49.6 mL after 0.80 h. The remaining 1.2 mL comes from the effluent segment's under-filling, which grows as the added effluent flow draws the jacket down from −1.5 mmHg to about −15 mmHg at the peak, so the pump delivers less than each added milliliter of command. With the fluid segments' under-filling set to zero the step response peaks at 48.4 mL after 0.786 h, as predicted, while setting the loading errors to zero changes nothing. The shipped case 2 makes the step at 1 h with noise and a trim bound of 300 mL/h, and its true error peaks 65.8 mL behind at 1.782 h, as the pumps restart after the effluent bag change that began at 1.722 h. At that instant the controller's own error is 53.4 mL, the scales read the running total 2.8 mL low, and the working target trails the trajectory by the 15.2 mL of backlog that the pause has just accrued. Raising the bound to 1000 mL/h leaves that peak unchanged. The time scales are well separated, since the drives settle in seconds, the jacket in 62 to 327 s and the balance in about an hour, so each layer can be tuned as if the faster ones were instantaneous.

5.4.3 A Loop That Holds a Drifting Measurement

The integrator in the plant and the integral in the controller make the balance a type-2 loop, which follows a ramp without steady error. A scale bias that grows at a constant rate is such a ramp in the measurement, and the loop follows it exactly by moving the true removal instead. In case 5 the effluent scale drifts 60 g/h from 1 h, the loop keeps its measured error within 28.4 mL for the whole day, and the patient is 37.4 mL behind at 2 h, 304.3 mL at 6 h, 647.8 mL at 12 h and 1358.9 mL at 24 h. By 6 h the effluent scale has counted 292.3 mL of its 300 g bias into the running total. Of the other 7.7 g, 6.0 g accrued while its bag was off the scale during the changes at 1.72 and 4.20 h, and the rest is reading noise. No tuning of the balance loop removes the ramp-following error, since any stable PI loop on this integrator follows the ramp, and a slower loop only follows it later.

🕮 **Pause.** Try Exercise 5.1 before reading on.

The plausibility check of Section 5.5 compares each scale's flow over 20 min with its pump's calibrated count and flags a disagreement above 20 mL/h plus 1 % of the flow at two checks in a row, about 40 mL/h for the effluent scale. Table 5.4 shows what that tolerance lets through. A drift of 40 g/h or more is flagged within 25 min and the error at 12 h stays within 46.4 mL, while 20, 30 and 35 g/h are never flagged and leave the patient 223.3, 322.7 and 379.3 mL behind. Below the tolerance

the drift is also taught to the pump count. At each clean effluent bag change the calibrated stroke is re-learned from the scale, by at most 1 % per bag, so at 35 g/h it climbs from 0.995 to 1.012 by 4.23 h and the reference follows the fault. Near the tolerance the outcome depends on the noise. At one of four other seeds 30 g/h is flagged 51 min after the onset, at the check after the first effluent change, whose pause shrinks the flow part of the tolerance and whose final weighing and re-tare each add one noisy reading to the window, and the substitute count then carries the part of the drift already learned, so the error still reaches 181.5 mL by 12 h.

> **TRY IT IN THE APP** **5 · Scale Drift**
>
> Load scenario 5 and press Run, then open the Balance view. From 1 h the effluent scale drifts. Watch the first trend: the true removal error leaves the ±60 mL band while the error by the scales, dashed, stays inside it. By 6 h the alarm badge reads a removal error of −304 mL.

5.4.4 One Pump for Every Flow

The effluent pump carries the dialysate, the replacement and the net removal at once, so its speed limit is shared by three loops. Below the limit more dialysate is matched by more effluent, and the steady gain of the dose loop is the diffusive slope $\partial K/\partial Q_d = 1 - \mathrm{e}^{-x}(1+x)$, with $x = K_0 A\, a/Q_d$, which is 0.897 at the reset of case 1. At the limit the allocation of Section 5.5 protects the net removal, so each added milliliter of dialysate displaces a milliliter of replacement and, with $S = 1$, a milliliter of convective clearance, and the gain becomes roughly the diffusive slope less one. At the moment the effluent pump reaches full speed in case 8, at 3.330 h, the steady gain computed from the model is 0.427 while the effluent follows the dialysate and -0.493 once the replacement gives way, and at 7 h, with the dialysate at its limit, it is -0.820. A dose loop whose process gain changes sign becomes a positive feedback that drives the dialysate up while the dose falls. The remedy is structural, a cap on the dialysate that keeps effluent headroom, and no tuning of the dose loop supplies that headroom, since a slower dose loop only postpones the fall. Allocation of one limited resource among several demands is the subject of Volume 4, Chapter 12.

🕮 **Pause.** Try Exercise 5.3 before reading on.

5.5 Controller as Implemented

The engine of Section 1.6.1 runs the controller once per second, every second physics step, before the integration and the discrete events of that step. The dose loop acts only when the clearance monitor reports, every 10 min. The historian logs one row every 20 s, 4320 rows of 147 columns in a day, among them the true running total beside the totals by the scales and by the pump counts, the pump commands, speeds and true deliveries, the pressures, the three scale readings with the effluent scale's bias, and the supervisor's state. The application runs from 1 to 3600 times real time, 300 by default, at which a four-hour case takes 48 s and the day-long challenge 4.8 min.

5.5.1 Instruments

No loop reads a bag volume, a flow or the patient. Four pressure transducers read the access, pre-filter, return and effluent pressures through a 1 s lag with colored noise of 1.5 mmHg correlated over 2 s, and TMP is formed from them as in (5.4) and filtered through 20 s for the override. Four encoders count revolutions exactly. The scales follow (5.9) with 2 g of noise correlated over 3 s and a 250 g tare. An idealized clearance monitor reports the true clearance with 3 % noise every 10 min while the circuit runs, never within 5 min of a restart. Two detectors report an effluent bag at its limit and air from an empty supply bag.

5.5.2 The PI Algorithm

Every loop is an instance of the framework's velocity-form PID of Section 1.6.2, Equations (1.12) to (1.15), with the derivative time at zero and no rate limit. The chapter writes the action as $\mathscr{A}$, the

Table 5.5 The Loops as Configured. Every Loop Is PI, With No Derivative Action and No Rate Limit

Loop	PV → MV (limits)	K_c	τ_I	T_t, $\mathscr{A}$
Blood speed	$\hat{n}_b \to u_b$ (0–100 %)	2.0 %/rpm	3 s	τ_I, +
Fluid speeds	$\hat{n}_k \to u_k$ (0–100 %)	1.5 %/rpm	3 s	τ_I, +
Access	$P_{acc,m} \to$ speed cap	0.10 rpm/mmHg	10 s	10 s, −
TMP	$\text{TMP}_f \to$ UF ceiling	20 mL/h/mmHg	600 s	60 s, +
Balance	$V_{net,m} \to \Delta$ (±300)	2.0/h	1 h	τ_I, +
Dose	$D_m \to Q_d$ (0–$Q_{d,lim}$)	60 kg	1 h	τ_I, +

a of that section: $\mathscr{A} = +1$ for a positive process gain, which is reverse acting, and $\mathscr{A} = -1$ for a negative one, which is direct acting. Only the access override is direct acting, since a faster blood pump draws the access pressure lower. The tracking time T_t of the back-calculation is set per loop in Table 5.5. The control period Δt_c is 1 s for every loop but the dose loop, which uses 600 s at each monitor sample. In MANUAL the four drive stations belong to the operator, each speed loop sets its integral so that its output equals its station, the balance and dose loops track what the stations amount to and the two overrides wait at their ceilings, so the return to AUTO is bumpless. Table 5.5 lists the loops.

5.5.3 Speed Loops and the Access Override

Each pump holds its speed on its encoder. A fluid pump's setpoint is its flow command divided by the nominal stroke, and the blood pump's is the lower of $Q_{b,SP}/V_{s,b}$ and the speed ceiling from the access override, a PI on the measured access pressure against its limit of −150 mmHg. The speed loops remove the droop of (5.2), and they cannot remove the under-filling of (5.1), since the pump turns at the right speed and moves less. An override that is not limiting waits just above the demand it may have to take over. The access ceiling sits at 110 % of the speed the flow setpoint asks for, held there by back-calculation with a 10 s tracking time, so it binds within seconds of the pressure crossing its limit instead of unwinding from full speed. Section 1.4.6 places such overrides on the limits of a living plant, and Volume 4, Chapter 2 treats a protective override whose output rests on its clamp until a limit nears.

5.5.4 Balance, Feedforward and Backlog

The balance loop compares the measured running total $V_{net,m}$, by the scales or by the pump counts at the operator's choice, with the working target $V_w = V_{Rx} - B$, and its output Δ trims the effluent command, as shown in Equation (5.11).

$$Q_{e,cmd} = Q_d + Q_r + Q_{ff} + \Delta, \qquad Q_{ff} = Q_{net,Rx} + \dot{B}_c \tag{5.11}$$

with Q_{ff} the feedforward from the prescription, zero when the feedforward is off, and $\dot{B}_c$ the catch-up rate. While the fluid pumps are paused, the backlog B grows at $Q_{net,Rx}$, so the working target stops and the loop does not wind up against a stopped pump. After the restart, with catch-up on, B is repaid at up to $r_c = 60\,\text{mL/h}$ and the repayment is fed forward. The trim is bounded to ±300 mL/h and to the room that the effluent pump and the ultrafiltration ceiling leave. When the feedforward is turned on or off, the integral absorbs the change in the feedforward term at the same instant, and the effluent command stays where it was.

5.5.5 Filter Protection and the Allocation of the Effluent Pump

Ultrafiltration is capped by the lower of two ceilings, the output U_T of a PI on the filtered TMP against its setpoint of 200 mmHg, and the filtration-fraction ceiling $\text{FF}_{max}\hat{Q}_p$ with $\hat{Q}_p = \hat{Q}_b(1 - \text{Hct}_{ent}) + Q_{r,pre}$ built from the blood pump's count and an entered hematocrit. When not limiting, the TMP override waits 300 mL/h above the ultrafiltration it may have to cap, with a 60 s tracking time, and it

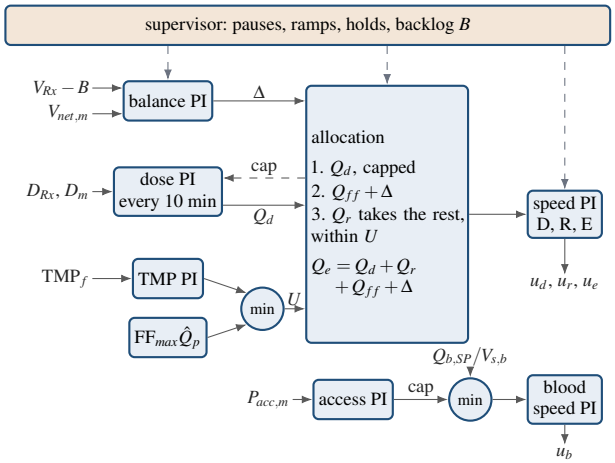


Figure 5.2 Loops above the pump drives. The balance trim Δ, the dialysate from the dose loop and the ultrafiltration ceiling U meet in a fixed allocation of the effluent pump, in which the replacement receives only what the dialysate, the net removal and the ceiling leave, and the headroom option caps the dialysate. The measured running total comes from the scales or the pump counts and, with scale supervision on, a flagged or silent scale is replaced by its calibrated count.

is frozen for 3 min after a restart while the jacket refills. The effluent pump's capacity is then allocated in a fixed order, as shown in Figure 5.2 and Equation (5.12).

$$\begin{aligned} Q_r &= \mathrm{clamp}\big(Q_r^*,\ 0,\ Q_{r,max}\big), \quad U = \min(U_T,\ \mathrm{FF}_{max}\hat{Q}_p) \\ Q_r^* &= \min\{Q_{r,Rx},\ U - Q_{ff} - \Delta^+,\ R_e - \Delta\} \\ R_e &= Q_{e,max} - Q_d - Q_{ff} \end{aligned} \tag{5.12}$$

where $\Delta^+ = \max(0, \Delta)$. The dialysate is set first, the net removal and its trim come next, and the replacement receives what remains up to its prescription, so an override or a full effluent pump takes convection before it takes removal. With the headroom option on, the dialysate is capped at $0.9\,Q_{e,max} - Q_r - Q_{ff}$ using the replacement of the previous scan, which keeps 10 % of the effluent rating for the trim, the wear and the suction.

5.5.6 Dose Tracking and Honest Shortfalls

The dose loop moves the dialysate to hold the measured dose $D_m = K_m/W$ at its prescription, updating once per monitor sample and holding its output between samples and through pauses. At the operating point of case 1 each 1 mL/h of dialysate adds 0.897 mL/h of clearance, 0.0112 mL/(kg h) for the 80 kg patient, so the default gain gives a proportional loop gain of 0.67 per sample. The machine reports a shortfall instead of forcing it. It declares the removal unreachable after 5 min with the ultrafiltration room exhausted or the effluent pump and trim at their limits, and the dose unreachable after 30 min with the dialysate at its limit and the monitor below 97 % of the prescription.

5.5.7 Scale Supervision

With scale supervision on, each scale's running total is compared once a minute with its pump's calibrated count, as shown in Equation (5.13).

$$\delta q = \frac{(S_b - S_a) - (C_b - C_a)}{W_c}, \quad \tau_q = q_{abs} + q_{rel}\,\frac{|C_b - C_a|}{W_c} \tag{5.13}$$

where S and C are one-minute means of the scale's running total and of the calibrated count at the two ends of the window $W_c = 20$ min, with $q_{abs} = 20$ mL/h and $q_{rel} = 1$ %. Two consecutive checks with

$|\delta q| > \tau_q$ flag the scale, and the balance then uses the calibrated count for that bag, rebuilt from the start of the window. A silent scale is replaced by the count at once, and a recalibration re-bases every scale on the total in use. The calibrated count is the nominal stroke times a factor learned from the scale at each clean bag change of at least an hour, limited to 1 % per bag, times the suction correction $1 + k_{in,f} \min(0, P_{eff,m})$ for the effluent pump, so it follows the wear that the raw count misses. With supervision off, a silent scale holds its last reading.

5.5.8 Supervisor

A supervisor sequences the pumps through five states, RUN, STOPPING, CHANGING, ALARM HOLD and RESTART. A supply bag at 150 mL or an effluent bag at 4800 mL, as the machine estimates them, starts a bag change, and with the change-together option any other bag due within 45 min joins the same pause. The fluid pumps stop, the final weights are recorded, the bags are swapped after 180 s and re-tared, and the pumps ramp back over 30 s, about 3.7 min in all, while the blood keeps circulating. A measured access pressure below −250 mmHg for 2 s, or a pump-count blood flow under 30 % of its setpoint for 20 s, stops every pump until the operator restarts. A measured TMP above 350 mmHg for 10 s stops the fluid pumps, a pre-filter pressure above 450 mmHg for 5 s stops the circuit as clotted, and a filter change stops it for 20 min. While the state is not RUN the upper loops hold their outputs, the TMP override freezes and the missed removal accrues as backlog.

5.5.9 Simulation, Control and Internal Models

Three models are kept apart here. The simulated plant is the model of Section 5.3. The control model of the balance is the integrator of Section 5.4, whose gain is slightly below one because part of each added milliliter of effluent is lost to suction. The control model of the TMP override is the jacket lag with a steady gain from ultrafiltration to TMP that rises from 0.065 to 0.413 mmHg per 1 mL/h as the open area falls from 1 to 0.15, while its time constant rises from 62.4 to 326.6 s, so the override meets a plant that stiffens and slows as the filter fouls. The controller's internal models are nominal by design. The feedforward assumes every pump delivers its nominal stroke, the filtration-fraction ceiling trusts the pump-count blood flow and the entered hematocrit, the headroom trusts the nominal effluent rating, and the calibrated count trusts the learned stroke and the nominal suction coefficient. None of them knows the patient.

5.6 Cases and Measured Outcomes

Table 5.6 gives what the eight cases produced, and the subsections below give their scripted events. Each case begins from a reference configuration, with the balance on the scales, scale supervision, feedforward, catch-up at 60 mL/h, filter protection, dose tracking with effluent headroom and the access override on, and a removal of 100 mL/h, 930 mL/h of dialysate, 1000 mL/h of replacement and a dose of 25 mL/(kg h), apart from the changes named. Every value below comes from a headless run of the simulation engine at its default seed, with any setting changed before the run, and each is a true value rather than an instrument reading unless stated. Four other seeds move the noise-dependent results only modestly, the worst error of case 1 over 4 h between 7.9 and 10.9 mL and the flag of the case 5 drift between 14 and 16 min. Errors are quoted as the largest deviation of the true running total from its trajectory, which, unlike an integrated error, stops growing when a transient ends.

Each case is a comparison of two runs: a run as shipped, a Reset, which restarts from $t = 0$ and keeps the learner's settings, one changed setting and a second run. The challenge is judged on the case as given, and changing a model parameter or a disturbance voids the attempt. The briefings, completion rules and the challenge follow Section 1.6.4.

5.6.1 The Balanced Circuit and a New Target

On the scales, case 1 holds the running total within 7.9 mL of its trajectory over 4 h, through two bag changes that pause the fluid pumps for 7.4 min, and within 22.3 mL over a day of 18 changes. The day's worst comes at 14.41 h, as the third of three bag changes within 20 min ends with 11.4 mL of

Table 5.6 Measured Outcomes of the Eight Cases

Case	Measured outcome
1, scales	worst 7.9 mL in 4 h, 22.3 mL in 24 h
1, pump count	48.6 mL behind at 4 h, 466.3 mL at 24 h
2, feedback only	65.8 mL behind at 1.78 h, 62.2 mL ahead at 5.01 h
2, feedforward	worst 18.2 mL, 46 min after the change
3, protection on	TMP at most 207.5 mmHg, Q_r 797 at 6, 503 at 8 h
3, dose	25.0 to 19.6 at 6 and 13.7 mL/(kg h) at 7.99 h
3, protection off	TMP 250 mmHg at 6.05 h, 350.9 mmHg at 7.70 h, stop
4, override on	−150 mmHg, count 111.6 mL/min, true 103.5 mL/min
4, filtration fraction	true 25.4 %, estimated 24.0 %, peak 26.3 %
4, override off	−198.7 mmHg, count 150 mL/min, true 135.5 mL/min
4, kinked line	stop at 2.00 h, 14.9 min, area 93.3 to 88.2 %
5, supervision off	measured within 28.4 mL, true −304.3 mL at 6 h
5, supervision on	flagged after 16.0 min, −12.4 mL at 6 h
6, catch-up off	10 pauses, 37.2 min, −88.1 mL at 8 h
6, catch-up on	worst 18.5 mL, together 5 pauses in 18.6 min
7, hold last reading	measured −651 mL, true +79.8 mL
7, substitute count	worst 6.8 mL
8, no headroom	full speed at 3.33 h, 21.4 mL/(kg h) over hours 5 to 8
8, headroom	27.8 mL/(kg h) over hours 5 to 8, shortfall at 3.17 h

backlog to repay, while the noisy final weighings and re-tares of the earlier changes have left the scale total about 12 mL above the true one. At four other seeds it is 12.2 to 19.9 mL, and noise-free 10.8 mL. The trim averages 12.3 mL/h over the first four hours, close to the 11.2 mL/h shortfall at reset, and it grows with the wear to 34.5 mL/h over the last four hours of the day (12.9 and 29.8 mL/h noise-free). On the pump count the patient is 48.6 mL behind at 4 h and 466.3 mL behind at 24 h. In the first four hours the open area falls to 92.3 % at the baseline fouling rate, TMP rises to 83.6 mmHg and urea falls from 30.0 to 26.6 mmol/L.

Case 2 raises the removal to 250 mL/h at 1 h and lowers it to 50 mL/h at 4 h. On feedback alone the patient falls 65.8 mL behind at 1.78 h, the trim reaches its 300 mL/h bound at 1.86 h and stays at or near it until about 4 h, and when the target drops the integral is still carrying the old rate, so the patient runs 62.2 mL ahead at 5.01 h and 31.4 mL ahead at 6 h. With the bound raised to 1000 mL/h that overshoot is 63.1 mL, close to the linear peak of 64.5 mL that the step down of 200 mL/h alone gives, so the bound does not cause it. Fed forward, the prescription leaves the loop only the pump errors, and the worst error is 18.2 mL, 46 min after the first change. In that run the filtration fraction climbs to 22.3 % at 250 mL/h, the plasma falls to 2.89 L from 3.20 L at reset and the hematocrit rises from 30 to 32 %, because the interstitium refills the plasma more slowly than the circuit takes it.

5.6.2 Fouling and the TMP Override

Case 3 fouls the filter twelve times faster than the baseline. The open area falls to 61.9, 38.3 and 23.7 % at 2, 4 and 6 h, and TMP rises from 77.1 to 111.4 and 163.3 mmHg at 2 and 4 h, since TMP less the oncotic pressure grows about inversely with the open area. TMP reaches the override setpoint at 4.99 h, and from 5.16 h the override cuts the replacement command, to 797 mL/h by 6 h and 503 mL/h by 8 h, with TMP at most 207.5 mmHg. Dose tracking meanwhile raises the dialysate from 930 to 1630 and 2729 mL/h, and the dose falls anyway, from 25.0 to 19.6 at 6 h and 13.7 mL/(kg h) just before the bag change at 7.99 h, because diffusion through a clogged membrane saturates at $K_0 A\, a$, 8.8 mL/min at 8 h. At 8.2 h tracking is worth 1.0 mL/(kg h), 13.2 against 12.2 without it. An effluent-based dose would read 41.7 mL/(kg h) at 7.99 h, since the effluent carries 3339 mL/h, while the clearance delivers 13.7. The pre-filter pressure passes the 300 mmHg filter-spent warning at 7.81 h, the filter is changed at 8.25 h, the 20 min stop takes the removal from 13.6 mL ahead to 20.2 mL behind, and TMP is back to 84.0 mmHg at 8.7 h. With the protection off TMP passes 250 mmHg at 6.05 h and reaches 350.9 mmHg at 7.70 h, where the fluid pumps stop until the filter change, and the

removal is 103.5 mL behind when they restart at 8.59 h. With the protection on, the removal error stays within 13.9 mL until the filter change, because the replacement gives way first.

5.6.3 Access Restriction

Case 4 raises the access resistance 2.6-fold at 0.5 h. The override holds the measured access pressure at −150 mmHg by slowing the blood pump to a mean of 111.6 mL/min by count between 1.0 and 1.7 h, while 103.5 mL/min actually passes, since each revolution under that suction moves 7.3 % less. The filtration-fraction ceiling uses the count, so the machine's estimate averages 24.0 % against a true 25.4 %, and the true fraction peaks at 26.3 %. The ceiling is only as good as the blood-flow estimate. Without the override the pump keeps its 150 mL/min count, delivers 135.5 mL/min and holds the access at −198.7 mmHg. At 2 h the line kinks to 20 times its resistance, the access pressure falls to the −450 mmHg collapse pressure and the pumps stop at 2.00 h. The blood stands still for 14.9 min, the open area falls from 93.3 to 88.2 % and the removal falls 28.7 mL behind. The circuit is back in RUN at 2.26 h, catch-up repays the 25.8 mL backlog over the next 25.8 min, and without catch-up the patient would still be 37.4 mL behind at 4 h.

5.6.4 Drift and Dropout

With scale supervision on, the drift of case 5 is flagged at 1.27 h, 16.0 min after it began and with 16 g of bias, the balance is rebuilt on the pump count from the start of the window, and the patient is 12.4 mL behind at 6 h. The count cannot follow the wear between recalibrations, and the error reaches 36.8 mL by 12 h. The case is shown in Figure 5.3 as shipped, with the supervision off.

Case 7 silences the effluent scale for 20 min at 1 h. With supervision off the last reading is held, the measured removal falls by the inflow, and the loop sees the patient 651 mL behind by 1.33 h and drives the trim to its 300 mL/h bound from 1.07 h. The patient is in fact being over-dried, 79.8 mL beyond the prescription when the reading returns and 10.6 mL at 2.5 h. With the bound at 1000 mL/h the same dropout over-removes 209.0 mL, so the harm scales with the authority given to the outer loop. With supervision on the count stands in for the missing scale and the error never exceeds 6.8 mL.

> **TRY IT IN THE APP** **7 · Sensor Dropout**
>
> Load scenario 7 and press Run. At 1 h the effluent scale stops reporting for 20 min. With Scale supervision off the loop drives the effluent trim to its 300 mL/h bound, and the patient is 79.8 mL beyond the prescription when the reading returns. Press Reset, set Scale supervision to On and run again: the error never exceeds 6.8 mL.

5.6.5 Bag Changes

Case 6 prescribes 1800 mL/h of dialysate, 1200 mL/h of replacement and a 150 mL/h removal, so the effluent pump draws about 3150 mL/h and the effluent bag fills about every 1.6 h. Without catch-up, ten changes pause the fluid pumps for 37.2 min in 8 h and leave the patient 88.1 mL behind. With catch-up the backlog is repaid and the worst error is 18.5 mL. Changing every bag due within 45 min in the same pause halves the pauses to five and the paused time to 18.6 min, and raises the delivered dose from 33.13 to 34.33 mL/(kg h). The ledger closes within 3.1×10^{-10} mL through every change.

5.6.6 The Effluent Pump at Its Limit

Case 8 treats a 120 kg patient at 250 mL/min of blood, 1500 mL/h of dialysate, 2000 mL/h of replacement and a 150 mL/h removal, 3650 mL/h of effluent by count and a dose of 29.0 mL/(kg h) against a prescribed 30, and raises the prescribed dose to 40 mL/(kg h) at 1 h with the headroom off. The dose loop raises the dialysate until the effluent drive reaches full speed at 3.33 h, and from 3.42 h the effluent command sits at its 4500 mL/h limit. The replacement then gives way milliliter for milliliter, its command falling to 368 mL/h by 8 h, the dialysate reaches its 4000 mL/h limit at 6.44 h, and the

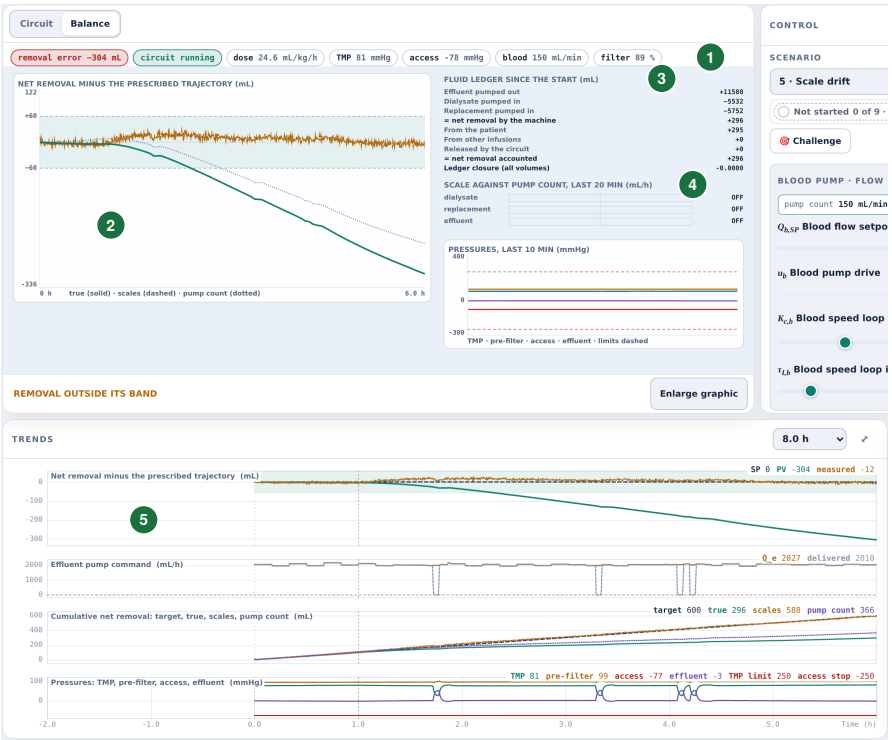


Figure 5.3 The application in case 5, scale drift, at $t = 6.00$ h with scale supervision off as shipped, five hours after the effluent scale began to drift at 60 g/h, on the Balance view, cropped to the plant view and the trends. Callouts: (1) the status badges, (2) the plot of the removal error in the scene, (3) the fluid ledger, (4) the scale checks and (5) the first trend. The plot in the scene and the first trend show the true removal error leaving the ±60 mL band while the error by the scales, dashed on the trend, stays inside it, and the alarm badge reads a removal error of −304 mL. The fluid ledger beside the plot balances 11 580 mL of effluent against 5532 mL of dialysate and 5752 mL of replacement, a net removal of 296 mL with 295 mL of it from the patient, and closes to 0.0000 mL, while the three scale checks read OFF. The trends over an 8 h window show the true error at −304 against a measured error of −12, the effluent command at 2027 mL/h (2010 mL/h delivered) with four dips for bag changes, and the running totals parting, target 600 mL, true 296 mL, scales 588 mL and pump count 366 mL. TMP is 81 mmHg on a filter 89 % open.

dose delivered over hours 5 to 8 is 21.4 mL/(kg h), below the 29.0 of the start. The worst balance error, 32.5 mL ahead at 7.11 h, builds while the dialysate drive runs at 100 % against a pressure rise of 53.8 mmHg and turns at the speed for 3936 mL/h of its 4000 mL/h command, since the allocation gives the replacement the room the command leaves. With the headroom on, the dialysate is capped near 1900 mL/h, the dose over hours 5 to 8 is 27.8 mL/(kg h), the machine reports at 3.17 h that the prescribed 40 is out of reach, and the worst balance error is 20.3 mL.

5.6.7 A Day on the Circuit

The challenge runs 24 h with the removal rising from 100 to 150 mL/h at 8 h, fouling at 3.2 times the baseline, a partly blocked catheter from 5 to 7 h, a kinked line at 12 h, an effluent-scale drift of 60 g/h from 15 h, a routine recalibration at 17 h and 40 mL/h of other infusions. It is met with the true running total within ±60 mL of its trajectory at every instant, at least 85 % of the prescribed dose delivered, TMP at most 250 mmHg, a filtration fraction at most 25 % and the ledger closed within 1 mL. Table 5.7 compares the configurations, and Figure 5.4 draws three of its columns as bars. In its

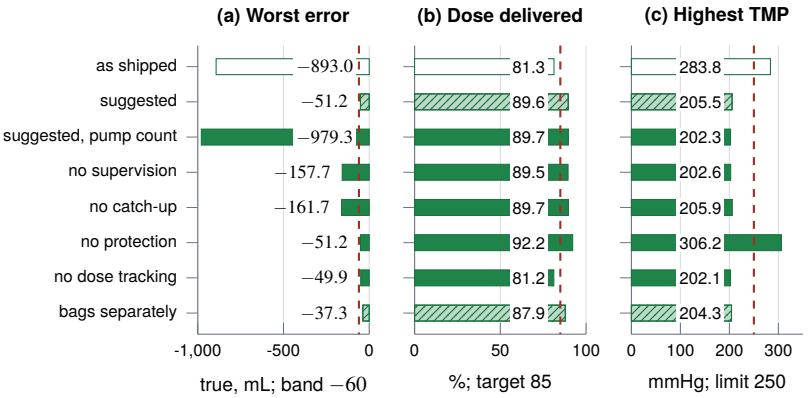


Figure 5.4 The eight configurations of Table 5.7 over the day of the challenge: (a) the worst deviation of the true running total against the ±60 mL band, (b) the dose delivered against the 85 % target and (c) the highest TMP against 250 mmHg, limits dashed. Rows after the second change one element of the suggested configuration. White: the shipped configuration. Solid: the other rows that fail. Hatched: the two rows that pass. Three of the failing rows fail on the band alone, the run without protection on TMP and the run without dose tracking on the dose.

shipped form the circuit balances on its pump counts with 700 mL/h of dialysate and no catch-up, filter protection or dose tracking, and it ends 893.0 mL short of its 3200 mL target, with TMP past 250 mmHg from 21.92 h and 81.3 % of the dose. The suggested configuration adds the scales with supervision, catch-up, protection, dose tracking and bags changed together. Its worst error is 51.2 mL behind at 12.26 h, as the circuit restarts after the access stop, 8.8 mL inside the band. It ends 24.1 mL short of the target, flags the drift at 15.27 h, holds TMP to 205.5 mmHg by cutting the replacement command to 617 mL/h by the end of the day while the dialysate rises to 3070 mL/h, pauses 14 times, 13 of them for bag changes and one for the access stop, for 63.7 min, keeps the filtration fraction at most 22.5 % and delivers 89.6 % of the dose. The application gives this one run whether the changes are made before Run or carried through a Reset. At four other seeds the suggested configuration still meets the target, with worst errors from 31.1 to 59.3 mL, the largest 0.7 mL inside the band, so its margin depends on the noise. Removing any single element fails except changing the bags together. Without the filter protection the dose is higher, 92.2 %, because the replacement is never cut, while TMP reaches 306.2 mmHg.

> ✦ DID YOU KNOW? **Changing Bags One at a Time**
>
> From the simulator: in the day-long challenge the worst error of the suggested configuration is 51.2 mL behind. The same configuration with the bags changed separately, not together, has a smaller worst error, 37.3 mL behind, and still passes, at 87.9 % of the dose against 89.6 % (Table 5.7).

5.7 Technology in Practice

The roller pumps, scales and pressure transducers of Figure 5.1 stand for the hardware of a machine for continuous kidney replacement. Figure 5.5 draws the loops of Section 5.5 with that hardware in place of the schematic's symbols.

What Measures It

The balance is measured by weight. Each bag hangs on a scale, and the weight gained by the effluent bag less the weight lost by the supply bags gives the net removal. Four transducers read the access,

Table 5.7 Challenge Configurations Over 24 h. Errors Are the Worst and the Final Deviation of the True Running Total

Configuration	Worst (mL)	Final (mL)	Dose (%)	TMP (mmHg)	Pass
as shipped	−893.0	−893.0	81.3	283.8	no
suggested[a]	−51.2	−24.1	89.6	205.5	yes
suggested, pump count	−979.3	−979.3	89.7	202.3	no
suggested, no supervision	−157.7	−134.8	89.5	202.6	no
suggested, no catch-up	−161.7	−157.5	89.7	205.9	no
suggested, no protection	−51.2	−27.3	92.2	306.2	no
suggested, no dose tracking	−49.9	−31.9	81.2	202.1	no
suggested, bags separately	−37.3	+27.4	87.9	204.3	yes

[a] scales, supervision, catch-up, protection, dose tracking, bags together

pre-filter, return and effluent pressures, and the transmembrane pressure is formed from three of them by the same combination that clinical studies of filter clotting use when they read the pressures a machine has recorded [18]. The chapter's scales follow Equation (5.9), with 2 g of correlated noise, a 250 g tare, a 1 g resolution and a bias that can drift, and its transducers read through a 1 s lag.

> **TECHNOLOGY Sensor: Bag Scales and Gravimetric Balancing**
>
> A machine for continuous therapy keeps its fluid balance by weight, by volume in balancing chambers, by flow meters, or by a combination. Weighing with one or more scales is the most common method, because it is the most reliable over long treatments. The effluent and the supply fluids are weighed continuously, the weight stands in for the flow, and the software adjusts the pump rates by servo feedback when prescribed and actual values disagree. Its weak points are the capacity of the scales, user errors and disturbances from the surroundings. Balancing by volume collects systematic, cumulative errors over a long treatment, since it lacks that continuous feedback [19]. Conventional machines control the balance to about ±30 mL/h, and one machine designed for newborns reports an ultrafiltration error of 1 g/h [20].

What Moves It

The final elements are four roller pumps, each turned by a drive under a speed loop. In the chapter a pump's delivery is its stroke times its speed, less the loading error of its tubing segment, its wear and its under-filling at a suction, as in Equation (5.1). The encoder on each pump counts revolutions exactly, and that count is the second, independent measurement of every flow that the scale supervision of Section 5.5 compares with the scales.

> **TECHNOLOGY Final Element: Roller Pumps on Tubing Segments**
>
> Every flow of the circuit is set by a pump. The blood pump sets the blood flow, the dialysate and replacement pumps feed their fluids, and the effluent pump sets the total fluid removed from the filter. Each works on a segment of tubing held between its rotor and stator, and pressure sensors watch the line before the blood pump, the line before the filter and the effluent compartment [19]. A blood pump of the usual kind has two rollers set 180° apart that squeeze a flexible tube. With a small catheter the circuit pressures must rise in proportion to keep the flow, and past a safe threshold the machine alarms stop the treatment [20]. United States regulation classes the delivery machine that pumps the blood and dialysate and controls the removal as a Class II device [21].

Where the Controller Runs

The controller is part of the delivery machine, which in United States regulation carries an ultrafiltration controller and mechanisms that monitor or control the fluid balance [21]. Its software reads

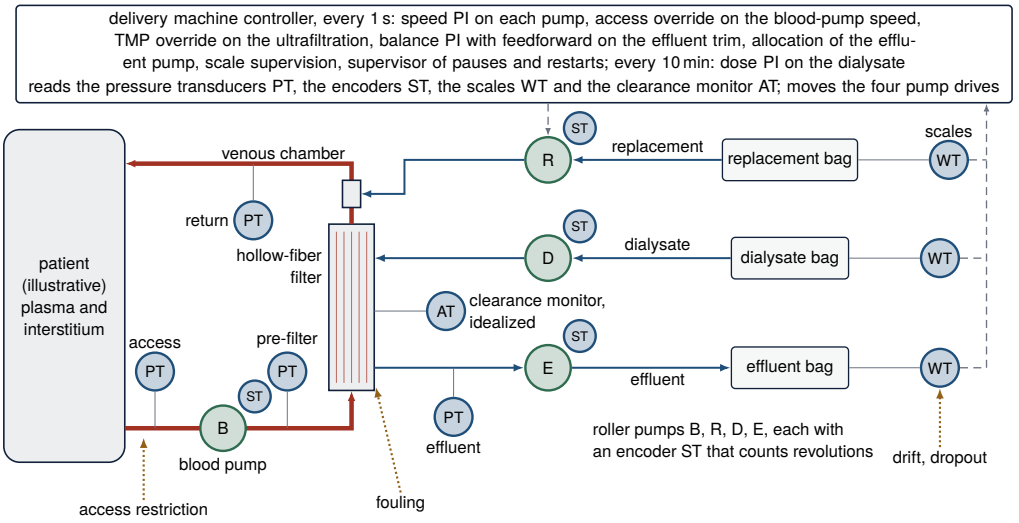


Figure 5.5 The loops of Section 5.5 with the hardware of a machine for continuous kidney replacement. Four roller pumps are the final elements: the blood pump B, and the replacement pump R, dialysate pump D and effluent pump E on the fluid side. The machine reads pressure transducers PT on the access, pre-filter, return and effluent lines, an encoder ST on every pump, a scale WT for every bag and, in the chapter only, an idealized clearance monitor AT. Thick red arrows carry blood, thin blue arrows the fluids, dashed lines the signals, dotted arrows the disturbances. The scales and the encoders measure the same flows twice, which is the redundancy the scale supervision uses.

the scales continuously and adjusts the pump rates by servo feedback [19], and the machine records its pressures and alarms for later review [2, 18]. No opened source gives the scan period of such a machine, so none is given here. The chapter's engine runs the speed loops, the two overrides and the balance loop once a second, the dose loop at each 10 min monitor sample, and the supervisor's bag changes, holds and restarts as a layer above them.

> **IN PRACTICE** **Watching Downtime and Alarms**
>
> Downtime, the share of treatment time with the therapy paused, is a quality indicator of continuous kidney replacement. In one cohort, access, return and filter pressure alarms were each associated with more downtime, and downtime widened the gap between prescribed and achieved fluid balance only in patients with six or more alarms a day [2]. A 2025 national guideline notes that the delivered dose often falls below the prescribed one, through interruptions, predilution and a less permeable membrane, so that 25 to 30 mL/kg/h is typically prescribed for a target of 20 to 25. It suggests a specialized or multidisciplinary team to lengthen filter life and reduce downtime [22].

What Is Modernizing It

Three developments bear on these loops. The first is data. Records from the machine and the health record can be screened by programs that alert clinicians to dangerous trends, and their suggestions can be fed back to the pumps, with or without a clinician's approval [9]. The second is a measurement of the patient. Bioimpedance and online hematocrit have been proposed for integration into the machine [9], and a multicenter randomized trial adjusted the ultrafiltration by bioelectrical impedance analysis [22]. Either would close a loop on the patient, which no loop of this chapter reads. The third is prediction. Early filter clotting is hard to detect, since the alarms sound only after the resistance has risen substantially, and resistances computed from the recorded pressures and flows have been proposed to predict it [18]. Machine learning has been tried for clotting, hypotension and alarm

management, though most such studies are retrospective and many lack external or prospective validation [23].

5.8 Limitations

The most basic limitation concerns purpose. The application is an educational equipment-control simulation of how a machine tracks a prescribed removal and dose with pumps, scales and pressure limits. Its patient model, its filter coefficients and every threshold it displays, the TMP limit and stop, the filtration-fraction ceiling, the access and pre-filter stop limits and the dose, are illustrative teaching values that have not had clinical review. It models no particular device, the prescriptions are scenario inputs that it never chooses, and nothing in it or in this chapter should be used to plan, adjust or judge the treatment of a patient.

> **OPEN CHALLENGE** **How Fast to Take the Water**
>
> The chapter takes the removal rate as a prescription it never chooses, and medicine has not settled that rate either. A 2025 guideline found no prospective intervention study comparing net ultrafiltration rates, rested its suggestion on retrospective studies and called for randomized trials [22]. A cohort of 973 patients in two centers then found the rate not independently associated with 28-day mortality once the achieved fluid balance was accounted for, while the balance itself was. Progress would be prospective studies that set patient-specific balance targets and use the rate as the means to reach them [24].

The technical limitations follow the assumptions. The patient has no blood pressure and no tolerance of removal, so the application cannot show a removal that the patient does not bear. Only urea is followed, and electrolytes and acid-base are absent. The membrane is lumped, its fouling is prescribed and the clearance monitor is idealized. The alarms are those the lessons need, and there is no air detector on the blood line, no blood-leak detector and no return-pressure alarm. The removal-unreachable report is implemented but no shipped scenario triggers it. Repeatability is itself unrealistic, since a fixed step and a fixed seed give the same day on every computer.

> **OPEN CHALLENGE** **Measuring the Dose Delivered**
>
> Dose estimates from the flows, the effluent or the pump settings, are easy at the bedside but often overestimate the dose found from urea clearance. Access dysfunction, recirculation, problems with anticoagulation and progressive fouling of the membrane lower the delivered dose without showing in the flows. Urea has limits of its own as a marker, and most solutes cannot be measured in routine practice [25]. The clearance monitor of Section 5.5 is idealized for that reason. Progress would be marker solutes measured in real time, at low cost, at the bedside [25].

5.9 Conclusion

A continuous kidney replacement circuit controls a small difference of large flows, and the simulator of this chapter carries it as a running total built from four imperfect pumps and three scales while it records the truth separately. On the pump counts the tubing tolerances, wear and suction leave the patient 466.3 mL behind after a day, and on the scales the balance stays within 22.3 mL. The balance loop, a PI loop on an integrator whose step response comes within 3 % of its linear prediction, passes a 60 g/h scale drift to the patient, 304.3 mL in 6 h, while its own error stays within 28.4 mL. A check against the pump counts flags that drift in 16.0 min but never flags a steady 35 g/h drift below its tolerance, which leaves the patient 379.3 mL behind in 12 h. At the effluent pump's limit the gain of the dose loop changes sign, from 0.427 to -0.493, and a cap that keeps effluent headroom holds the dose at 27.8 against 21.4 mL/(kg h) over hours 5 to 8. The challenge moves from 893.0 mL short and 81.3 % of the dose to within 51.2 mL and 89.6 %. The patient model is illustrative and the clearance monitor idealized. A fuller model would add the hemodynamic tolerance of removal, electrolytes,

clearance inferred from samples and a population of patients. The simulator is a browser plant for learning that a loop holding its own measurement need not be holding the patient.

KEY POINTS

- The water a patient loses is a small difference of large pumped flows, so a small pump error grows large in it.
- On the pump counts the patient ends a day 466.3 mL behind, and on the scales the balance stays within 22.3 mL.
- A balance loop on a drifting scale keeps its own error within 28.4 mL while the patient falls 304.3 mL behind in 6 h. A check against the pump counts flags that drift in 16.0 min.
- At the effluent pump's limit the gain of the dose loop changes sign, and a cap that keeps effluent headroom holds the dose at 27.8 against 21.4 mL/kg/h.

Where else this applies. A running total that only an independent measurement can check returns in Volume 5, Chapter 13, and a weighed flow held while the product is not in Volume 12, Chapter 12.

References

[1] J. Bouchard, S. B. Soroko, G. M. Chertow, J. Himmelfarb, T. A. Ikizler, E. P. Paganini, and R. L. Mehta, "Fluid accumulation, survival and recovery of kidney function in critically ill patients with acute kidney injury," *Kidney International*, vol. 76, no. 4, pp. 422–427, 2009.

[2] C. Braun, T. Takeuchi, J. Lambert, L. Liu, S. Roberts, S. Carter, W. Beaubien-Souligny, A. Tolwani, and J. A. Neyra, "Association of continuous renal replacement therapy downtime with fluid balance gap and clinical outcomes: a retrospective cohort analysis utilizing EHR and machine data," *Journal of Intensive Care*, vol. 12, p. 55, 2024.

[3] J. L. Teruel, M. Fernández Lucas, R. Marcén, J. R. Rodríguez, J. López Sánchez, M. Rivera, F. Liaño, and J. Ortuño, "Differences between blood flow as indicated by the hemodialysis blood roller pump and blood flow measured by an ultrasonic sensor," *Nephron*, vol. 85, no. 2, pp. 142–147, 2000.

[4] C. M. Crowe, "Data reconciliation — progress and challenges," *Journal of Process Control*, vol. 6, no. 2–3, pp. 89–98, 1996.

[5] R. Isermann, "Model-based fault-detection and diagnosis – status and applications," *Annual Reviews in Control*, vol. 29, no. 1, pp. 71–85, 2005.

[6] K. J. Åström and T. Hägglund, *Advanced PID Control*. Research Triangle Park, NC: ISA – The Instrumentation, Systems, and Automation Society, 2006.

[7] A. Santoro, E. Mancini, C. Basile, L. Amoroso, S. Di Giulio, M. Usberti, G. Colasanti, G. Verzetti, A. Rocco, E. Imbasciati, G. Panzetta, R. Bolzani, F. Grandi, and M. Polacchini, "Blood volume controlled hemodialysis in hypotension-prone patients: A randomized, multicenter controlled trial," *Kidney International*, vol. 62, no. 3, pp. 1034–1045, 2002.

[8] P. Kramer, W. Wigger, J. Rieger, D. Matthaei, and F. Scheler, "Arteriovenous haemofiltration: A new and simple method for treatment of over-hydrated patients resistant to diuretics," *Klinische Wochenschrift*, vol. 55, no. 22, pp. 1121–1122, 1977, in German.

[9] C. Ronco, "Continuous renal replacement therapy: Forty-year anniversary," *The International Journal of Artificial Organs*, vol. 40, no. 6, pp. 257–264, 2017.

[10] P. Kramer, W. Wigger, D. Matthaei, C. Langescheid, J. Rieger, C. Fuchs, K.-W. Rumpf, and F. Scheler, "Clinical experience with continuously monitored fluid balance in automatic hemofiltration," *Artificial Organs*, vol. 2, no. 2, pp. 147–149, 1978.

[11] M. H. Sigler and B. P. Teehan, "Solute transport in continuous hemodialysis: A new treatment for acute renal failure," *Kidney International*, vol. 32, no. 4, pp. 562–571, 1987.

[12] C. Ronco, R. Bellomo, P. Homel, A. Brendolan, M. Dan, P. Piccinni, and G. La Greca, "Effects of different doses in continuous veno-venous haemofiltration on outcomes of acute renal failure: a prospective randomised trial," *The Lancet*, vol. 356, no. 9223, pp. 26–30, 2000.

[13] RENAL Replacement Therapy Study Investigators, R. Bellomo, A. Cass, L. Cole, S. Finfer, M. Gallagher, S. Lo, C. McArthur, S. McGuinness, J. Myburgh, R. Norton, C. Scheinkestel, and S. Su, "Intensity of continuous renal-replacement therapy in critically ill patients," *New England Journal of Medicine*, vol. 361, no. 17, pp. 1627–1638, 2009.

[14] D. Schneditz, J. Roob, M. Oswald, H. Pogglitsch, M. Moser, T. Kenner, and U. Binswanger, "Nature and rate of vascular refilling during hemodialysis and ultrafiltration," *Kidney International*, vol. 42, no. 6, pp. 1425–1433, 1992.

[15] M. Joannidis and H. M. Oudemans-van Straaten, "Clinical review: Patency of the circuit in continuous renal replacement therapy," *Critical Care*, vol. 11, no. 4, p. 218, 2007.

[16] R. Murugan, E. Hoste, R. L. Mehta, S. Samoni, X. Ding, M. H. Rosner, J. A. Kellum, C. Ronco, and on behalf of the Acute Disease Quality Initiative (ADQI) Consensus Group, "Precision fluid management in continuous renal replacement therapy," *Blood Purification*, vol. 42, no. 3, pp. 266–278, 2016.

[17] C. Ronco, P. M. Ghezzi, A. Brendolan, C. Crepaldi, and G. La Greca, "The haemodialysis system: basic mechanisms of water and solute transport in extracorporeal renal replacement therapies," *Nephrology Dialysis Transplantation*, vol. 13, no. Suppl. 6, pp. 3–9, 1998.

[18] Y. Hu, Z. Hu, C. Li, X. Yang, Y. Ge, and D. Gong, "Prediction of filter clotting using longitudinal trends of circuit pressure parameters during continuous renal replacement therapy (CRRT): an exploratory study," *Scientific Reports*, vol. 16, p. 17411, 2026.

[19] G. Villa, M. Neri, R. Bellomo, J. Cerda, A. R. De Gaudio, S. De Rosa, F. Garzotto, P. M. Honore, J. Kellum, A. Lorenzin, D. Payen, Z. Ricci, S. Samoni, J.-L. Vincent, J. Wendon, M. Zaccaria, and C. Ronco, "Nomenclature for renal replacement therapy and blood purification techniques in critically ill patients: practical applications," *Critical Care*, vol. 20, p. 283, 2016.

[20] M. Parolin, G. Ceschia, and E. Vidal, "New perspectives in pediatric dialysis technologies: the case for neonates and infants with acute kidney injury," *Pediatric Nephrology*, vol. 39, no. 1, pp. 115–123, 2024.

[21] U.S. Food and Drug Administration, "High permeability hemodialysis system," Code of Federal Regulations, Title 21, Section 876.5860, revised as of April 1, 2024, U.S. Government Publishing Office, govinfo.gov, 2024.

[22] J. Lee, D. Kang, J. H. Paek, J. Lee, J. N. An, J. Jeon, K. Lee, H. R. Jang, J. H. Jhee, H. J. Kim, H. Rhee, S. Y. Lim, J. Yang, S. G. Kim, S. S. Han, S. Y. Ahn, S. Park, H. K. Lee, H. Cho, Y. Lee, H. W. Kim, K. Kim, M. Choi, and S. Kim, "Korean society of nephrology 2025 evidence-based clinical practice guideline for continuous kidney replacement therapy," *Kidney Research and Clinical Practice*, vol. 45, no. Suppl. 1, pp. S1–S54, 2026.

[23] X. Li and L. Zhang, "Artificial intelligence and continuous renal replacement therapy," *Renal Failure*, vol. 48, no. 1, p. 2750759, 2026.

[24] N. C. Monai, C. A. Pfortmueller, J. C. Schefold, L. Cioccari, and M. Faltys, "Net ultrafiltration rate and mortality in critically ill patients: A multicenter cohort study with analysis of fluid balance changes," *Critical Care Medicine*, vol. 54, no. 10, pp. 2648–2659, 2026.

[25] G. Villa, S. Fabbri, S. Samoni, M. Cecchi, A. Fioccola, C. Scirè-Calabrisotto, G. Mari, D. Pomarè Montin, and S. Romagnoli, "Methods for dose quantification in continuous renal replacement therapy: Toward a more precise approach," *Artificial Organs*, vol. 45, no. 11, pp. 1300–1307, 2021.

5.10 Exercises

5.1 ⌕ Concept check · 2 min

In scenario 5 the effluent scale starts to read 60 g/h high at 1 h, and scale supervision is off. The balance loop holds the running total measured by the scales on its target. What do the trends show over the next five hours?

(a) The measured error stays within a few tens of milliliters while the true removal falls behind by about the drift rate

(b) The measured error grows at about the drift rate, and the badge shows the band crossed after about an hour

(c) Both errors stay small, because the integral action of the loop cancels a constant bias of the scale

(d) The measured and true errors each grow at about half the drift rate, because the loop splits the bias

5.2 ⌕ Concept check · 2 min

In case 1 the pumps count 930 mL/h of dialysate and 1000 mL/h of replacement into the circuit and 2030 mL/h of effluent out of it, for a net removal of 100 mL/h. Each pump in turn is given a stroke error of 1 %. Which statement is right?

(a) Each pump moves the net removal by the same amount, one percent of the removal
(b) The effluent pump's error is about a fifth of the removal, twice the effect of either supply pump
(c) The errors of the two supply pumps cancel the effluent pump's error when all three are present
(d) The errors move the balance only during bag changes, when the scales are off the bags

5.3 Concept check · 2 min

In scenario 8 the dose loop raises the dialysate until the effluent pump runs at full speed. Every further milliliter of dialysate then displaces a milliliter of replacement, and the dose falls while the loop pushes harder. Which change keeps the dose from collapsing, however long the run?

(a) Raise the dose-loop gain so that the loop reaches the prescribed dose before the pump fills
(b) Raise the replacement prescription so that convection makes up the clearance lost to diffusion
(c) Cap the dialysate below the effluent rating so that the pump keeps headroom for the other flows
(d) Balance on the pump counts so that the effluent command is no longer trimmed by the scales

5.4 Theory and calculation · 6 min

Case 1 prescribes a removal of 100 mL/h with 930 mL/h of dialysate, 1000 mL/h of replacement and 2030 mL/h of effluent by pump count.

(a) Compute the amplification $Q_k/Q_{net,Rx}$ of a relative error in each pump.*Checkpoint:* 20.3, 10.0 and 9.3

(b) The effluent pump moves 15.6 L in 8 h. Predict how much the removal error by pump count at 8 h moves per percent of effluent stroke error, and compare with Table 5.3. *Checkpoint:* 156 mL per percent

(c) At reset the true removal falls 11.2 mL/h short of the pump count. Predict the deficit after 24 h if the shortfall stayed constant, compare with the reported 466.3 mL, and name the two terms of Equation (5.1) that make the difference.*Checkpoint:* 268.8 mL

5.5 Theory and calculation · 6 min

The balance loop sees an integrator, $V_{net,m}(s) = \Delta(s)/s$ in the trim Δ, and a PI controller with gain K_c and integral time τ_I.

(a) For $K_c = 2/\text{h}$ and $\tau_I = 1\,\text{h}$, find the closed-loop poles, the damping ratio and the natural frequency. *Checkpoint:* $-1 \pm \text{j}$ per hour, 0.707, 1.41 rad/h

(b) An unmeasured removal rate steps by 150 mL/h. Find the peak error and its time after the step.*Checkpoint:* 48.4 mL at 0.785 h

(c) Repeat (b) for $K_c = 4/\text{h}$ and the same integral time. *Checkpoint:* 27.6 mL at 0.5 h

5.6 Theory and calculation · 6 min

The plausibility check of Equation (5.13) compares each scale with its pump's calibrated count over a window of 20 min, with $q_{abs} = 20\,\text{mL/h}$ and $q_{rel} = 1\,\%$, and flags a scale after two consecutive checks above the tolerance, one check a minute. All fluids weigh 1.00 kg/L.

(a) Compute the tolerance for the effluent scale at 2030 mL/h, the dialysate scale at 930 mL/h and the replacement scale at 1000 mL/h. *Checkpoint:* 40.3, 29.3 and 30.0 mL/h

(b) What is the smallest steady drift of the effluent scale that the check can flag, and how does Table 5.4 bear this out? *Checkpoint:* about 40 g/h

(c) A drift of 60 g/h starts at t_0. While the window still reaches back before t_0, the disagreement is $\delta q = \beta\,(t - t_0)/W_c$. Find the earliest time after the onset at which it exceeds the tolerance, and the earliest flag. *Checkpoint:* 13.4 min, flag from about 14.4 min

5.7 App challenge · 7 min

Open scenario 5, Scale drift, at the default seed and the default speed of 300 ×. Make no change before the first run.

(a) Run to 6 h with Scale supervision off as shipped. Read the Worst removal error on the scorecard, within 3 mL, and the range of the measured error on the first trend. *Checkpoint:* −304 mL; measured within about ±30 mL

(b) Press Reset, set Scale supervision to On before pressing Run, and run to 6 h. Read the Worst removal error and the effluent pump card. *Checkpoint:* −13 mL, effluent scale IMPLAUSIBLE

(c) Run on to 12 h and read the Worst removal error again. Explain why it grows although the drifting scale has been replaced. *Checkpoint:* −37 mL

5.8 App challenge · 8 min

Open scenario 7, Sensor dropout, at the default seed and 300 ×. The effluent scale is silent for 20 min from 1 h, and Scale supervision is off as shipped. Read each value on the Worst removal error tile after a run to 2.5 h, within 3 mL.

(a) Run as shipped. Read the tile and the lowest measured error on the first trend.*Checkpoint:* +80 mL; measured about −650 mL

(b) Press the reload button. On the Gravimetric balance card open the advanced settings, set Largest effluent trim to 1000 mL/h, and run. Reload again, set it to 100 mL/h, and run. *Checkpoint:* +209 mL and +24 mL

(c) Reload, set Scale supervision to On, and run. *Checkpoint:* −7 mL

5.9 App challenge · 8 min

Open the challenge, A day on the circuit, at the default seed. Run each day to 24 h at 900 ×, about 1.6 min a run, and read the scorecard.

(a) Run the day as shipped. Read the Worst removal error, the Delivered dose in percent, the Highest TMP and the Highest FF, and say which criteria fail.*Checkpoint:* −893 mL, 81 %, 284 mmHg, 21.1 %

(b) Reload, press Reveal suggested tuning, and run. Read the same four tiles.*Checkpoint:* −51 mL, 90 %, 206 mmHg, 22.5 %

(c) Reload, reveal the suggested tuning, set Change bags together to Off, and run. Does the day still pass?*Checkpoint:* yes, −37 mL, 88 %

5.10 Application area · 4 min

In case 4 the access override holds the measured access pressure at −150 mmHg. The blood pump then counts 111.6 mL/min while 103.5 mL/min actually passes, and the machine's filtration fraction reads 24.0 % against a true 25.4 %. Explain the discrepancy, say why it matters for the filtration-fraction ceiling, and name two ways a machine could narrow it.

5.11 Application area · 4 min

In case 6 ten bag changes pause the fluid pumps for 37.2 min in 8 h at a prescribed removal of 150 mL/h. Estimate the removal lost if nothing is paid back, compare with the reported deficit, and say what catch-up and changing the bags together each do about it.

5.12 Review the agent · 5 min

An AI assistant reviews scenario 5 and writes: *The balance loop is too slow to reject the scale drift. Raise the Balance loop gain from 2 to 6/h and cut its integral time from 1 to 0.3 h, and the true*

removal error will stay inside the band. Find the flaw, and say how you would check the claim in the app.

CHAPTER 6

Treadmill Pacing Simulator for Constraint Selection and Fixed-Horizon Optimization



The controller moves 1 motor that sets the belt speed

The controller reads 1 heart rate, from a chest strap and its receiver 2 blood lactate, which the chapter estimates

What disturbs it 1 tilt of the deck

One motor sets the belt speed for two readings of the runner, heart rate and blood lactate, while the tilt of the deck changes what every speed costs. Illustration, not to scale.

IN PLAIN TERMS Pacing a Runner on a Treadmill

The main idea. A treadmill has one control, the speed of its belt. A controller sets that speed so that a runner covers as much ground as possible in one hour without passing safe limits of the body.

The challenge. Heart rate, body heat, stored fuel and blood lactate, a sign of hard effort, all climb with speed at different rates, and the first to reach its limit must set the speed. Extra effort left unused when the hour ends is distance thrown away.

Why it matters. The same problem appears wherever one control has to respect several limits before a fixed deadline.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	The first page, the fact box of Section 6.1, the history boxes of Sections 6.2 and 6.3, the try-it boxes of Sections 6.3 and 6.7, the fact box of Section 6.7 and the key points
Study	60 min	The whole chapter with its pause points and the twelve exercises
Practice	20 min	The first page, Section 6.7, Section 6.8, Section 6.9 and the open challenge

The control question Belt speed is the only actuator of a treadmill, while blood lactate, heart rate, core temperature and fuel limit how fast it may run and answer on different time scales. How should one actuator be shared among those limits, and what is a finite store worth when the hour has a fixed end?

Learning objectives

- Compute the static gain from belt speed to heart rate, and explain why gains carried as a fraction of threshold pace serve every runner.
- Predict the time a runner lasts above critical power from the reserve and the excess power.
- Explain how a low selector with an external-reset band shares one actuator, and why tracking the losing loop to the applied output deadlocks.
- Measure what incline feedforward and an aggressive heart-rate tuning do to distance, command travel and lactate.
- Divide the predictive controller's advantage into a tuning choice and an architectural part, and judge what its terminal value is worth.

Prerequisites Sections 1.6.2 (velocity form), 1.4.4 (scheduling on a known individual number), 1.4.6 (overrides and stores) and 1.6.4 (scenarios and data). Section 1.6.3 is needed only for the predictive controller.

The app https://apmonitor.com/apps/treadmill-runner/

See also Volume 4, Chapter 4 and Volume 4, Chapter 3 for one actuator shared among limits through a selector whose losing loops track the applied output. Volume 9, Chapter 12 for slack spent against a fixed deadline, and Chapter 7 for the same athlete model behind an actuator whose authority is a store.

6.1 Introduction

Most loops in an introductory course pair one manipulated variable with one controlled variable and one setpoint. A treadmill and the person on it do not divide that way. Belt speed is the only actuator, and the quantities that decide how fast the belt may run are blood lactate, heart rate, core temperature and the amount of carbohydrate left, none of which can be moved directly and no two of which respond on the same time scale. Raising the belt raises all of them. Which one arrives at its limit first depends on the runner rather than on the tuning, and the controller is required to hold whichever one is currently binding while the others are left wherever the actuator puts them.

That structure is the ordinary case wherever one input serves several constrained outputs. A compressor holds discharge pressure until the surge line binds, a furnace holds coil outlet temperature until a tube metal thermocouple binds, and a battery charger holds current until cell voltage or cell temperature binds. The standard answer is a selector, which computes one output per constraint and passes the most conservative of them to the actuator [1]. The part a block diagram omits is what the losing controller does while it waits, and that omission is not cosmetic. A selector without external reset winds its unselected integrators to the actuator limit, and a selector whose external reset is implemented by loading the losing integrator with the applied output deadlocks instead, because the loser then reproduces the applied output exactly and wins the next comparison by an arbitrarily small margin.

The objective adds a second layer. Distance over a fixed hour grows monotonically with the belt speed that every constraint forbids, the horizon is fixed and known, and two of the stores are worth nothing once the clock stops. An anaerobic work reserve that is still full at the finish line is distance that was available and was not taken, which makes the problem an economic optimization with a terminal condition [2, 3] rather than a regulation problem with a setpoint. A receding-horizon controller whose horizon is shorter than the run has to be told what the state it leaves behind is worth, and how that valuation is scaled turns out to decide whether it helps or hurts.

Simulators of exercise are usually built to answer physiological questions, and control studies of treadmills are usually built to hold one variable. Feedback control of heart rate on a treadmill has an established literature with identified first- and second-order models and quantified disturbance

rejection [4–6], and oxygen uptake has been controlled on the same machine [7]. However, those systems regulate one physiological variable to a reference and none of them optimizes distance, so the selector between competing limits and the value of a store at a deadline never appear. This chapter works with a one-hour treadmill in which one belt speed is driven against four physiological constraints, so that constraint selection, external reset and a terminal value can be exercised on a plant whose variables a student already has an intuition about. The physiological relations are published work with recalibrated coefficients, and the control structures are standard.

> ✦ DID YOU KNOW? **A Heart Rate Held Within Two Beats**
>
> In a 2023 laboratory study, 23 healthy participants ran 35-minute bouts on a treadmill whose speed a feedback controller set to hold a constant heart rate. The mean root-mean-square tracking error was 1.98 bpm with a second-order design and 2.13 bpm with a first-order one [8].

Section 6.2 traces closed-loop treadmills and pacing models. Section 6.3 carries every equation and every constant a reader would need to build the fourteen-state plant again, with its instrument chain. Section 6.4 measures the lags of the four constrained outputs, the static gain from belt speed to heart rate and how far the sustainable speed moves in an hour. Section 6.5 sets out the two regulators, the low selector with external reset and the receding-horizon alternative. Section 6.6 defines the scorecard, which keeps the competing costs of a run apart instead of averaging them. Section 6.7 reports the sixteen cases, from a stopped belt to a timed distance target, and measures what the terminal value is worth. Section 6.8 describes the real heart-rate straps, treadmill drives and laboratory controllers. Sections 6.9 and 6.10 state where the model stops holding and where the same structure appears in other plants.

6.2 Background: Controlling a Person on a Belt

The motor-driven treadmill entered laboratory use as an instrument for imposing a known and repeatable work rate. Graded protocols made that quantitative by stepping belt speed and grade together on a fixed schedule, and the Bruce protocol and the nomogram built from it turned the time a subject lasted into an estimate of maximal oxygen intake [9]. The arrangement is open loop by design. The schedule is fixed in advance, the subject is the disturbance, and the endpoint is the moment the subject stops. Nothing in the protocol adapts to the individual, which is what makes the test comparable across subjects and also what makes it unable to hold any one of them at a chosen physiological intensity.

Closing a loop around the belt came later and came from rehabilitation and exercise-prescription requirements, where the quantity that matters is the intensity the patient is working at rather than the speed the machine is running. Heart rate was the first controlled variable, because it is the one an inexpensive instrument reports continuously. How its models and controllers developed, from identified models to predictive control, is told in the history box of this section. Later work addressed the disturbance that dominates the loop in practice, which is very-low-frequency heart-rate variability rather than measurement noise, by shaping the input sensitivity directly [6]. Oxygen uptake has been controlled on the same machine, and it is the harder loop because the measurement is slower and noisier than heart rate [7].

What that literature does not address is the question a runner actually asks, which is how fast to go for a fixed time. Pacing has its own modeling tradition, built on an energy-flow account in which a finite anaerobic store is spent against an aerobic supply and the optimal strategy follows from how the two are scheduled over the event [3]. Those models optimize and do not regulate. The application of this chapter joins the two, putting a regulator and a receding-horizon optimizer on the same actuator and the same plant so that the difference between holding a constraint and scheduling a store can be measured on one run.

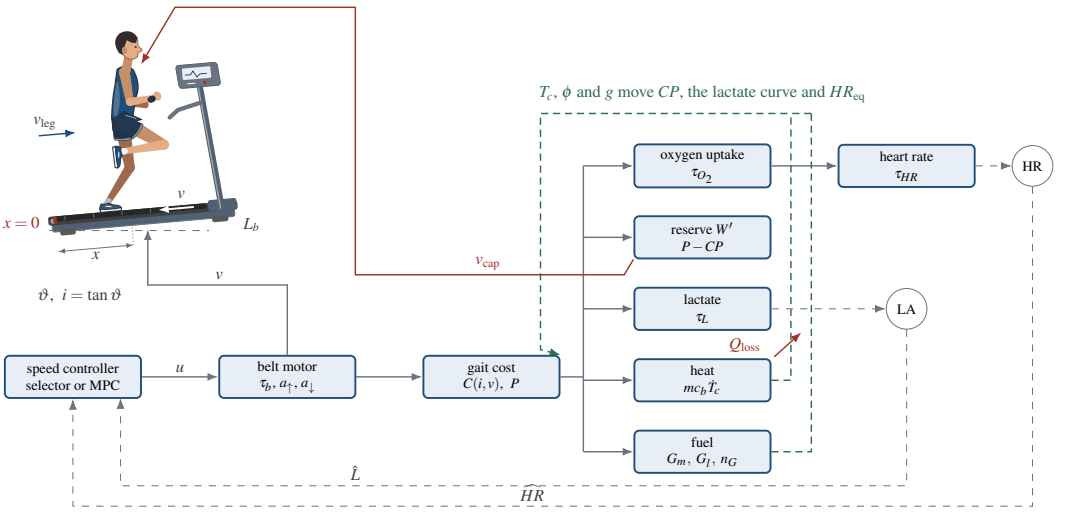


Figure 6.1 One runner on one treadmill. The belt motor follows the controller output u with lag τ_b and rate limits $a_\uparrow$ and $a_\downarrow$, the deck sits at angle ϑ of fractional grade i, and the runner stands a distance x from the back edge of a belt of usable length L_b. The legs run at v_{leg}, which equals the belt unless the belt asks for more than the capability v_{cap} the anaerobic reserve currently supports, and the run ends at $x = 0$. Gait cost $C(i, v)$ sets the metabolic demand P, which drives all five subsystems in parallel. Core temperature, accumulated fatigue and the remaining carbohydrate fraction feed back into the sustainable power, the lactate curve and the heart-rate equilibrium, drawn here as one bundle. The controller reads a noisy filtered heart rate $\widehat{HR}$ and a lactate estimate $\hat{L}$ and commands belt speed alone.

> **HISTORY** **Closing the Loop Around the Belt (2007 to 2021)**
>
> In 2007 a computer-controlled treadmill held a person's heart rate on a preset profile by changing the speed of its belt. Its designers described the heart-rate response as a static nonlinearity at the input followed by linear dynamics, identified it with a pseudorandom test signal, and compared a robust controller designed on that model with a conventional PID loop [4]. A year later a nonlinear model, identified from subjects walking on a treadmill at different speeds, carried a controller with feedforward and feedback parts that followed two different heart-rate profiles [5]. Predictive control came in 2021, on models estimated from twelve participants with a short protocol, and with the speed and acceleration of the treadmill kept within a safe range [10].

6.3 Plant Model

The plant is one runner on one treadmill for 3600 s from a standing start. Belt speed v is the only manipulated variable and is bounded at 6.7056 m/s, which is 15 mph. The deck is tilted to an angle ϑ that is either held or redrawn every 300 s from a seeded schedule. The runner occupies a position x measured from the back edge of a 1.50 m belt, and the run ends when x reaches zero. Figure 6.1 shows the arrangement and names every symbol that appears in the equations. Units are SI internally, with heart rate kept in beats per minute and lactate in mmol/L because every limit quoted for them is in those units. Display converts speed to mph, km/h and pace. The physiology below the legs, from oxygen uptake to the fuel stores, is a shared athlete model that the cycling simulator of Chapter 7 also runs. The treadmill supplies its gait cost as the metabolic demand and charges the anaerobic reserve in metabolic watts, and moving the equations into the shared model reproduced 132 recorded checkpoints of this application over 23 runs bit for bit.

The roles in Table 6.1 divide the fourteen states as follows. Two belong to the machine, being

Table 6.1 Classification of Variables for the Recreational Male Preset at Age 35 on the Level. Nominal Values Are the Threshold-Pace Steady State Where One Exists and the Measured Value After One Hour Under the Selector Otherwise.

Symbol	Role	Nominal	Range
HR	CV, measured	160 bpm	154–170 bpm band
L	CV, estimated	4.0 mmol/L	1.5–6 mmol/L SP
v	MV	2.787 m/s	0–6.7056 m/s
s	objective	9585 m	—
ϑ	DV, measured	0°	−3 to +8°
T_{air}, φ, a	DV, measured	20 °C, 0.45, 0.60	see Table 6.2
T_c	state, shown	38.57 °C	34–41.5 °C
W'	state, shown	189 m	0–190 m
$G_m + G_l$	state, shown	222 g	0–361 g
x	state, shown	0.75 m	0–1.50 m
$\dot{V}_{O_2}$	state, shown	959 W	91–1176 W
ϕ	state, hidden	0.005	0–0.40
n_G	state, shown	93.8 mmol	$\geq$ 42.2 mmol
G_u	state, shown	0 g	40 g per feed
f_R	state, shown	46.9/min	5–75/min

the belt speed and the runner's place on it, five are metabolic, four are fuel, one is thermal, one is respiratory and one is the distance integral. The fourth fuel state is carbohydrate eaten during the run and still in the gut. Only accumulated fatigue ϕ is withheld from every display, because its effect is meant to be inferred from the pace that a runner can still hold rather than read off a gauge.

6.3.1 Cost of Locomotion and Metabolic Demand

The energy cost of covering ground is taken from the measured relation for walking and running against gradient, which is a fifth-order polynomial in the fractional grade with a published level value of 3.6 J/(kg m) for running [11]. Each runner carries one economy parameter C_r that stretches the whole curve, and the running and walking polynomials are blended below 2.4 m/s so that the transient from a standing start is not an extrapolation of a running fit through a gait nobody is using, as shown in Equation (6.1).

$$\begin{aligned} C_R(i) &= 3.6 + 19.5i + 46.3i^2 - 43.3i^3 - 30.4i^4 + 155.4i^5 \\ C_W(i) &= 2.5 + 19.6i + 51.9i^2 - 76.8i^3 - 58.7i^4 + 280.5i^5 \\ w(v) &= \xi^2(3 - 2\xi), \qquad \xi = \operatorname{clamp}\left(\tfrac{v-1.6}{0.8}, 0, 1\right) \\ C(i,v) &= \max\left(0.6,\ \tfrac{C_r}{3.6}\left[0.9\,(1-w)\,C_W(i) + w\,C_R(i)\right]\right) \end{aligned} \tag{6.1}$$

where $i = \tan\vartheta$ is the fractional grade, not the angle and not a percentage, and C carries units of J/(kg m). The distinction matters numerically. At $\vartheta = 5^\circ$ the grade is 8.75 % and (6.1) returns 6.49 J/(kg m) for a runner whose level economy is 4.15 J/(kg m), which is 56 % more energy for the same meter covered.

Metabolic demand is the resting turnover plus the cost of the ground covered, as shown in Equation (6.2).

$$P = P_{\text{rest}} + m\,C(i, v_{\text{leg}})\,v_{\text{leg}}, \qquad P_{\text{rest}} = \frac{3.5\,\varepsilon_{O_2}}{60}\,m \tag{6.2}$$

where m is body mass in kg, $\varepsilon_{O_2} = 20.9\,\text{J/mL}$ is the energy equivalent of oxygen and 3.5 mL/(kg min) is one metabolic equivalent. Aerobic capacity follows the same conversion, as shown in Equa-

tion (6.3).

$$\begin{aligned} P_{\max} &= \frac{m\dot{V}_{O_2,\max}\,\varepsilon_{O_2}}{60} \\ f(P) &= \max\!\left(0,\ \frac{P-P_{\text{rest}}}{P_{\max}-P_{\text{rest}}}\right) \end{aligned} \tag{6.3}$$

where f is the fraction of the oxygen-uptake reserve a metabolic power represents and is the single intensity variable the rest of the model is written against. For the recreational male preset, $m = 75\,\text{kg}$ and $\dot{V}_{O_2,\max} = 45\,\text{mL}/(\text{kg}\,\text{min})$ give $P_{\text{rest}} = 91.4\,\text{W}$ and $P_{\max} = 1175.6\,\text{W}$.

6.3.2 Oxygen Uptake, Critical Power and the Anaerobic Reserve

Oxygen uptake is carried as a delivered power and follows demand with a first-order lag, saturating at the aerobic ceiling, as shown in Equation (6.4).

$$\tau_{O_2}\frac{\mathrm{d}\dot{V}_{O_2}}{\mathrm{d}t} = \min(P,\ P_{\max}) - \dot{V}_{O_2} \tag{6.4}$$

Anything demanded above $P_{\max}$ is supplied anaerobically and is accounted separately. The sustainable power is the critical-power construct, which originated as the asymptote of the work-time relation for a muscle group [12] and is now the standard two-parameter description of exercise tolerance, with a finite work reserve W' above that asymptote [13]. The application makes the sustainable power a function of the damage already done, as shown in Equation (6.5).

$$\begin{aligned} CP &= P_{\text{rest}} + f_{CP}\,(P_{\max}-P_{\text{rest}})\,(1-\phi)\,\Psi \\ \Psi &= \Psi_G\,\Psi_c, \qquad g = \frac{G_m+G_l}{G_{m,0}+G_{l,0}} \\ \Psi_G &= 0.80 + 0.20\,\text{clamp}\!\left(\tfrac{g}{0.15},0,1\right) \\ \Psi_c &= 0.85 + 0.15\,\text{clamp}(c_G-3,\,0,\,1) \end{aligned} \tag{6.5}$$

where f_{CP} is fresh critical power as a fraction of the oxygen-uptake reserve, $\phi \in [0, \phi_{\max}]$ is accumulated fatigue, g is the fraction of the starting carbohydrate store remaining and c_G is blood glucose in mmol/L. The fuel factor Ψ is unity until the store falls below 15 %, which is why an hour begun with full stores has an almost fixed sustainable power and an hour begun short of fuel does not. Peak power and the power the runner can actually produce at an instant follow in Equation (6.6).

$$\begin{aligned} P_{\text{peak}} &= P_{\text{rest}} + \mu\,(P_{\max}-P_{\text{rest}})\left(1-\tfrac{\phi}{2}\right) \\ P_{\text{av}} &= CP + (P_{\text{peak}}-CP)\left(\frac{W'}{W'_{\max}}\right)^{0.4} \end{aligned} \tag{6.6}$$

where $\mu = 1.35$. The two-parameter model in its textbook form holds any power above CP until W' reaches zero and then drops to CP instantly [13]. That is an accurate description of time to exhaustion and an unusable one here, because a hard switch moved the runner from holding the belt to off the back of it inside two integration steps with no trend showing why. The fractional exponent in (6.6) spreads the loss of capability over most of the reserve. The reserve itself is declared as a distance D', because that is the unit it is spent in, and drains at exactly the excess power, as shown in Equation (6.7).

$$\begin{aligned} W'_{\max} &= m\,C_r\,D' \\ \frac{\mathrm{d}W'}{\mathrm{d}t} &= \begin{cases} -(P-CP), & P > CP \\ (CP-P)\,k_r\left(1-\dfrac{W'}{W'_{\max}}\right), & P \le CP \end{cases} \end{aligned} \tag{6.7}$$

where k_r is the recovery efficiency. Above CP the time to an empty reserve is W' divided by the excess power, which is the arithmetic the belt turns into a consequence.

6.3.3 Heart Rate

Heart rate is driven by the oxygen uptake actually delivered rather than by the demand, using the equivalence of heart-rate reserve and oxygen-uptake reserve [14], with separate onset and recovery lags and two drift terms, as shown in Equations (6.8) and (6.9).

$$\begin{aligned} HR_{\text{eq}} &= HR_{\text{rest}} + (HR_{\max} - HR_{\text{rest}})\, f(\dot{V}_{O_2}) \\ &\quad + k_T \max(0, T_c - 37.8) + k_F \frac{\phi}{\phi_{\max}} \end{aligned} \tag{6.8}$$

$$\tau_{HR} \frac{\mathrm{d}HR}{\mathrm{d}t} = \text{clamp}(HR_{\text{eq}},\ 35,\ 1.02\, HR_{\max}) - HR \tag{6.9}$$

where τ_{HR} is $\tau_{HR\uparrow}$ when HR_{eq} exceeds HR and $\tau_{HR\downarrow}$ otherwise, and $HR_{\max} = \text{clamp}(220 - \text{age} + \Delta_{HR},\ 120,\ 220)$. The two drift terms in (6.8) are the reason a constant belt speed does not produce a constant heart rate. The estimator $220 - \text{age}$ is retained because every band quoted in practice is written against it, and an editable offset Δ_{HR} is provided because that estimator carries a population standard deviation near 11 bpm and systematically misestimates older adults [15]. The threshold band the application draws is 83 to 92 % of $HR_{\max}$, which is 154 to 170 bpm at age 35.

6.3.4 Blood Lactate

Blood lactate is one well-mixed compartment whose concentration is the balance of appearance and disappearance rather than an indicator of an oxygen shortfall, which is the modern reading of lactate kinetics [16]. The steady-state concentration is a power law in the intensity above a knee, calibrated so that it balances at the resting value with the runner standing still and crosses the limit exactly at the preset's own threshold intensity, as shown in Equation (6.10).

$$\begin{aligned} L_{\text{ss}}(f) &= L_{\text{rest}} + (L_4 - L_{\text{rest}}) \left[\max\left(0, \tfrac{f - f_0}{f_{T,\text{eff}} - f_0}\right) \right]^q \Lambda \\ \tau_L \frac{\mathrm{d}L}{\mathrm{d}t} &= L_{\text{ss}}(f(P)) - L \end{aligned} \tag{6.10}$$

where f_0 is the intensity below which lactate does not rise, q places the knee, $\Lambda = 0.6 + 0.4\,\text{clamp}(g/0.2, 0, 1)$ carries the effect of a depleted store and $f_{T,\text{eff}} = \max\big(0.40,\ f_T(1-\phi)\Psi\big)$ is the intensity at which the curve crosses the limit, with f_T the fresh lactate marker. The marker is scaled by the same fatigue and fuel factors as critical power in (6.5), so that a tired runner's lactate does not read like a fresh one's. The constraint $L_4 = 4\,\text{mmol/L}$ is the conventional working threshold [17]. The marker f_T and the critical-power fraction f_{CP} are separate parameters, because critical power and the lactate threshold markers correlate closely without being interchangeable [18], and every preset sets them equal. For the presets the lactate constraint and the reserve constraint are then one constraint, which is the teaching simplification that keeps the selector legible, and editing either parameter separates them.

> ⧖ HISTORY **Four Millimoles on the Treadmill (1976 to 1985)**
>
> The lactate limit of 4 mmol/L has a laboratory history. In 1976 a group of physiologists in Cologne placed the change from purely aerobic to partially anaerobic metabolism at an earlobe blood lactate of 4 mmol/L, without data to support it. From 1979 the group tested the claim on a treadmill. Each participant ran 25 min at one constant speed, stopping for 30 s every 5 min for an earlobe sample, and several such runs near the threshold were needed. In 15 participants lactate stayed constant at intensities where it reached 4.0 ± 0.7 mmol/L. The data appeared in German in 1982, and the English term maximal lactate steady state followed in 1985. The test is too long for routine testing and remains the gold standard against which quicker tests are checked [19].

6.3.5 Fatigue, Heat and Respiration

Accumulated fatigue is the mechanism by which sustained lactate and sustained thermal strain reduce the sustainable power in (6.5), as shown in Equation (6.11).

$$\begin{aligned} \Gamma &= k_L \frac{\max(0, L - L_4)}{L_4} + k_{TF} \max(0, T_c - 38.5) \\ \frac{\mathrm{d}\phi}{\mathrm{d}t} &= \Gamma \left(1 - \frac{\phi}{\phi_{\max}}\right) - k_\phi \, \phi \, \mathrm{clamp}\left(1 - \tfrac{P}{CP}, 0, 1\right) \end{aligned} \tag{6.11}$$

The thermal term is present because a rising core temperature reduces the work a person can sustain independently of the metabolic state [20]. The heat balance itself is one lumped body node written about the resting equilibrium, as shown in Equation (6.12).

$$\begin{aligned} m\, c_b \frac{\mathrm{d}T_c}{\mathrm{d}t} &= \max(0,\ P - P_{\mathrm{mech}}) - Q_{\mathrm{loss}} \\ Q_{\mathrm{loss}} &= P_{\mathrm{rest}} + G_d \,(T_c - 37.0) \\ &\quad + \min\left(G_e (T_c - 37.05),\ E_{\max}\right) \end{aligned} \tag{6.12}$$

where $P_{\mathrm{mech}} = m\, g\, v_{\mathrm{leg}} \sin\vartheta$ for $\vartheta > 0$ is the mechanical work leaving the balance against gravity, and c_b is the specific heat of the body. Writing the loss term about 37.0 °C makes a standing runner in a 20 °C room exactly steady rather than drifting, which is verified as an equality in the application's own test suite. The conductances scale with the runner and the room, as shown in Equation (6.13).

$$\begin{aligned} \Phi &= \frac{m \dot{V}_{O_2,\max}}{75 \times 45} \\ G_d &= (g_0 + g_a a) \frac{m}{75} \, \mathrm{clamp}\left(1 + \tfrac{20 - T_{\mathrm{air}}}{25}, 0.3, 1.8\right) \\ G_e &= k_s \, \Phi \, (1 - 0.55\varphi)(0.45 + 0.55a) \\ E_{\max} &= E_0 \, \Phi \, (1 - 0.6\varphi)(0.45 + 0.55a) \end{aligned} \tag{6.13}$$

where $a \in [0, 1]$ is airflow over the runner, φ is relative humidity and Φ is aerobic capacity relative to the default preset. Both sweating terms scale with Φ because an elite runner produces about 1.6 times the heat of a recreational runner in a smaller body, and a fixed evaporative ceiling left that runner with no thermal steady state at their own threshold pace. Respiratory frequency is a first-order response to intensity with additive terms in lactate and temperature, as shown in Equation (6.14), and is displayed rather than controlled.

$$\begin{aligned} f_{R,\mathrm{eq}} &= f_{R,0} + k_R f(\dot{V}_{O_2}) \\ &\quad + 2\max(0, L - 2) + 3\max(0, T_c - 38) \\ \tau_R \frac{\mathrm{d}f_R}{\mathrm{d}t} &= \min(70,\ f_{R,\mathrm{eq}}) - f_R \end{aligned} \tag{6.14}$$

6.3.6 Substrate Selection and the Carbohydrate Stores

Fat and carbohydrate are oxidized concurrently with a split that moves with intensity, and fat oxidation carries a rate ceiling that carbohydrate does not [21]. The split and the three fuel states are given by Equations (6.15) and (6.16).

$$\begin{aligned} \chi &= \mathrm{clamp}(0.20 + 0.78 f,\ 0.15,\ 0.98) \\ P_F &= \min\left(P_{F,\max}\left[0.7 + 0.3\left(1 - \tfrac{G_m}{G_{m,0}}\right)\right],\ (1 - \chi) P_{\mathrm{del}}\right) \\ P_C &= P_{\mathrm{del}} - P_F, \qquad \dot{n}_C = \frac{P_C + P_{\mathrm{an}}}{e_G} \end{aligned} \tag{6.15}$$

where $P_{\text{del}} = \min(P, P_{\max})$, $P_{\text{an}} = \max(0, P - P_{\text{del}})$ is the anaerobic contribution, and $e_G = 3153\,\text{J/mmol}$ is the metabolizable energy of a millimole of glucose equivalent. Counting P_{an} as carbohydrate is what keeps the anaerobic reserve from being free energy.

$$
\begin{aligned}
\beta &= \operatorname{clamp}\Big(0.18 + 0.35\left(1 - \tfrac{G_m}{G_{m,0}}\right), 0,\ 0.6\Big) \\
\frac{\mathrm{d}G_m}{\mathrm{d}t} &= -\dot{n}_C(1-\beta)M_G \\
\frac{\mathrm{d}G_l}{\mathrm{d}t} &= -\dot{n}_H M_G, \qquad \frac{\mathrm{d}G_u}{\mathrm{d}t} = -\frac{G_u}{\tau_u} \\
\frac{\mathrm{d}n_G}{\mathrm{d}t} &= \dot{n}_H + \dot{n}_g - \dot{n}_C\beta + \frac{G_u}{\tau_u M_G} \\
&\quad - k_u \max\left(0,\ \frac{n_G}{V_G m} - c_{G,0}\right)
\end{aligned}
\tag{6.16}
$$

where $M_G = 0.180\,16\,\text{g/mmol}$, β is the share of the carbohydrate flux drawn from circulating glucose and rises as muscle glycogen falls, and the endogenous supply is gluconeogenesis first and hepatic glycogenolysis for the remainder, as shown in Equation (6.17).

$$
\begin{aligned}
\dot{n}_w &= \max\big(0,\ \dot{n}_C\beta + 0.02(c_{G,0}V_G m - n_G)\big) \\
\dot{n}_g &= \min(k_g,\ \dot{n}_w) \\
\ell &= \operatorname{clamp}\Big(\tfrac{G_l}{0.15\,G_{l,0}}, 0, 1\Big) \\
\dot{n}_H &= \operatorname{clamp}(\dot{n}_w - \dot{n}_g,\ 0,\ \dot{n}_{H,\max}\,\ell)
\end{aligned}
\tag{6.17}
$$

Writing gluconeogenesis as an unconditional source instead let blood glucose climb while the runner stood still, which is an unclosed balance rather than a teaching point. Carbohydrate eaten during the run enters the gut store G_u and empties into the circulating pool with $\tau_u = 600\,\text{s}$, and circulating glucose above its starting concentration is taken up at $k_u = 0.12\,\text{L/s}$ per unit of excess. This sink stands in for the hormonal response without modeling it. Its only job is to stop a meal from piling up in the small circulating pool, and it does nothing while glucose sits at or under the starting concentration, which is the case for the whole of every scenario without a feed. Glucose it clears is burned instead of fat, and since fat is not a tracked store the carbohydrate accounting still closes. Each feed places 40 g in the gut. In prolonged exercise a carbohydrate feed keeps plasma glucose from falling without sparing muscle glycogen [22], and the gut store behaves the same way. The store is calibrated so that level running at threshold pace exhausts it after a documented distance, which is what the fuel gauge reads at full scale, and the exhaustion distance is checked against the integrated model rather than asserted. Progressive glycogen depletion at a fixed work rate and the pace decline that follows it are the behavior this reproduces [22].

6.3.7 Belt, Position and the Two Moving Ceilings

The belt is a first-order motor with separate acceleration and deceleration limits, and the runner's position on it is the state that ends the run, as shown in Equation (6.18).

$$
\begin{aligned}
\frac{\mathrm{d}v}{\mathrm{d}t} &= \operatorname{clamp}\left(\frac{u - v}{\tau_b},\ -a_{\downarrow},\ a_{\uparrow}\right), \qquad v \in [0,\ v_{\max}] \\
v_w &= v + \operatorname{clamp}\big(k_c(x_0 - x),\ -0.25,\ 0.7\big) \\
v_{\text{leg}} &= \max\big(0,\ \min(v_w,\ v_{\text{cap}})\big) \\
\frac{\mathrm{d}x}{\mathrm{d}t} &= \operatorname{clamp}(v_{\text{leg}} - v,\ -3,\ 3), \qquad \frac{\mathrm{d}s}{\mathrm{d}t} = v_{\text{leg}}
\end{aligned}
\tag{6.18}
$$

where v_{cap} is the speed corresponding to P_{av} of (6.6) on the grade underfoot, k_c is how hard the runner works to stay centered, and x_0 is the starting position. When the belt asks for more than the runner can produce, v_{leg} falls short of v, the runner travels backward along the deck, and the run ends at $x = 0$. The failure mode is therefore a position and not an overshoot on a trend.

> **TRY IT IN THE APP** **2 · A Speed Step by Hand**
>
> Load scenario 2 and press Run, then watch the runner's position on the belt. The belt steps to 9.36 km/h at 60 s and everything settles. At 15 min it steps to 11.16 km/h, above critical power, and nothing settles: the runner drifts back along the deck and goes off the back of the belt at 24:28 with 3.93 km covered.

Both scored ceilings open over a finishing window, as shown in Equation (6.19), because a reserve that is still full when the clock stops was distance that was available and not taken.

$$\begin{aligned} \zeta &= \operatorname{clamp}\left(\frac{t-(T-t_w)}{t_w},\ 0,\ 1\right) \\ L_{\text{lim}}(t) &= L_4 + (L_f - L_4)\,\zeta^2 \\ HR_{\text{lim}}(t) &= HR_{\text{hi}} + (HR_{\text{max}} - HR_{\text{hi}})\,\zeta^2 \end{aligned} \tag{6.19}$$

With $T = 3600\,\text{s}$ and $t_w = 300\,\text{s}$, the lactate ceiling holds at 4.0 mmol/L for 55 min and reaches 6.0 mmol/L at the finish, and the heart-rate ceiling rises from 170 bpm to 185 bpm over the same window at age 35. The quadratic shape gives the ceiling a continuous value and slope where it starts to open.

6.3.8 Parameters, Presets and Numerics

Table 6.2 lists the constants of the default preset. Every one of them is adjustable in the application and all but one are displayed. Four presets write mass, aerobic capacity, economy, the lactate marker, the critical-power fraction, the reserve, the lag constants and the fuel stores in one action, and the level threshold pace each of them implies is measured from the model rather than asserted at it, at 10.03 km/h for the recreational male preset, 8.72 km/h for the recreational female preset, 20.26 km/h for the elite male preset and 17.20 km/h for the elite female preset. The relative intensity at 4 mmol/L is deliberately not given a difference between the male and female presets.

The engine of Section 1.6.1 advances the state vector by fourth-order Runge-Kutta at a fixed 0.25 s step. The regulators execute once per simulated second, the optimizer replans once per five, and the historian keeps one sample per second. Two independent seeded streams supply the randomness, one for the incline schedule and one for the instrument noise, so turning the noise up leaves the hills exactly where they were. The playback multipliers are 1, 5, 20 and 60.

6.3.9 Assumptions

The model is a teaching representation of a control problem. It is not a training plan, not a fitness assessment and not medical advice, and no number in it describes any individual. Each simplification fails somewhere.

Critical power and the 4 mmol/L intensity are equal in every preset. They are close in practice and are not identical, and neither construct has a single agreed determination [13, 17]. Setting them equal collapses the lactate constraint and the reserve constraint into one constraint for the presets, which is the simplification that makes the selector legible. The two are separate parameters, so a student who edits one can study the difference, and the presets cannot show it on their own.

Lactate is one well-mixed compartment. There is no muscle-to-blood transport delay, no shuttling between fiber types and no distinction between arterial and venous sampling, all of which are real and all of which would make the measurement later than the 250 s already modeled [16]. The application understates how far behind the legs the measurement is.

Table 6.2 Model Parameters of the Recreational Male Preset. All Are Adjustable. Only Accumulated Fatigue Is Withheld From Every Display.

Symbol	Meaning	Value	Unit
m	body mass	75	kg
age	age	35	years
$\dot{V}_{O_2,\max}$	maximal oxygen uptake	45	mL/(kg min)
C_r	level running economy	4.15	J/(kg m)
f_T	intensity at 4 mmol/L	0.80	—
f_{CP}	critical power, fraction of reserve	0.80	—
D'	anaerobic distance reserve	190	m
HR_{rest}	resting heart rate	60	bpm
Δ_{HR}	offset from 220 − age	0	bpm
τ_{O_2}	oxygen-uptake lag	35	s
$\tau_{HR\uparrow}$	heart-rate onset lag	45	s
$\tau_{HR\downarrow}$	heart-rate recovery lag	70	s
τ_L	lactate clearance	250	s
τ_R	respiratory lag	20	s
k_T	drift with core temperature	7	bpm/°C
k_F	drift at full fatigue	10	bpm
f_0	lactate knee	0.35	—
q	lactate curvature	3.0	—
k_r	reserve recovery efficiency	0.5	—
μ	peak power multiple	1.35	—
k_L	fatigue from lactate	1.7×10^{-4}	1/s
k_{TF}	fatigue from heat	4.6×10^{-5}	1/(s °C)
k_ϕ	fatigue recovery	2.0×10^{-4}	1/s
$\phi_{\max}$	largest power loss	0.40	—
T_{air}	room temperature	20	°C
φ	relative humidity	0.45	—
a	airflow	0.60	—
c_b	specific heat of the body	3500	J/(kg K)
k_s	evaporative gain	900	W/K
E_0	evaporative ceiling	1820	W
g_0	dry loss in still air	16	W/K
g_a	extra dry loss at full airflow	22	W/K
$G_{m,0}$	working-muscle glycogen	295	g
$G_{l,0}$	liver glycogen	66	g
$c_{G,0}$	starting blood glucose	5.0	mmol/L
V_G	glucose distribution volume	0.25	L/kg
k_g	gluconeogenesis	0.02	mmol/s
$\dot{n}_{H,\max}$	largest hepatic output	0.18	mmol/s
$P_{F,\max}$	fat oxidation ceiling	420	W
τ_u, k_u	gastric emptying, glucose uptake	600, 0.12	s, L/s
$v_{\max}$	top belt speed	6.7056	m/s
$a_\uparrow$, $a_\downarrow$	belt rate limits	0.35, 0.50	m/s^2
τ_b	belt motor lag	0.6	s
L_b, x_0	belt length, start position	1.50, 0.75	m
k_c	recentering effort	0.5	1/s
L_4, L_f	lactate constraint, finish	4.0, 6.0	mmol/L
t_w	finishing window	300	s
T	length of the run	3600	s

Heart rate responds to intensity, temperature and fatigue only. Hydration, posture, caffeine, altitude and the last meal do not exist here, and the very-low-frequency variability that dominates a real heart-rate control loop is represented only as filtered measurement noise [6]. The estimator 220 − age places an individual's band up to ten beats either side of where the application draws it [15].

Running economy is one coefficient. Real economy varies with speed, surface, footwear and fatigue

Table 6.3 Response to a Belt Step From Rest to 9.4 km/h at $t = 60$ s, Recreational Male Preset on the Level. Final Values Are Read at 60 min, and Times From the Step to 63 % of the Total Change Are Read at the 0.25 s Integration Step.

Output	Initial	Final	t_{63} (s)
Oxygen uptake (mL/(kg min))	3.50	34.48	40
Heart rate (bpm)	60.0	158.2	96
Blood lactate (mmol/L)	1.00	3.05	257
Core temperature (°C)	37.00	38.50	475

in ways a single multiplier on the gradient curve cannot carry, and (6.1) is a fit to treadmill data at steady speeds [11].

The thermal model has one body node. There is no skin compartment, no sweat-rate dynamics, no dehydration and no cardiovascular consequence of fluid loss, so a hot hour goes wrong more slowly here than it does in practice [20]. The coefficients in (6.13) are teaching calibrations and are not validated fits.

Feeding is carbohydrate alone. A feeding event empties from the gut with one time constant, glucose above the resting value is removed by a proportional sink rather than an insulin response, and there is no drink and no refilling of muscle glycogen during the run, so a fed hour postpones the fall in sustainable pace rather than removing it [22].

Falling off the back is modeled as a position reaching zero. No real treadmill permits that, because a safety key stops the belt first. The state is in the model because the consequence of demanding more than the runner can produce has to land somewhere other than a performance metric.

Repeatability is unrealistic. Every run of a given scenario returns the same numbers to the last digit, which is convenient for marking and is something no physiological experiment has ever managed.

6.4 Dynamic Analysis

6.4.1 Four Lags in Series on One Actuator

The distinctive dynamic feature of this plant is that one input reaches four constrained outputs through lags that differ by an order of magnitude. Table 6.3 reports the time each output takes to complete 63 % of its total change after the belt is stepped from rest to 9.4 km/h, measured in the application from a single run, and Figure 6.2 sets them on one time axis.

The ordering in Table 6.3 is the control problem. The legs are doing the new work within a second, the fastest instrument that reports anything is 96 s behind, the constraint that actually binds is 257 s behind, and the constraint with no actuator behind it is 475 s behind. By the time the fast variable reports that the belt went too high, the slow one has already begun to charge for it. A loop closed on lactate cannot be faster than 257 s without chasing a measurement that has not happened yet, and a loop closed on heart rate sees 35 s and 45 s of physiological lag in series with a 6 s measurement filter.

6.4.2 Static Gain From Belt Speed to Heart Rate

Linearizing (6.2), (6.3) and (6.8) at a steady speed above the walking blend gives the static gain in Equation (6.20), because C is independent of speed once $w = 1$.

$$\frac{\partial HR}{\partial v} = \frac{HR_{\max} - HR_{\text{rest}}}{P_{\max} - P_{\text{rest}}}\, m C_r \tag{6.20}$$

Equation (6.20) evaluated at the recreational male preset's threshold speed of 2.787 m/s gives $125/1084.2 \times 75 \times 4.15 = 35.9$ bpm per m/s, confirmed by a central difference on the model at 35.9 bpm per m/s. The same expression gives 19.3 bpm per m/s for the elite male preset and 41.3 bpm per m/s for the recreational female preset, a spread of 2.1 across the four presets. Mass and economy

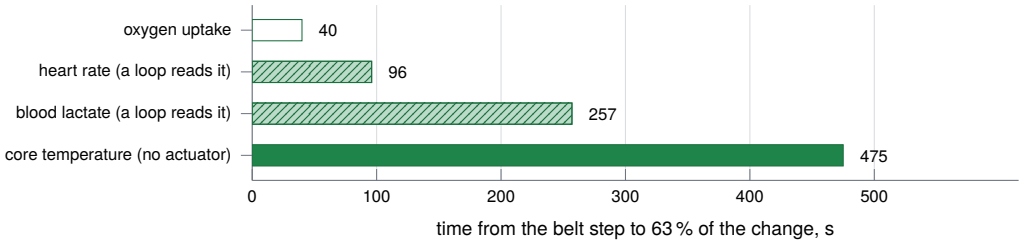


Figure 6.2 The four lags of Table 6.3 on one time axis, after a belt step from rest to 9.4 km/h. Hatched bars are the two outputs the loops read, the white bar is oxygen uptake and the solid bar is core temperature. Notice that the constraint that binds, lactate, answers 257 s after the belt moves, and a loop closed on it cannot be faster than that without chasing a measurement that has not happened yet.

Table 6.4 Movement of the Sustainable Speed Over One Hour, Recreational Male Preset Under the Selector, Read at 10 min and 60 min.

Condition	v_{sus} (km/h)	Change	T_c (°C)	$\phi / \phi_{\max}$
Level, 20 °C, fueled	10.03 → 9.98	−0.5 %	38.57	1 %
28 °C, 60 % RH, $a = 0.15$	10.03 → 8.88	−11.4 %	40.26	29 %
One third of the stores	10.03 → 8.00	−20.2 %	38.30	0 %

enter the numerator and aerobic capacity enters the denominator, so a fitter runner in a lighter body presents a plant gain less than half that of a recreational runner. A fixed controller gain in m/s per bpm therefore cannot serve both, which is the reason the tuning is scheduled.

6.4.3 When the Constraint Moves and When It Does Not

The sustainable speed is the speed at which (6.10) settles exactly on the limit, and it moves only through ϕ and Ψ in (6.5). Table 6.4 measures how far it actually travels over an hour in three conditions, read at 10 min and at 60 min.

Table 6.4 corrects an intuition that the application is often used to test. On a level hour in a cool room with full stores the constraint is very nearly fixed, falling 0.5 % over the whole hour, because the fatigue term reaches only 1 % of its ceiling and the fuel factor Ψ in (6.5) stays at unity while the store remains above 15 %. The constraint moves when a store is actually being exhausted. In the warm room the thermal term in (6.11) takes 29 % of the sustainable power and the sustainable speed falls 11.4 %, and starting with one third of the stores costs 20.2 % with the core temperature almost unchanged. The two mechanisms are separable on the trends and they are not separable on the pace, which is the diagnostic point.

The under-fueled hour also shows that the mechanism is a rate ceiling and not a shortage of energy. The belt averages 9.69 km/h from 10 to 20 min, exactly as it does with full stores, with the gauge falling steadily and nothing else out of range. At 30 min the muscle store has reached 57 g, blood glucose has left 5.00 mmol/L and the sustainable speed has begun to fall. By 35 min the belt is at 8.50 km/h and blood glucose is at 2.55 mmol/L. Fat oxidation continues throughout and the fat store is effectively unbounded, but the ceiling $P_{F,\max}$ in (6.15) caps what it can deliver per second, so the pace falls with the store and does not recover inside the hour.

6.5 Controller Design

Three modes drive the same actuator. Manual passes an operator command through to the belt. The regulatory mode runs two proportional-integral loops through a low selector. The predictive mode replaces both with a receding-horizon search. All three share the actuator handling and the instrument chain.

6.5.1 Instruments

The heart-rate measurement carries colored noise and a first-order filter of 6 s, as shown in Equation (6.21).

$$\begin{aligned} \widehat{HR}_k &= \alpha_H \widehat{HR}_{k-1} + (1-\alpha_H)\,(HR + \sigma_H n_k) \\ \alpha_H &= \mathrm{e}^{-\Delta t_c/\tau_H} \end{aligned} \tag{6.21}$$

where σ_H defaults to 0.8 bpm and n_k is drawn as white noise and then passed through a low-pass filter of its own, with the amplitude rescaled afterward so that the stationary standard deviation of the result is still unity. Advancing α_H by the exponential rather than by a forward difference keeps the filter stable at any sample period. The lactate signal is produced the same way with a 20 s filter and a default error of 0.06 mmol/L, and it is labeled on every display as an estimate. There is no continuous blood-lactate sensor and no glycogen sensor, and the application does not imply either instrument.

6.5.2 Two Regulators and a Low Selector

Both loops produce a belt speed and both are bounded by the belt. Each is the shared incremental, or velocity, form algorithm, with its derivative acting on the measurement rather than the error, as shown in Equation (6.22).

$$\begin{aligned} e_k &= \mathrm{SP}_k - \mathrm{PV}_k, \qquad p_k = K_c e_k \\ d_k &= -K_c \tau_D \frac{\mathrm{PV}_k^f - \mathrm{PV}_{k-1}^f}{\Delta t_c} \\ s_k &= s_{k-1} + (p_k - p_{k-1}) + (d_k - d_{k-1}) \\ c_k &= u_b + s_k + I_{k-1} \\ u_k &= \mathrm{clamp}(c_k,\ 0,\ v_{\max}) \\ I_k &= I_{k-1} + \left[\frac{K_c}{\tau_I} e_k + \frac{u_k - c_k}{T_t}\right] \Delta t_c \end{aligned} \tag{6.22}$$

where u_b is the bias and PV^f is the filtered measurement. These are Equations (1.12) to (1.15) of Section 1.6.2 in the loop's own symbols, with $s_k = S_k$ and $c_k = v_k$, and anti-windup is back-calculation on I_k alone with the tracking constant $T_t = \tau_I$. Both loops are reverse acting, $a = +1$ in the convention of Section 1.3, because a faster belt raises heart rate and lactate. The increments difference the fully formed terms, so a gain changed in automatic moves the output at once by $(K_{c,\mathrm{new}} - K_{c,\mathrm{old}})\, e_k$. That change is deliberately not bumpless. Restoring the scenario 6 heart-rate loop from 11 to 1.1 % and from 30 to 150 s at 15 min, with the loop 59 bpm below its reference, stops the belt within 5 s and holds it stopped for 243 s. In manual, I_k is solved backward from whatever the operator has set, which makes both directions of transfer bumpless. The heart-rate loop uses $K_c = 1.1\,\%$ of threshold pace per bpm, $\tau_I = 150\,\mathrm{s}$ and $\tau_D = 0$. The lactate loop uses $K_c = 20\,\%$ of threshold pace per mmol/L and $\tau_I = 420\,\mathrm{s}$. Derivative action defaults to zero on both, because Table 6.3 shows nothing fast for a derivative to anticipate and (6.21) supplies noise for it to amplify.

Both gains are carried as a fraction of the runner's own nominal threshold pace and multiplied by it at execution, which is gain scheduling on the one number that says how fast this runner operates. Equation (6.20) gives a factor of 2.1 in plant gain across the four presets against a factor of 2.3 in operating speed, and the schedule is what allows one tuning to serve all of them.

Both loops also trim around a feedforward rather than building their entire output from an integral, as shown in Equation (6.23).

$$u_b = v_{\mathrm{nom}}\, \mathrm{clamp}\left(\frac{t}{t_r/2}, 0, 1\right) \frac{C(0, v_{\mathrm{nom}})}{C(i, v_{\mathrm{nom}})} \tag{6.23}$$

where v_{nom} is the runner's nominal level threshold pace computed from the selected preset and t_r is the reference ramp length. The second factor is the optional incline feedforward, which converts a

level-ground speed into the speed that costs the same energy per second on the grade underfoot. It is a measured feedforward and not a predicted one, because the angle is available the instant it changes while the feedback path behind it is 257 s long.

The selector and its external reset are given in Equation (6.24).

$$\begin{aligned} u &= \min(u_{HR},\ u_L) \\ &\text{track } u_j \text{ to } u+\delta \text{ whenever } u_j > u+\delta \end{aligned} \tag{6.24}$$

where j indexes the loop that was not selected and $\delta = 0.05\,\mathrm{m/s}$, which is 0.7 % of the belt's range. The band in (6.24) is the part that is easy to get wrong. Tracking the losing loop to exactly u puts its next output exactly on u, where it wins the following comparison by an arbitrarily small margin and discards the winner's integral increment before it reaches the belt. The two loops then exchange the selection every scan and the actuator does not move. During development that fault held the belt at a walking pace for a whole hour with an 86 bpm heart-rate error and both trends looking reasonable. Letting the losing loop sit up to δ above the applied output leaves it free to fall through the selection, which is the entire purpose of a selector, while preventing it from winding to the actuator limit. Volume 4, Chapter 4 shares a ground speed among several demands through the same structure, with an incumbent margin that keeps two nearly equal demands from trading the drive.

🕮 **Pause.** Try Exercise 6.1 before reading on.

A lost heart-rate signal changes the selection rather than the loops. The monitor holds its last reading, the heart-rate request leaves the comparison, and a cap at 85 % of the nominal threshold pace takes its place, so that $u = \min(u_L,\ 0.85\,v_{\mathrm{nom}})$ until the signal returns. A frozen reading is not a measurement, and a loop that kept acting on it would hold the runner to a heart rate that is no longer there. The cap is a speed the runner can hold by construction.

6.5.3 Receding-Horizon Alternative

The predictive controller is the application's own nonlinear search and not the linear option of Section 1.6.3. It replans every 5 s over a horizon of 300 s, integrating a reduced copy of the selected preset forward on a 5 s step from its own estimate of the runner. The reduction drops respiration, which constrains nothing, carries blood glucose without the gut store or the uptake sink, and replaces the 0.6 s motor lag with a rate limit, because 0.6 s of motor against a 5 s prediction step is not worth integrating. The decision variables are four blocked speeds with block edges at 0, 10, 30 and 60 % of the horizon. The objective is Equation (6.25).

$$\begin{aligned} J = s_N + \Phi(\mathbf{z}_N) + \tfrac{1}{2}\frac{W'_N}{mC_r}\,\mathrm{clamp}\!\left(\frac{T-t_N}{300\,\mathrm{s}},0,1\right) \\ - 20\lambda\sum_k |u_k - u_{k-1}| - 12\,\Omega \end{aligned} \tag{6.25}$$

where s_N is the distance predicted inside the horizon, $\mathbf{z}_N$ is the predicted end state, λ is a move penalty and Ω accumulates weighted constraint violation. The last two of the leading terms are the terminal value. The first of them, Φ, is the distance still available after the horizon, obtained by rolling the reduced plant forward from $\mathbf{z}_N$ under the base policy of holding sustainable pace on the grade underfoot, in at most sixteen steps of nominally 150 s, as shown in Equation (6.26). This Φ is not the capacity ratio of Equation (6.13).

$$\Phi(\mathbf{z}_N) = \sum_{j=1}^{n} v_{\mathrm{sus}}\big(\mathbf{z}_{N+j-1}\big)\,\Delta t_j, \qquad \sum_j \Delta t_j = T - t_N \tag{6.26}$$

The stores drain along that rollout as they would in the plant, which is what distinguishes Φ from the product $(T-t_N)\,v_{\mathrm{sus}}(\mathbf{z}_N)$ that the same idea suggests. Section 6.7.8 measures the difference and it

Table 6.5 The Application's Own Performance Measures. All Are Accumulated at the 0.25 s Integration Step Over the Whole Run and Are Frozen When the Run Ends.

Measure	Definition	What it is for
Distance	$s(T)$	the objective
Average pace	s/t including the start	the unit a runner reads
Compliant	both integrated excesses below 1 and 10	one pass or fail over both ceilings
Lactate excess	$\int_0^T \max(0, L - L_{\text{lim}})\,\mathrm{d}t$	area above a moving ceiling
Above the ceiling	time with $HR > HR_{\text{lim}}$	duration rather than area
Peak core	$\max_t T_c$	the constraint with no actuator
Fuel left	$F\,g(T)$ in equivalent miles	what the hour cost in store
Reserve left	$W'(T)/(mC_r)$ in meters	whether the deadline was used
Min margin	$\min_t x$	how close the run came to ending

is not small. The second terminal term credits the reserve the base policy never spends, at its own length in meters, ramped to zero over the last 300 s so that a reserve still held at the finish line is worth nothing, which is where the finishing effort comes from. That ramp is a fixed 300 s, equal to the default t_w, and does not follow an edited finishing window.

Constraint violation is evaluated against margins inside the true limits, at 0.25 mmol/L inside the lactate ceiling and 2 bpm inside the heart-rate ceiling, to cover the hill the controller cannot see coming. Candidates are compared on feasibility first and on J second, so a plan that violates less always beats a plan that goes further and the application never softens a constraint to buy distance. The search is a bounded coordinate refinement over deltas of 0.60, 0.22 and 0.08 m/s from four warm starts, capped at 260 plan evaluations. Over the flat hour it uses a mean of 37 evaluations per replan with a maximum of 68, and the mean solve time including the rollout is 0.56 ms in a local Chromium measurement with a maximum of 3.5 ms against a replan period of 5 s. When no feasible plan exists, the search still returns the least-violating one and the applied speed is also limited to the nominal sustainable pace, which is survivable by construction, and the application reports the event rather than hiding it.

In both automatic modes the command itself is rate limited to 1.5 times the belt's own acceleration and deceleration limits, which makes every transfer between modes and architectures bumpless by construction, including the transfer from predictive to regulatory control that no optimizer warm start guarantees on its own.

6.6 Performance Measures

The application reports nine measures beside the per-loop integral absolute error, overshoot, settling time, actuator travel and time saturated that the shared framework computes for every declared loop at the logging rate (Section 1.6.4). The set is deliberately a list of competing costs rather than one composite score, and Table 6.5 gives each one with what it is computed from and what it is for.

Four properties of that set are worth stating because they are design decisions rather than conventions. Compliance is judged on integrated excess rather than on any excursion at all, so that a sample or two of numerical overshoot at a constraint the run is deliberately riding does not read as a violation and a genuine excursion cannot hide behind a short duration. The lactate measure is an area and the heart-rate measure is a duration, because the lactate ceiling is a hard constraint whose magnitude matters and the top of the heart-rate band is a target whose duration matters. Reserve left is reported without a pass band until the run ends, because a full reserve is correct at 20 min and is a failure at 60 min. The distance and the costs are shown together and are never combined, so that a faster run and a compliant run can be compared on the record rather than through a weighting nobody agreed to.

The integral absolute error the framework reports for the heart-rate loop accrues for the whole 3600 s and includes the approach from rest, so it is quoted here only as a comparison between runs of identical length and never as a bare number. Actuator travel, here called command travel, is the total variation $\sum_k |u_k - u_{k-1}|$ of the speed command in m/s, which the application records with the run. The framework's own travel figure differences the logged belt speed instead and reads slightly lower. Command travel is the measure that separates a controller buying distance with continuous speed changes from one that is not.

6.7 Results

The sixteen cases the application ships with are listed in Table 6.6, ordered the way a course would take them, with the mechanism each exercises and what completes it. Loading one writes the mode, the architecture, the runner preset, the tuning, the environment and the event schedule together, and states in advance what the run is meant to show. Completion and the certificates work as Section 1.6.4 describes. Reset restarts the run from $t = 0$ and keeps every setting the learner changed, which is how the before-and-after comparisons are made, while reloading the case restores its own settings. The challenge is judged on the case as given, so changing a model parameter or a disturbance voids the attempt.

6.7.1 Manual Operation and the Failure Mode

Scenario 2 makes the plant's character clear before any controller is involved. The belt steps to 9.36 km/h at 60 s, which is 0.67 km/h below the recreational male preset's threshold pace of 10.03 km/h, and everything settles with lactate at 2.98 mmol/L at 15 min. The belt then steps to 11.16 km/h, which is above critical power, and nothing settles at all. Lactate reaches 4.62 mmol/L three minutes later and 6.10 mmol/L at 24 min, the reserve drains at the excess power of (6.7), and the runner reaches the back edge of the belt at 24:28 with 3.93 km covered. The same experiment at four constant speeds gives failures at 52 s at 16.2 km/h, 261 s at 12.6 km/h and 577 s at 11.2 km/h, the last with 1.78 km covered, and a completed hour of 9.39 km at 9.4 km/h. A speed 11 % above threshold is not an offset that a controller trims out. It is a countdown whose length is the reserve divided by the excess.

6.7.2 Gradient Disturbances and Measured Feedforward

Scenario 4 steps the deck to 5° at 10 min with the controller settled. The belt speed does not change at that instant and the energy every meter costs rises 56 %, so the disturbance is multiplicative on the plant gain and on the operating point at once. Table 6.7 gives the measured result with and without the incline feedforward of (6.23).

> **TRY IT IN THE APP 4 · A Hill Arrives**
>
> Load scenario 4 and press Run with Incline feedforward off. At 10 min the deck tilts to 5°, and lactate peaks at 4.52 mmol/L, past its 4.0 mmol/L limit. Press Reset, open Advanced, set Incline feedforward to On and press Run again. The peak is now 3.50 mmol/L and the lactate excess is zero.

The feedforward acts at the instant the angle changes and the feedback path behind it is 257 s long, so the single hill of scenario 4 costs 29.2 mmol s/L of integrated excess with feedback alone and nothing at all with the feedforward on, while the peak falls from 4.52 to 3.50 mmol/L. Repeating that disturbance twelve times over an hour in scenario 10 multiplies the difference, at 200.3 mmol s/L of excess and a peak of 5.32 mmol/L with feedback alone against zero excess and a peak of 3.86 mmol/L with the feedforward. The shipped noise seed sits at the top of the range for feedback alone. Five other instrument-noise seeds on the same course give 11.7 to 27.5 mmol s/L for the single hill and 175 to 196 mmol s/L for the hour, and the feedforward holds the excess at zero on every one. The course seed's twelve draws average +0.99° and the steepest, +4.42°, arrives at 40 min. Because (6.1)

Table 6.6 Scenarios and Completion Requirements

#	Scenario	What it exercises	Required to complete
1	Standing on a stopped belt *experiment*	the single equilibrium, and basal fuel turnover	Hold the belt at zero in manual for 5 min with every state at rest, then set it to 0.8–1.0 m/s (about 2 mph) and hold it 10 min while the lags answer in order.
2	A speed step by hand *demonstration*	four lags in series, and time to exhaustion	Let both scripted manual steps land (9.4 km/h at 1 min, 11.2 km/h at 15 min) and run on until the runner goes off the back of the belt.
3	Three degrees, held *experiment*	degrees against percent against energy	Under the selector with incline feedforward off, hold the belt within 0.12 m/s of 2.18 m/s for 5 min after the first 5 min, then Reset, switch the feedforward on before Run and settle the same way again.
4	A hill arrives *experiment*	a multiplicative disturbance, and measured feedforward	With feedforward off, see lactate pass 4.0 mmol/L after the hill at 10 min, then Reset, switch the feedforward on before the hill and reach 25 min with peak lactate below 4.0 mmol/L.
5	Downhill, and the distance given away *experiment*	the asymmetry a lactate loop is worst at	Run the selector to 40 min with feedforward off, then Reset, select the predictive controller within the first minute and run the same course to 40 min.
6	Tuning the heart-rate loop into a hunt *experiment*	a hunt whose cost is not on the trend being watched	Run 15 min with the loaded tuning until command travel exceeds 12 m/s, then Reset, restore $K_{c,HR}$ to 0.9–1.3 % and $\tau_{I,HR}$ to 120–180 s before Run, and run 15 min with travel below 7.5 m/s.
7	Two runners, two binding constraints *experiment*	which constraint binds, and why it is not tuning	With the elite preset, see the heart-rate loop selected for 1 min after 10 min, then Reset, switch to the recreational preset and see the lactate loop selected for 5 min after 10 min.
8	Starting the hour under-fueled *experiment*	a constraint that is invisible until it binds	Finish the hour under the selector, then Reset, select the predictive controller within the first minute and finish the hour again.
9	A warm room and a dead fan *experiment*	a constraint with no actuator behind it	Finish the hour as loaded with peak core temperature above 39.5 °C, then Reset, restore the airflow to 0.55–0.65 and finish with the peak below 39.0 °C.
10	One hour over a rolling course *experiment*	a disturbance that is measured and not predictable	Finish the hour under the selector with feedforward off, then Reset, switch the feedforward on within the first minute and finish the same course again.
11	Overground, with a pursuer *demonstration*	a high selector over the low selector	Run under the selector until the pursuer closes the gap, then Reset, select the predictive controller within the first minute and reach 60:00 still ahead.
12	Intervals and incomplete recovery *experiment*	three stores recovering at three rates	Let the five scripted manual speed changes land, then hold the 7.9 km/h recovery for 5 min.
13	Under-fueled with a timed snack *experiment*	delayed appearance of glucose	Let the 40 g feed land at 30 min, see blood glucose rise above 5.0 mmol/L and finish the hour in automatic.
14	Belt beyond the runner's capacity *demonstration*	failure as a drift in position	Let the belt step to 15.1 km/h at 1 min in manual and run until the runner goes off the back.
15	Heart-rate monitor dropout *experiment*	a fallback cap and the return	See the fallback cap selected while the signal is lost (6 to 11 min), then run to at least 15 min with a loop selected again and the belt within 0.1 m/s of 2.6 m/s.
C	One hour, flat, for distance *challenge*	the whole economic problem under a deadline	Cover more than 9.3 km in the hour, compliant on both integrated excesses of Table 6.5, and still on the belt at 60:00, with the case's model parameters and disturbances as given.[a]

[a] With the shipped targets (86 % and 2.6 mmol/L) the hour is 8.78 km and compliant, 0.52 km short. With the suggested targets (90 % and 3.7 mmol/L) it is 9.58 km and compliant, a margin of 0.28 km, and the same when the targets are carried through Reset.

is convex the mean energy cost is 4.61 J/(kg m) against 4.15 J/(kg m) on the level, which is 11 % more for every meter and is why the same runner covers 9.05 km here against 9.58 km on the flat hour.

Table 6.7 Gradient Disturbances Under the Selector, With and Without the Measured Incline Feedforward. Scenario 4 Is Read at 25 min and Scenario 5 at 40 min, Settings Changed Before Run.

	Feedback only	With feedforward	Predictive
Scenario 4, step to 5°			
Peak lactate (mmol/L)	4.52	3.50	—
Lactate excess (mmol s/L)	29.2	0.0	—
Scenario 5, up then down			
Distance at 40 min (km)	6.04	6.25	6.50
Scenario 10, rolling course			
Distance (km)	8.80	9.05	9.38
Lactate excess (mmol s/L)	200.3	0.0	0.0
Peak lactate (mmol/L)	5.32	3.86	5.90

Scenario 5 is the same argument in the other direction and is the one runners give away. Coming onto a −2° section the sustainable speed rises from 7.30 to 12.12 km/h at the instant the deck tilts, and lactate takes four minutes to admit it. The selector reaches only 9.7 km/h over the following two and a half minutes. With the feedforward the belt is at 12.9 km/h within 30 s. The predictive controller's margin over the feedforward in Table 6.7 is 0.25 km at 40 min, which is smaller than the usual expectation, because this disturbance is measured at the moment it happens and prediction adds only the scheduling of the reserve around it.

> ✦ **DID YOU KNOW? The Longest Rolling Hour Peaks Highest**
>
> From the simulator: on the rolling course of scenario 10 the predictive controller covers the most distance, 9.38 km, reaches the highest peak lactate of the three columns, 5.90 mmol/L, and records no lactate excess. Feedback alone peaks lower, at 5.32 mmol/L, and records 200.3 mmol s/L of excess (Table 6.7).

6.7.3 Which Constraint Binds

Scenario 7 runs the elite male preset with the same controller, the same targets and the same room as the recreational preset. Both presets reach 4 mmol/L at an intensity the model resolves to a heart rate, and the two answers fall on opposite sides of the heart-rate target. The recreational preset reaches it at 160 bpm, which is 86.5 % of a 185 bpm maximum and below the 90 % target, so the lactate loop is the lower request and is selected for 96 % of everything after 10 min. The elite preset reaches it at 169 bpm, which is 91.2 % and above the target, so the heart-rate loop is selected for 81 % of the same window and the elite finishes the hour at 3.11 mmol/L with lactate it was never allowed to use. Nothing about the tuning differs between the two runs. Which constraint binds is a property of the plant, and an application that only ever showed one of the two would teach the wrong lesson about selectors.

🕮 **Pause.** Try Exercise 6.2 before reading on.

6.7.4 Detuning, and Where the Cost Lands

Scenario 6 sets the heart-rate loop ten times too hard at $K_c = 11$ % of threshold pace per bpm with its integral five times too fast at $\tau_I = 30$ s, against the 80 s of series physiological lag that Table 6.3 implies plus the 6 s measurement filter. The belt hunts over 4.7 to 8.4 km/h after the first five minutes, and the strongest spectral component of the belt-speed record between 15 and 60 min is at 338 s. Read at 40 min, the detuned loop covers 4.52 km with 43.3 m/s of command travel against 6.34 km and 9.2 m/s for the same loop returned to $K_c = 1.1$ % and $\tau_I = 150$ s before Run with the scenario's 60 s reference ramp kept, a loss of 1.82 km and a factor of 4.7 in command travel. Restoring the 300 s

ramp as well gives 6.26 km and 8.4 m/s. Five other noise seeds give 3.23 to 4.36 km detuned and 6.30 to 6.33 km restored, so the shipped seed shows the smallest loss of the six.

The instructive part is where that cost does not appear. Peak lactate over the detuned run is 1.55 mmol/L against 3.55 mmol/L for the tuned one, so the constraint everybody is watching reports the hunting controller as the safer of the two. The average of a swing that spends half its time slow is a slow average, and on this plant the penalty for a badly tuned loop is collected in the objective and in the actuator rather than in the controlled variable.

6.7.5 Constraints With and Without an Actuator

Scenario 9 places the same runner in a room at 28 °C and 60 % relative humidity with the airflow at 0.15. The runner produces the same heat and the evaporative ceiling $E_{\max}$ of (6.13) falls below it, so (6.12) has no steady state and the core temperature climbs until the pace falls far enough to balance it. The measured peak is 40.26 °C against a stop threshold of 40.5 °C, the fatigue term reaches 29 % of its ceiling, the belt finishes at 8.62 km/h and the hour is 9.21 km. Restoring the airflow to 0.60 without changing the room gives a peak of 38.76 °C and 9.53 km, so 1.5 °C of core temperature and 320 m of distance were bought by a fan and not by a controller. The predictive controller in the same room reaches 40.28 °C, 0.22 °C short of a stop, and covers 9.52 km with 5.7 mmol s/L of lactate excess. No tuning constant creates evaporative capacity, and the only control action available is to produce less heat, which means less distance.

6.7.6 Intervals, Feeding, Capacity and a Lost Signal

Scenario 12 runs two hard two-minute intervals at 11.2 km/h, separated by four easy minutes at 7.9 km/h, in manual after four minutes at 9.7 km/h. The first interval takes heart rate from 155 to 168 bpm, lactate from 2.5 to 3.9 mmol/L and the reserve from 190 to 153 m. The four easy minutes bring heart rate back to 138 bpm and lactate to 2.4 mmol/L but return only 13 m of reserve, so the second interval starts from 166 m and ends at 129 m, and ten easy minutes after it restore the reserve only to 170 m. The three quantities recover at three rates, and the one a runner feels soonest, heart rate, is the fastest of them. The reserve recovers slowly because it refills only below critical power and in proportion to the deficit in (6.7).

Scenario 13 repeats the under-fueled hour of scenario 8 with 40 g of carbohydrate eaten at thirty minutes. Blood glucose is 3.65 mmol/L when the carbohydrate goes in and 5.88 mmol/L five minutes later. Muscle glycogen at thirty-five minutes is 49.9 g against 50.8 g unfed, lower because the fed runner is running faster, and the belt holds 9.8 km/h to forty minutes. The hour is 9.43 km and compliant, against 8.90 km for the unfed hour, which also spends 23.5 mmol s/L above the lactate limit. The snack postpones the fall in sustainable pace and refills none of the muscle store.

Scenario 14 sets the belt to 15.1 km/h after a minute of standing, half again the runner's threshold pace. The belt reaches that speed about fifteen seconds after the step, and the runner keeps up while the reserve drains from 190 to 90 m by 141 s, when the capability v_{cap} of (6.18) falls below the belt speed. The runner then drifts the 0.75 m from the middle of the belt to its back edge and is off at 154.0 s, with heart rate at 141 bpm, still well behind the effort, and lactate at 6.8 mmol/L. Failure is a trajectory in position, and the reserve is the only displayed quantity that anticipates it.

In scenario 15 the chest strap stops reporting six minutes into an hour under the selector and resumes five minutes later. Meanwhile the cap holds the belt at 8.53 km/h against 9.44 km/h before the loss, and the true heart rate falls from 153 to 146 bpm while the monitor still shows 153 bpm. When the signal returns the lactate loop is selected, the belt is back at 9.36 km/h four minutes later, and the hour is 9.44 km and compliant.

6.7.7 The Full Hour, Regulator Against Optimizer

Table 6.8 reports the scorecard for the flat hour under five configurations of the same plant, read at 3600 s, and Figure 6.3 draws three of its rows.

Table 6.8 Measured Scorecard for the Flat Hour, Recreational Male Preset, Read at $t = 3600$ s, Settings Changed Before Run. The Heart-Rate IAE Accrues Over the Whole Run and Is Comparable Only Between These Columns.

	Loaded PID	Selector PID	Selector t_r=120 s	Selector no ff	Predictive MPC
HR target (%)	86	90	90	90	90
L target (mmol/L)	2.6	3.7	3.7	3.7	3.7
Distance (km)	8.784	9.584	9.677	8.676	9.947
Compliant	yes	yes	yes	yes	yes
Lactate excess (mmol s/L)	0.0	0.0	0.0	0.0	0.0
Peak lactate (mmol/L)	2.63	3.75	3.75	3.74	5.87
Peak *HR* (bpm)	155	164	164	164	182
Peak core (°C)	38.44	38.57	38.57	38.57	38.66
Reserve left (m)	190	189	185	190	36
Fuel left (miles)	12.1	11.1	11.0	12.0	10.6
Command travel (m/s)	6.3	6.5	7.0	5.2	32.9
HR IAE (bpm s)	24524	19073	23342	59603	21824
In the band (%)	36	89	89	72	96

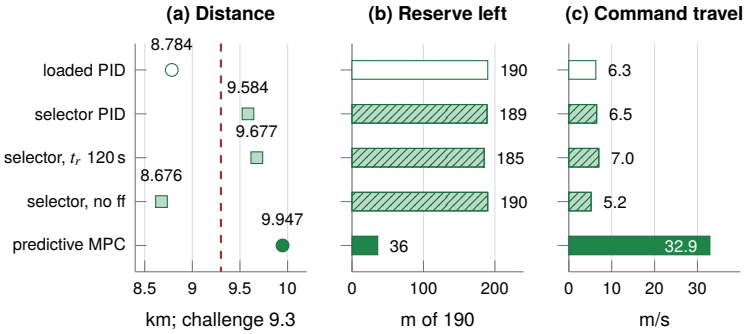


Figure 6.3 The flat hour of Table 6.8 for its five configurations, read at 3600 s: distance as dots against the challenge's 9.3 km (dashed), and the reserve left and the command travel as bars. Open marks and white bars are the loaded configuration, squares and hatched bars the three selector variants, and the solid ones the predictive controller. Notice that the predictive controller goes furthest by spending most of its reserve and by moving the belt far more than any selector.

Reading Table 6.8 across gives the expected ordering. The conservative configuration the scenario loads is comfortably inside every limit and 0.80 km short of the regulator's answer. The regulator at the suggested targets covers 9.584 km against a threshold pace that would cover 10.03 km in the hour if the runner were already at that speed at $t = 0$, and the acceleration from standing and the reference ramp account for the difference. Disabling the nominal pace feedforward of (6.23) costs 0.909 km and multiplies the heart-rate IAE by 3.1, from 19073 to 59 603 bpm s, because a proportional-integral loop starting from a stopped belt has to integrate its way up to the operating speed and the loop that gets there slowest wins the low selector. With the lactate loop's 420 s integral that is a quarter of an hour spent discovering a pace the runner's own profile already states.

The predictive controller covers 9.947 km, which is 363 m and 3.8 % more than the regulator, with every constraint respected. Five other instrument-noise seeds give gaps of 340 to 384 m. Settings made before Run and carried through Reset give identical trajectories and 32.9 m/s of command travel. Where that margin is earned is more interesting than its size, and Table 6.9 decomposes it.

Table 6.9 contradicts the intuitive account. More than half of the advantage, 210 m of 363 m, exists by 3 min, because the predictive controller reaches 14.1 km/h within 15 s and is still at 11.1 km/h at one minute while the regulator walks a 300 s heart-rate reference ramp up from rest. Between

Table 6.9 Where the Predictive Controller's 363 m Advantage Over the Selector Is Accumulated, Flat Hour, Identical Runner and Identical Targets.

Elapsed	Selector (m)	Predictive (m)	Gap (m)
1 min	42	204	162
3 min	331	541	210
10 min	1429	1690	261
30 min	4677	4959	282
55 min	8765	9025	260
60 min	9584	9947	363

Table 6.10 The Same Hours Under Three Terminal Values. *Frozen* Is the Product $(T - t_N)\, v_{\mathrm{sus}}(\mathbf{z}_N)$, *Rollout* Is Φ of (6.26), *None* Deletes Both Terminal Terms. Every Run Completes the Hour, Compliant Unless Starred (Counterfactual Controller Copy, Shipped Noise Seed, Settings Changed Before Run).

	Frozen			Rollout			None		
Hour	km	W'	tr	km	W'	tr	km	W'	tr
Flat, 300 s	9.95	46	30	9.95	36	33	9.94	38	33
Flat, 60 s	9.95	59	18	9.95	59	21	9.92*	53	25
Rolling course	9.38	38	52	9.38	42	45	9.37	31	62
Warm room	9.52*	14	68	9.52*	11	62	9.52*	21	68
Under-fueled	8.35	60	37	9.09	51	13	9.10	3	25

W' is the reserve left in meters of 190, tr is the command travel in m/s.

3 min and 55 min the predictive controller runs on average 0.06 km/h faster, at most 2.12 km/h apart, and the gap grows by only 49 m. The last five minutes add 103 m, which is the finishing allowance of (6.19) being spent. The predictive controller reaches 11.89 km/h and 5.82 mmol/L at 59.9 min while the regulator holds 9.80 km/h and 3.73 mmol/L, and the reserves end at 36 m and 189 m of 190 m respectively.

Two consequences follow. The startup component is a tuning choice rather than an architectural advantage, and shortening the reference ramp from 300 to 120 s recovers 93 m of it with no constraint violation, as the third column of Table 6.8 shows. What remains is architectural. A regulator holding a setpoint has no representation of a deadline, so it cannot decide when to stop holding, and the 153 m of reserve it carries across the finish line beyond the predictive controller's is the price of that.

6.7.8 What the Terminal Value Is Actually Worth

The usual argument for a terminal value is that without one a five-minute optimizer spends stores whose cost lands outside its horizon. That argument says nothing about whether a particular terminal value is correctly built, and it is possible to settle the question by running the same hours three ways. Table 6.10 does so, with Φ as written in (6.26), with Φ replaced by the frozen product $(T - t_N)\, v_{\mathrm{sus}}(\mathbf{z}_N)$, and with the terminal terms deleted altogether.

On four of the five hours the three forms give distances within 33 m of one another, while the distance, the reserve left and the command travel move by amounts comparable to the spread between instrument-noise seeds, and the reason is in Table 6.4. The terminal value is a function of the predicted end state through the sustainable speed, and on a level hour with full stores that speed moves 0.5 %, so the term takes very nearly the same value for every candidate plan and cannot rank them. A terminal cost that does not vary between candidates is not doing anything, whatever its magnitude.

The fifth hour is where the term is live, and the frozen form fails on it. The mechanism is a kink. The fuel factor Ψ_G in (6.5) is flat until the store falls through 15 % and slopes afterward, so $\partial v_{\mathrm{sus}} / \partial (G_m + G_l)$ is exactly zero above 54.1 g of store and 0.0103 m/s per gram below it. Multiplying

a kink by $T - t_N$, which reaches 3300 s, turns it into a cliff in J worth about 34 m per gram of predicted end-state glycogen. Above the kink a plan is charged nothing at all for burning fuel and below it the charge is abrupt, so successive replans settle on opposite sides. Between 21 and 40 min of the under-fueled hour the frozen form moves the belt between 6.2 and 10.8 km/h from one minute to the next, with 9.8, 6.4, 8.0 and 7.1 km/h at 20, 25, 30 and 35 min, which is not a pacing strategy but a controller alternating between two plans.

🕮 **Pause.** Try Exercise 6.3 before reading on.

Integrating through the kink instead of sampling one side of it removes it. The rollout in (6.26) is a distance rather than a speed times a duration, it is smooth in every state it depends on because the stores drain along it, and on the same four marks it gives 9.9, 9.8, 8.7 and 8.4 km/h. The hour recovers 746 m, from 8.35 to 9.09 km, and the command travel falls from 37 to 13 m/s, both runs compliant. The complete solve with rollout averages 0.56 ms in the local Chromium measurement over 719 advancing replans, against a 5 s replan period. These wall times depend on the host and browser.

Two conclusions follow and the second is the more useful. A terminal cost is a modeling choice with a shape, and a shape that is flat on one side of a threshold and steep on the other will be sampled rather than integrated by a receding-horizon search, which is a failure that shows up as actuator motion long before it shows up in the objective. The deeper point is in the third group of columns. With the corrected terminal value the controller is still within 31 m of having no terminal value at all on every hour tested, and on the under-fueled hour the version without one is the 6 m longer. The term is not what makes this predictive controller beat the regulator. What does that is the constraint handling and the finishing allowance of (6.19), and the honest description of the terminal value here is that its job is to avoid doing harm.

6.8 Technology in Practice

The heart-rate monitor, the lactate estimate and the belt motor of Figure 6.1 stand for the equipment of an exercise laboratory. Figure 6.5 draws the selector loop of Section 6.5 with that equipment in place of the schematic's symbols.

What Measures It

Heart rate comes from a chest strap that records the heart's electrical signal and reports a rate over a wireless link. In one laboratory heart-rate loop the rate arrived once a second and the controller used the average of five readings every 5 s [8]. The chapter's measurement in Equation (6.21), with 0.8 bpm of colored noise and a 6 s filter, stands for that strap and its averaging, and the held reading of scenario 15 stands for a strap that stops reporting. Lactate has no such instrument: a threshold test draws blood at the fingertip or the earlobe [23]. The lactate signal the chapter's loop reads every second is therefore an estimate, and the chapter labels it so.

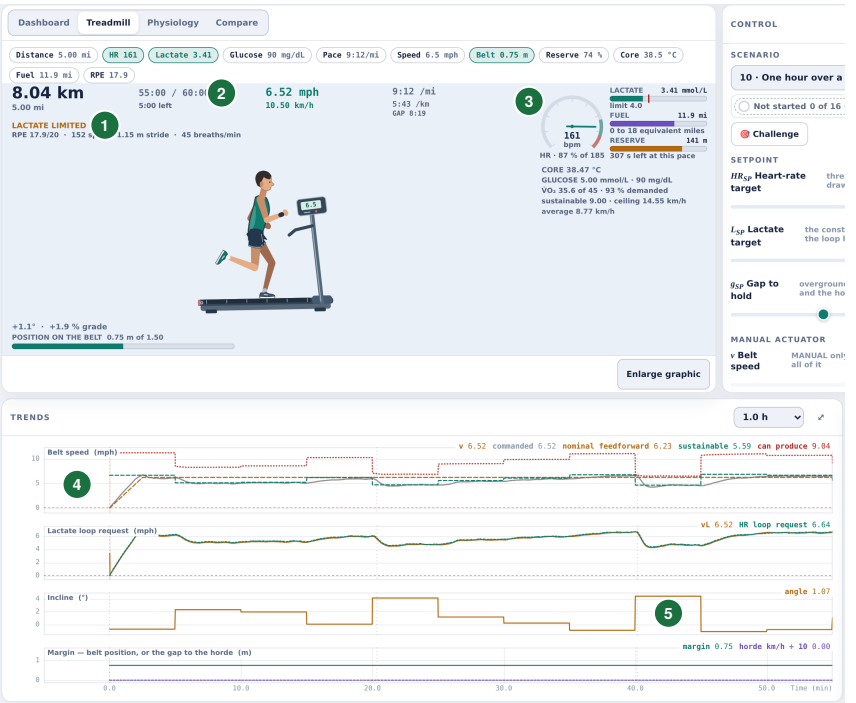


Figure 6.4 Scenario 10 at $t = 55$ min, the rolling course under the selector on feedback alone, cropped to the treadmill view and the trends. The deck is tilted to +1.07° and the runner has covered 8.04 km with lactate at 3.41 mmol/L. The incline strip records the five-minute redraws over the whole run. The +4.42° block that lands at 40 min takes the sustainable pace down, drives blood lactate above its 4.0 mmol/L constraint within 9 s and pulls the lactate loop's request and the belt down with it, the belt to 6.83 km/h at 41:13, and that block adds 169 of the 200 mmol s/L of lactate excess the run records. The belt-speed strip carries the commanded speed, the nominal feedforward of (6.23), the sustainable pace and the capability v_{cap} alongside the belt itself, and the scene names the selected constraint in words above the machine. Callouts: (1) the selected constraint named in words, (2) distance and elapsed time, (3) the heart-rate gauge with the lactate, fuel and reserve bars, (4) the belt-speed strip, (5) the +4.42° block on the incline strip.

TECHNOLOGY **Sensor: The Chest-Strap Heart-Rate Monitor**

A chest strap holds electrodes against the skin and records the heart's electrical signal, from which it reports the heart rate and the intervals between beats over a wireless link. In a 2025 systematic review the most studied straps correlated strongly with electrocardiography at rest and in light-to-moderate activity, and chest straps were noted for better accuracy than the more common wrist-worn devices. The limitations reported were motion artifacts, poor strap placement, sweating and a degrading contact between skin and electrode, and none of the devices reviewed was approved for clinical use [24]. A clinical cardiac monitor that measures heart rate from an electrocardiograph's signal, and may alarm outside preset limits, is a Class II device in United States regulation [25].

What Moves It

The final element is the belt drive, which brings the belt to the speed a computer commands [26]. The chapter models that inner loop in Equation (6.18) as a first-order motor with a 0.6 s lag and rate limits of 0.35 and 0.50 m/s^2, and its automatic modes limit the command itself to 1.5 times the belt's rate limits. A runner who cannot hold the belt speed drifts back along the deck, which is the position state of the chapter and the failure that scenario 14 shows.

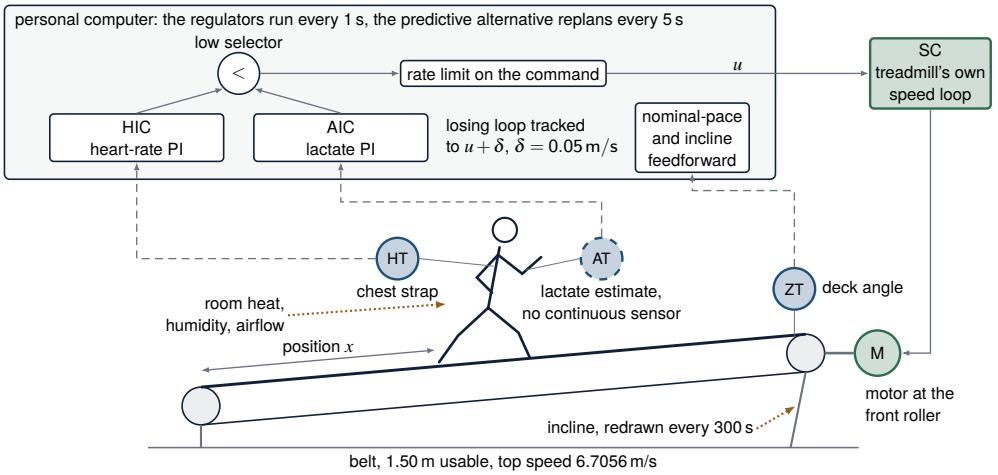


Figure 6.5 The selector loop of Section 6.5 with laboratory equipment in place of the schematic's symbols. A chest strap (HT) reports heart rate and the deck angle (ZT) is measured, while the lactate signal (AT, dashed) is an estimate, because the chapter implies no continuous blood-lactate sensor. Two PI loops (HIC, AIC) meet in a low selector on a personal computer, and its rate-limited command u goes to the treadmill's own speed loop (SC) and motor (M). Notice that one final element serves both loops, and that the runner's position x on the belt, the state that ends a run, is read by no instrument in this loop.

TECHNOLOGY **Final Element: The Treadmill Belt Drive**

A treadmill in a control study takes its speed from a computer, and its own drive holds the belt there. In an adaptive running platform the computed speed went to the platform's own microcontroller, which set it directly, many times a second, and the platform accelerated and stopped at a constant 1 m/s^2 [26]. A predictive heart-rate controller kept both the speed and the acceleration of its treadmill within a safe range for vulnerable exercisers [10]. United States regulation lists a powered treadmill intended for medical purposes among powered exercise equipment, a Class I device under general controls [27].

Where the Controller Runs

In the laboratory the heart-rate controller ran as real-time software on a personal computer beside the treadmill and sent its speed commands over a serial link every 5 s [8]. The chapter's regulators run once per simulated second and its optimizer replans every 5 s. The two proportional-integral loops, the low selector with its external-reset band, the two feedforward terms and the fallback cap of Section 6.5 form one supervisory layer above the drive's own speed loop.

IN PRACTICE **Running a Heart-Rate Loop in the Laboratory**

A laboratory test of a heart-rate loop is a protocol as much as a controller. In one study each test began with 10 min of warm-up under the controller, then 10 min of rest, then a 35 min measurement, and the target heart rate was set from the age-predicted maximum. The tracking error was scored over the 30 minutes from 295 to 2095 s. Two participants had measurement artifacts in their final five minutes, and their scoring stopped five minutes early. Beside the tracking error, the study scored the controller's effort as the average power of the changes in the speed command [8], the concern that command travel answers in Table 6.5.

What Is Modernizing It

Three developments bear on this loop. The first is predictive control on models identified from the person: a predictive heart-rate controller ran on models estimated with a short protocol and kept the

treadmill's speed and acceleration within a safe range [10]. The second is a treadmill that follows the runner. An adaptive running platform reads the person's position on the belt from trackers or a camera and sets the belt speed from it, and a position error up to 5 cm, 30 Hz updates and 100 ms of total latency were judged acceptable [26]. That loop is closed on the position x that the chapter uses only as its failure state. The third is lactate read from the skin. A wearable sensor on the forehead monitored sweat lactate continuously during arm crank exercise in 27 healthy young adults, and the threshold read from it correlated strongly with the first ventilatory threshold [28]. A sensor of that kind that held for every runner would replace the chapter's lactate estimate with a measurement.

6.9 Limitations

The model is not a physiological instrument and cannot be used as one. The coefficients of (6.5), (6.10), (6.11) and (6.13) are teaching calibrations chosen so that four presets behave plausibly, and they are not fitted to any individual or population. The relations they are built on are published, and the parameterizations that carry them here are not. Nothing in the application is validated for prescribing training, assessing fitness or guiding anybody's running.

Critical power and the 4 mmol/L intensity are equal in every preset, which removes a distinction the literature maintains, so the difference between the two constructs can be studied only by editing one of them. Lactate has no transport delay beyond its clearance time constant, so the application understates the measurement problem rather than overstating it. The thermal model has one node and no dehydration, so a hot hour degrades more gently here than in practice. Feeding is carbohydrate alone, and nothing refills muscle glycogen during the run.

The predictive controller finds an approximate local solution with a bounded coordinate search over four blocked moves and a rolled-out terminal value. It is not a global optimum and no claim of optimality is made, and the comparison in Table 6.9 is a comparison between two implemented controllers rather than between an architecture and a bound. The teaching switch that hands the predictive controller the hidden incline schedule is an advantage no controller has, and any comparison that uses it is labeled on screen and in the exported record. The determinism is unrealistic in its own right, since a person running the same hour twice does not produce the same hour twice. Section 6.3.9 names the point at which each modeling assumption stops holding.

> **OPEN CHALLENGE** **A Lactate Reading Without a Blood Sample**
>
> The chapter labels its lactate signal an estimate and implies no continuous blood-lactate sensor. In practice most tests of the anaerobic threshold still draw blood at the fingertip or earlobe, which is costly and time-consuming. A 2025 study of 55 subjects on a cycle ergometer read thresholds from sweat electrolytes instead. They agreed with the blood lactate threshold in the high and middle fitness groups but may not predict it in the low group, and the authors called for further research [23]. Progress would be a reading taken from sweat or skin that holds for every fitness level, which is the gap that study left open.

6.10 Discussion

The structure of this chapter is not particular to running. Any plant with a stored resource, a fixed deadline and an objective that grows with the thing the constraints forbid has the same shape, and the list includes a battery-electric vehicle on a long climb, a peaking generator working against a fuel contract, a batch reactor with a fixed cycle time and a spacecraft with a propellant budget. Volume 9, Chapter 12 spends a train's spare seconds between stations against its energy meter on the same kind of deadline. In each of them the value of the store at the deadline is nothing like its value halfway through, which is the property that a regulator cannot represent and an economic formulation can [2]. Pacing models in sport reached the same conclusion from the other direction, by writing the event as an energy flow with a finite anaerobic store and solving for the schedule [3]. Putting both on one actuator lets the 363 m in Table 6.9 be split into the part that is a tuning choice and the part that is architectural.

Tables 6.4 and 6.10 carry one caution between them. What makes scheduling a store valuable is a constraint that moves, and on the level hour the sustainable speed moves 0.5 %, which is why the flat hour moves by less than 12 m whether the terminal value is computed carefully, computed carelessly or deleted. Only in the warm room and the under-fueled hour, where the constraint falls 11.4 % and 20.2 %, does the forecast of a store have any work to do, and the under-fueled hour is the only one on which the careless form does damage. Establishing which regime a plant is in, before building the controller that assumes one of them, decides whether the terminal cost is worth writing and whether its shape matters.

Two further methods fit the same equations without changing them. The fourteen-state system with explicit constraints transfers directly to a nonlinear programming formulation in APMonitor [29] or the Gekko Optimization Suite [30], where the blocked search of (6.25) would be replaced by a gradient-based solve over a full discretization and the rollout of (6.26) by a cost-to-go the solver differentiates. A moving horizon estimator built on the same equations could recover ϕ, the one quantity the application keeps off every display, from the heart rate and the pace still available, which turns the heat-against-fuel diagnosis of Table 6.4 into a soft sensor instead of a reading of two trends. The heart-rate identification literature supplies the reduced models a practical implementation would use in place of the full plant [4, 5, 10].

6.11 Conclusion

This chapter presents one hour of running as a laboratory plant in which a single actuator serves several constrained outputs and the objective is bounded by a deadline. The simulator integrates fourteen states, covering the mechanical belt and the runner’s position on it, oxygen uptake, a finite anaerobic reserve, heart rate, blood lactate, accumulated fatigue, core temperature, respiration, three carbohydrate stores and a gut compartment for carbohydrate eaten during the run, with the energy cost of gradient taken from published measurement and the sustainable power written as a critical-power construct that falls with fatigue and fuel. The four constrained outputs answer a belt step with time constants of 40, 96, 257 and 475 s. Two proportional-integral loops sharing the actuator through a low selector with a 0.05 m/s external-reset band cover 9.584 km in the hour with every constraint respected, while a receding-horizon controller with a 300 s horizon and a terminal value covers 9.947 km and finishes with 36 m of its 190 m reserve against the regulator’s 189 m. Of that 363 m advantage, 210 m exists by three minutes and is recoverable by shortening the regulator’s reference ramp, and 103 m is taken in the last five minutes when the ceilings open. A measured incline feedforward removes the whole of the 200.3 mmol s/L of lactate excess that twelve hills cost feedback alone, and detuning the heart-rate loop costs 1.82 km of the first 40 min and 4.7 times the command travel while lowering the peak lactate, which is the reason the scorecard reports competing costs rather than one number. The terminal value is measured rather than assumed. Written as the end-state sustainable speed held for the remaining time it inherits a kink from the fuel factor, the search alternates across it, and the under-fueled hour loses 746 m and raises its command travel from 13 to 37 m/s. Written as a rollout of the base policy it does not, and it is then within 31 m of having no terminal value at all on every hour tested. The later cases show the reserve recovering far more slowly than heart rate between intervals, a snack at half time worth 535 m to an under-fueled runner without refilling the muscle store, a belt beyond capacity failing as a drift in position 94 s after the step, and a lost heart-rate signal handled by a conservative cap. A fuller model would add a skin compartment and dehydration in the thermal balance, presets that separate critical power from the lactate marker, a gradient-based solve in place of the blocked coordinate search, and estimation of the hidden fatigue state from the measurements the runner already has. A runner on a belt is an accessible plant for learning to hold whichever constraint binds and to work out when to stop holding it.

KEY POINTS

- One belt speed has to respect several limits of the runner's body, and which limit is reached first depends on the runner, not on the tuning.
- A low selector passes the lower of the two requests. Its losing loop is tracked to a band above the applied speed, because tracking it to the applied speed itself deadlocks the selector.
- A measured incline feedforward removes the whole 200.3 mmol s/L of lactate excess that twelve hills cost feedback alone.
- The predictive controller covers 9.947 km against 9.584 km. Of that 363 m, 210 m exists by three minutes and is a tuning choice.

Where else this applies. One actuator shared among limits by a selector returns in Volume 4, Chapter 4 and Volume 4, Chapter 3, and slack spent against a fixed deadline in Volume 9, Chapter 12.

References

[1] K. J. Åström and T. Hägglund, *Advanced PID Control*. Research Triangle Park, NC: ISA – The Instrumentation, Systems, and Automation Society, 2006.

[2] M. Ellis, H. Durand, and P. D. Christofides, "A tutorial review of economic model predictive control methods," *Journal of Process Control*, vol. 24, no. 8, pp. 1156–1178, 2014.

[3] J. J. de Koning, M. F. Bobbert, and C. Foster, "Determination of optimal pacing strategy in track cycling with an energy flow model," *Journal of Science and Medicine in Sport*, vol. 2, no. 3, pp. 266–277, 1999.

[4] S. W. Su, L. Wang, B. G. Celler, A. V. Savkin, and Y. Guo, "Identification and control for heart rate regulation during treadmill exercise," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 7, pp. 1238–1246, 2007.

[5] T. M. Cheng, A. V. Savkin, B. G. Celler, S. W. Su, and L. Wang, "Nonlinear modeling and control of human heart rate response during exercise with various work load intensities," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 11, pp. 2499–2508, 2008.

[6] K. J. Hunt and S. E. Fankhauser, "Heart rate control during treadmill exercise using input-sensitivity shaping for disturbance rejection of very-low-frequency heart rate variability," *Biomedical Signal Processing and Control*, vol. 30, pp. 31–42, 2016.

[7] K. J. Hunt, B. Ajayi, H. Gollee, and L. Jamieson, "Feedback control of oxygen uptake during treadmill exercise," *IEEE Transactions on Control Systems Technology*, vol. 16, no. 4, pp. 624–635, 2008.

[8] H. Wang and K. J. Hunt, "Feedback control of heart rate during treadmill exercise based on a two-phase response model," *PLOS ONE*, vol. 18, no. 10, p. e0292310, 2023.

[9] R. A. Bruce, F. Kusumi, and D. Hosmer, "Maximal oxygen intake and nomographic assessment of functional aerobic impairment in cardiovascular disease," *American Heart Journal*, vol. 85, no. 4, pp. 546–562, 1973.

[10] L. Wang, Y. Yang, and S. Su, "Nonparametric modelling based model predictive control for human heart rate regulation during treadmill exercise," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2021, pp. 148–151.

[11] A. E. Minetti, C. Moia, G. S. Roi, D. Susta, and G. Ferretti, "Energy cost of walking and running at extreme uphill and downhill slopes," *Journal of Applied Physiology*, vol. 93, no. 3, pp. 1039–1046, 2002.

[12] H. Monod and J. Scherrer, "The work capacity of a synergic muscular group," *Ergonomics*, vol. 8, no. 3, pp. 329–338, 1965.

[13] A. M. Jones, A. Vanhatalo, M. Burnley, R. H. Morton, and D. C. Poole, "Critical power: Implications for determination of $\dot{V}O_2$max and exercise tolerance," *Medicine & Science in Sports & Exercise*, vol. 42, no. 10, pp. 1876–1890, 2010.

[14] D. P. Swain and B. C. Leutholtz, "Heart rate reserve is equivalent to $\%\dot{V}O_2$ reserve, not to $\%\dot{V}O_2$max," *Medicine & Science in Sports & Exercise*, vol. 29, no. 3, pp. 410–414, 1997.

[15] H. Tanaka, K. D. Monahan, and D. R. Seals, "Age-predicted maximal heart rate revisited," *Journal of the American College of Cardiology*, vol. 37, no. 1, pp. 153–156, 2001.

[16] G. A. Brooks, "The science and translation of lactate shuttle theory," *Cell Metabolism*, vol. 27, no. 4, pp. 757–785, 2018.

[17] H. Heck, A. Mader, G. Hess, S. Mücke, R. Müller, and W. Hollmann, "Justification of the 4-mmol/l lactate threshold," *International Journal of Sports Medicine*, vol. 6, no. 3, pp. 117–130, 1985.

[18] P. L. Valenzuela, L. B. Alejo, A. Montalvo-Pérez, J. Gil-Cabrera, E. Talavera, A. Lucia, and D. Barranco-Gil, "Relationship between critical power and different lactate threshold markers in recreational cyclists," *Frontiers in Physiology*, vol. 12, p. 676484, 2021.

[19] H. Heck and H. Wackerhage, "The origin of the maximal lactate steady state (MLSS)," *BMC Sports Science, Medicine and Rehabilitation*, vol. 16, p. 36, 2024.

[20] J. González-Alonso, C. Teller, S. L. Andersen, F. B. Jensen, T. Hyldig, and B. Nielsen, "Influence of body temperature on the development of fatigue during prolonged exercise in the heat," *Journal of Applied Physiology*, vol. 86, no. 3, pp. 1032–1039, 1999.

[21] J. A. Romijn, E. F. Coyle, L. S. Sidossis, A. Gastaldelli, J. F. Horowitz, E. Endert, and R. R. Wolfe, "Regulation of endogenous fat and carbohydrate metabolism in relation to exercise intensity and duration," *American Journal of Physiology-Endocrinology and Metabolism*, vol. 265, no. 3, pp. E380–E391, 1993.

[22] E. F. Coyle, A. R. Coggan, M. K. Hemmert, and J. L. Ivy, "Muscle glycogen utilization during prolonged strenuous exercise when fed carbohydrate," *Journal of Applied Physiology*, vol. 61, no. 1, pp. 165–172, 1986.

[23] Z. Luo, M. Cao, L. Yan, and J. Wang, "Feasibility study of a novel wearable sweat sensor for anaerobic threshold determination," *Scientific Reports*, vol. 15, p. 30467, 2025.

[24] A. Machado, D. F. Ferreira, S. Ferreira, N. Almeida-Antunes, P. Carvalho, P. Melo, N. Rocha, and M. A. Rodrigues, "A systematic review of chest-worn sensors in cardiac assessment: Technologies, advantages, and limitations," *Sensors*, vol. 25, no. 19, p. 6049, 2025.

[25] U.S. Food and Drug Administration, "Cardiac monitor (including cardiotachometer and rate alarm)," Code of Federal Regulations, Title 21, Section 870.2300, revised as of April 1, 2024, U.S. Government Publishing Office, govinfo.gov, 2024.

[26] A. Obukhov and A. Volkov, "Development of control algorithms for an adaptive running platform for a musculoskeletal rehabilitation system," *Sensors*, vol. 25, no. 21, p. 6667, 2025.

[27] U.S. Food and Drug Administration, "Powered exercise equipment," Code of Federal Regulations, Title 21, Section 890.5380, revised as of April 1, 2024, U.S. Government Publishing Office, govinfo.gov, 2024.

[28] T. Sawada, H. Okawara, K. Minami, S. Narushima, A. Shiratori, Y. Katsumata, M. Nakamura, T. Nagura, and D. Nakashima, "Validation of aerobic threshold assessment using a sweat lactate sensor during arm crank exercise by healthy adults," *Physiological Reports*, vol. 14, no. 13, p. e71002, 2026.

[29] J. D. Hedengren, R. Asgharzadeh Shishavan, K. M. Powell, and T. F. Edgar, "Nonlinear modeling, estimation and predictive control in APMonitor," *Computers & Chemical Engineering*, vol. 70, pp. 133–148, 2014.

[30] L. D. R. Beal, D. C. Hill, R. A. Martin, and J. D. Hedengren, "GEKKO optimization suite," *Processes*, vol. 6, no. 8, p. 106, 2018.

6.12 Exercises

6.1 ○ Concept check · 2 min

Suppose the band δ of Equation (6.24) is set to zero, so that the loop that lost the comparison is tracked to exactly the applied speed u. What does the treadmill do?

(a) The losing integrator winds to the top of the belt's range, so the hand-over comes late

(b) The belt follows the higher of the two requests, so the binding limit is exceeded

(c) The selection settles on whichever loop has the shorter integral time

(d) The loops trade the selection every scan, and the belt stops moving toward either target

6.2 ○ Concept check · 1 min

Scenario 7 runs the elite male preset with the same tuning, targets and room as the recreational preset. Which loop holds the belt for most of the run after 10 min, and why?

(a) The heart-rate loop, because this runner reaches 4 mmol/L above the 90 % heart-rate target

(b) The lactate loop, because its gain is the larger of the two in percent of threshold pace

(c) The heart-rate loop, because this runner's heart rate lags the belt less than lactate does

(d) The lactate loop, because this runner's anaerobic reserve is larger than the other's

6.3 Concept check · 2 min

With the frozen terminal value $(T - t_N)\, v_{\text{sus}}(\mathbf{z}_N)$, the under-fueled hour moves the belt between 6.2 and 10.8 km/h from one minute to the next. What causes the alternation?

(a) The move penalty λ is too small to damp the plan between replans

(b) The fuel factor has a kink at 15 % of the store, which the remaining time turns into a cliff in J

(c) The lactate estimate's noise changes which plans are feasible at every replan

(d) The horizon cannot see the finishing window until the last five minutes

6.4 Theory and calculation · 6 min

Take the recreational male preset of Table 6.2: $m = 75\,\text{kg}$, $\dot{V}_{O_2,\max} = 45\,\text{mL/(kg min)}$, $C_r = 4.15\,\text{J/(kg m)}$, $f_{CP} = 0.80$, $HR_{\text{rest}} = 60\,\text{bpm}$ and age 35, with $\varepsilon_{O_2} = 20.9\,\text{J/mL}$.

(a) Compute P_{rest} and $P_{\max}$ from Equations (6.2) and (6.3). *Checkpoint:* 91.4 W and 1175.6 W.

(b) Evaluate the static gain of Equation (6.20). *Checkpoint:* 35.9 bpm per m/s.

(c) On the level with full stores and no fatigue, the threshold pace v_{nom} satisfies $mC_r\, v_{\text{nom}} = CP - P_{\text{rest}}$. Show that $v_{\text{nom}}\, \partial HR/\partial v$ equals $f_{CP}\,(HR_{\max} - HR_{\text{rest}})$. Then compute the loop gain $K_c\, v_{\text{nom}}\, \partial HR/\partial v$ of the heart-rate loop at $K_c = 1.1\,\%$ of threshold pace per bpm, for this preset and for the elite male preset ($f_{CP} = 0.87$, same age and resting heart rate). Give two decimals. *Checkpoint:* 1.10 and 1.20.

6.5 Theory and calculation · 6 min

The recreational male preset on the level has $CP = 958.8\,\text{W}$, $P_{\text{rest}} = 91.4\,\text{W}$, a reserve $W'_{\max} = mC_r D'$ with $D' = 190\,\text{m}$, and $\mu = 1.35$. Treat ϕ as zero and the belt as being at speed from the start.

(a) At a constant 11.2 km/h, compute the demand P of Equation (6.2), the excess $P - CP$ and the time to empty a full reserve. *Checkpoint:* 1059.8 W, 101.0 W, 586 s.

(b) The legs fall short of the belt when P_{av} of Equation (6.6) drops below P. Find the fraction $W'/W'_{\max}$ at which that happens at 11.2 km/h, and the time to reach it. Compare with the failure at 577 s that the chapter reports. *Checkpoint:* 0.0118 and 579 s.

(c) Repeat (b) for the 15.1 km/h belt of scenario 14 and give the reserve in meters at which the runner loses the belt. *Checkpoint:* 88 m.

6.6 Theory and calculation · 5 min

The frozen terminal value multiplies the sustainable speed of the predicted end state by the time left, $T - t_N$. Use the recreational male preset on the level with $\phi = 0$, a blood glucose above 4 mmol/L and the starting store $G_{m,0} + G_{l,0} = 361\,\text{g}$.

(a) Below what store, in grams, does Ψ_G of Equation (6.5) begin to fall? *Checkpoint:* 54.1 g.

(b) Below that store, compute $\partial v_{\text{sus}}/\partial(G_m + G_l)$, with $v_{\text{sus}} = (CP - P_{\text{rest}})/(mC_r)$. *Checkpoint:* 0.0103 m/s per gram.

(c) Multiply by the largest $T - t_N$, 3300 s, and explain what a bounded coordinate search does with an objective shaped this way. *Checkpoint:* about 34 m per gram.

6.7 App challenge · 8 min

Load scenario 4, A hill arrives, and keep the seed and every setting it loads (Auto · PID, recreational male, incline feedforward off). The deck tilts to 5° at 10 min. Set the units to SI (km/h and kilometers) and run to 25:00 at speed 20.

(a) Read the Lactate excess tile and the peak lactate, given in the tile's hover note and in the last row of column `metric_peak_lac` of the exported CSV file. Give the excess within 1 mmol s/L and the peak within 0.02 mmol/L. *Checkpoint:* 29 mmol s/L and 4.52 mmol/L.

(b) Reload the scenario, set Incline feedforward to On under Advanced before Run, and repeat the run and the readings. *Checkpoint:* 0 mmol s/L and 3.50 mmol/L.

(c) With $C = 4.15\,\mathrm{J/(kg\,m)}$ on the level and 6.49 J/(kg m) at 5°, compute the speed that Equation (6.23) sets on the hill from the threshold pace of 2.787 m/s. Check it against column `speed_ff_mps` in the row at 20 min of the second file, within 0.01 m/s. *Checkpoint:* 1.78 m/s.

6.8 App challenge · 8 min

Load scenario 6, Tuning the heart-rate loop into a hunt, and keep the seed and every setting it loads ($K_c = 11\,\%$, $\tau_I = 30\,\mathrm{s}$, a 60 s reference ramp, monitor noise 2 bpm). Set the units to SI (km/h and kilometers), run to 40:00 at speed 60 and export the CSV file.

(a) Read the Distance tile, the peak lactate (hover note of the Lactate excess tile) and the command travel, the last row of column `metric_mv_travel` in m/s. *Checkpoint:* 4.52 km, 1.55 mmol/L, 43.3 m/s.

(b) Reload the scenario, set Heart-rate loop gain to 1.10 % and Heart-rate integral time to 150 s before Run, leave the ramp at 60 s, and repeat. *Checkpoint:* 6.34 km, 3.55 mmol/L, 9.2 m/s.

(c) Compute the distance lost to the hunt and the ratio of the command travels. Which of the three readings would mislead an operator watching only the constrained variable? *Checkpoint:* 1.82 km and 4.7.

Tolerances: 0.02 km, 0.02 mmol/L and 0.5 m/s.

6.9 App challenge · 9 min

Load the challenge, One hour, flat, for distance, and keep its seed. Before Run set Heart-rate target to 90 % and Lactate target to 3.7 mmol/L, keep Auto · PID, set the units to SI (km/h and kilometers), run the hour at speed 60 and export the CSV file. Skip any lines that start with #. The file's time column is in hours.

(a) Read the last row of column `dist_km` and the Reserve left tile. *Checkpoint:* 9.584 km and 189 m.

(b) Reload, set the same two targets and Auto · MPC before Run, and repeat. *Checkpoint:* 9.947 km and 36 m.

(c) In both files read `dist_km` in the row at 0.05 h. What share of the final gap exists by three minutes? *Checkpoint:* 0.331 km and 0.541 km, a gap of 210 m, 58 % of 363 m.

(d) Using Table 6.8, say which part of the gap is a tuning choice and which is architectural. *Checkpoint:* The start is a tuning choice, and the finish is architectural.

Tolerances: 0.005 km on distances and 1 m on the reserve.

6.10 Application area · 3 min

A clinic wants to use the application's heart-rate band, 83 to 92 % of 220 − age, for a 60-year-old client.

(a) Compute the estimated maximum and the band. *Checkpoint:* 160 bpm, 133 to 147 bpm.

(b) The estimator carries a population standard deviation near 11 bpm. How far can the band sit from the client's true band, and what would you set in the application, from what measurement? *Checkpoint:* About ten beats. Set Δ_{HR} from a measured maximum.

6.11 Application area · 4 min

Scenario 9 puts the recreational runner in a room at 28 °C and 60 % relative humidity with the airflow at 0.15.

(a) From Equation (6.13) with $E_0 = 1820\,\mathrm{W}$ and $\Phi = 1$, compute the evaporative ceiling $E_{\max}$ at airflow 0.15 and at 0.60. *Checkpoint:* 620 W and 909 W.

(b) At threshold pace the runner must shed about $CP - P_{\mathrm{rest}}$, or 867 W, above the resting turnover. At which airflow can the heat balance reach a steady state, and what does that say about the actuator for core temperature? *Checkpoint:* Only at 0.60. The actuator is the fan, not the belt loops.

6.12 Review the agent · 5 min

An AI assistant reads Table 6.8 and writes: "The predictive controller covers 363 m more than the selector because its terminal value prices the reserve beyond the 300 s horizon. To close the gap, give the regulator the same terminal value, and to widen it, double the weight of the predictive controller's terminal terms." Find the flaw, and say how you would check the claim in the chapter and in the app.

CHAPTER 7

Cycling Exertion Simulator to Teach Control of a Depleting Actuator



The controller moves 1 how hard the rider tries, through the legs

The controller reads 1 crank power, from strain gauges in the crank 2 heart rate, from a chest strap

What disturbs it 1 the grade of the road 2 a headwind

Notice that the controller can only set how hard the rider tries, while the watts that effort buys come from the legs and fall when a hard effort is held. Illustration, not to scale.

IN PLAIN TERMS **Pacing a Rider Whose Strength Runs Down**

The main idea. A controller on a bicycle cannot set the rider's power, only how hard the rider tries. Above the level a rider can keep up, the same effort buys less power as a small store of energy runs low.

The challenge. Heart rate and blood lactate, a substance that rises in the blood with hard effort, warn of too much effort only minutes later. One effort setting has to serve a power target and these slow limits at once.

Why it matters. A battery that weakens as it is used poses the same problem.

THREE WAYS TO READ THIS CHAPTER

Explore	11 min	The first page, the fact box and the history box of Section 7.2, the two try-it boxes of Section 7.4, the fact box of Section 7.7 and the key points
Study	60 min	The whole chapter with its pause points and the twelve exercises
Practice	20 min	The first page, Section 7.7, Section 7.8, Section 7.9 and the open challenge of Section 7.10

The control question A rider sets exertion, and the watts that exertion buys fall as a finite work capacity drains. How should a pacing loop and two slow physiological limits share that one actuator, and what is the terrain or the end of the ride worth to a controller that plans ahead?

Learning objectives

- Compute the gain from exertion to power below and above the critical-power exertion, and the fade time constant of a held exertion.
- Distinguish saturation against a drained store from integral windup on a run of the power loop.
- Explain how a low selector with margin-scaled external reset shares one actuator, and why the margin is scaled and never negative.
- Compute how much speed a watt buys on the flat and on a climb, and judge what terrain preview is worth on a timed ride.
- Compare the regulator and the predictive controller on the challenge hour and say where the predictive lead is earned.

Prerequisites Sections 1.4.5 (late samples), 1.4.6 (stores and overrides), 1.6.2 (velocity form) and 1.6.4 (scenarios and data). Section 1.6.3 and Chapter 6, which shares the athlete model, are useful but not required.

The app https://apmonitor.com/apps/cycling-rider/

See also Volume 4, Chapter 4 and Volume 4, Chapter 3 for one actuator shared among a governing loop and its limits through a selector with external reset. Volume 4, Chapter 14 for a store that drains as it is used and a defense that ends when it runs out.

7.1 Introduction

An actuator is usually treated as a fixed map from the controller output to the process input. A valve at half travel passes about the same flow tomorrow as it does today, and a tuning written against it stays written. Some actuators do not keep their gain. A battery sags as it discharges, a heat exchanger loses duty as it fouls, a catalyst bed loses activity and an accumulator delivers less flow as it empties. In each case what a given actuator position buys depends on how hard the actuator has been driven in the recent past, and a controller that assumes a fixed gain saturates against a limit that is not a mechanical stop and winds its integral up while it waits [1].

A person on a bicycle is the most familiar actuator of that kind. The quantity a rider decides is effort, which has its own psychophysical scales of perceived exertion [2]. What the effort buys in watts is not fixed. Above critical power the rider draws on a finite work capacity, and the greatest power available at any instant falls in proportion to what remains of that capacity [3, 4]. Cyclists asked to hold a constant rating of perceived exertion let their power output decline steadily, and fastest in the heat [5, 6]. Power is therefore something a rider produces rather than something a rider sets, and effort is the natural manipulated variable.

Three control problems follow from that structure, and none of them is particular to cycling. The first is saturation against a store instead of a stop, where the actuator reaches its limit because the capacity behind it has run down and the integral has to be kept from winding up while it is there. The second is override selection, where one actuator serves a pacing loop and several limits whose measurements trail the effort by minutes, and the losing loops have to stay ready to take over without winding up and without deadlocking the selector [1]. The third is economic operation over a fixed horizon, where the objective grows with the quantity the limits forbid and a stored capacity is worth nothing once the ride ends [7]. A peaking battery system, a generator on a fuel allowance and a batch unit with a fixed cycle time present the same three.

Road-cycling power models predict the power a speed requires on a given road from aerodynamic drag, rolling resistance, gravity and drivetrain loss, and they reproduce measured road power closely [8, 9]. Optimal-pacing studies use such models to schedule power over a course, either against a fixed power budget [10] or against individualized dynamics of fatigue and recovery [11]. Feedback

control of heart rate on a cycle ergometer adjusts the work load to hold a heart-rate reference, and nonlinear and robust designs have been demonstrated in closed-loop experiments [12, 13]. However, both lines of work take power as the decision variable, so the loss of authority as the work capacity drains appears as a constraint on an optimizer or does not appear at all, and neither exercises the selector logic that one actuator shared by a pacing target and several slow limits requires.

This chapter works with a cycling simulator whose manipulated variable is exertion rather than power, so that pacing, override selection and anti-windup all have to work through an actuator whose gain depends on how it has been used. Three measured observations about that plant run through it. A held exertion above the sustainable level settles onto critical power with a closed-form time constant. An external reset whose margin is not scaled to each override's distance from its limit gives up more than a third of the hour's distance, and one whose margin is allowed to go negative ratchets the actuator down. Preview of the terrain is worth almost nothing on this plant, because on a timed ride the grade enters the objective and not the constraints. The road mechanics, the critical-power structure and the control laws are standard.

Section 7.2 traces ergometry, critical power and pacing. Section 7.3 carries the nineteen-state model of the rider, the bicycle and the road, with every equation and constant needed to reimplement it, and Section 7.3.11 says where each assumption breaks down. Section 7.4 measures the gain from exertion to power, the lags of the limits behind the effort and what the terrain changes. Section 7.5 sets out the instruments, the three proportional-integral loops on a low selector and the receding-horizon alternative, and Section 7.6 defines the scorecard. Section 7.7 reports the twenty cases, from open-loop mechanics through saturation and slow overrides to an hour on a rolling circuit and a time trial, and Section 7.7.5 measures what each selector rule is worth. Section 7.8 describes the real power meters, heart-rate straps and head units, and Sections 7.9 to 7.11 give the limitations, the discussion and the conclusion.

7.2 Background: Ergometry, Critical Power and Pacing

The bicycle ergometer and the critical-power construct came out of one laboratory tradition, whose early steps are told in the history box at the end of this section. In its two-parameter form the tolerable duration of a constant power above critical power is the work capacity W' divided by the excess, and critical power is the highest rate of work that can be sustained without drawing on W' [3].

Two refinements made the construct usable inside a simulation. The three-parameter model added the maximal instantaneous power and implied that the power available at any instant is proportional to the work capacity remaining [4], which is the structure of the exertion map in Section 7.3. The three-minute all-out test turned that structure into a measurement. An all-out effort long enough to exhaust W' finishes on a stable power equal to critical power, and the work done above that power matches W' [14]. For intermittent riding the W' balance models track depletion above critical power and reconstitution below it, in an integral and a differential form that rest on different assumptions about recovery [15]. Critical power and the blood-lactate threshold markers correlate closely in recreational cyclists, but with limits of agreement wide enough that the two cannot be used interchangeably [16], and the simulator carries them as separate parameters for that reason.

> **✦ DID YOU KNOW? A Store That Refills Slowly**
>
> In one laboratory study, the work capacity above critical power came back to 37, 65 and 86 % of its full value after 2, 6 and 15 min of recovery at a nominal 20 W. Its half time of 234 s was slower than that of oxygen uptake, 74 s, and faster than that of blood lactate, 1366 s [17].

The mechanics of the moving bicycle were written down separately. Towing experiments gave the resistance of a cyclist as a rolling term plus an aerodynamic term quadratic in air speed, and an equation of motion for the rider followed from it [8]. A validated model of road-cycling power later reproduced power measured on the road with $R^2 = 0.97$ and a standard error of 2.7 W [9]. Pacing

studies built on that model predict that varying a 289 W mean power by 10 % with gradient and wind saves 126 s on a hilly course against constant power [10], and optimal-control formulations treat the rider's maximal power as a time-varying constraint driven by fatigue and recovery [11]. Feedback control reached the ergometer before it reached the road. Heart rate has been regulated through the work load by a nonlinear generalization of PI control [12] and by robust linear designs with root-mean-square tracking errors of 2.5 to 3.1 bpm over 73 closed-loop experiments [13].

The experiments that fix effort instead of power are the ones closest to this simulator. At a fixed rating of perceived exertion, cyclists let power decline linearly in cool, normal and hot trials, fastest in the heat [5], and over one hour at a constant rating in the moderate domain, power output, oxygen uptake and ventilation all fell [6]. Those studies measure what a held effort buys and do not close a loop around it, and the controllers in the pacing and heart-rate literature do not take effort as their output. The application of this chapter does both on one plant, with the exertion map carrying the three-parameter structure and the controllers acting on the exertion alone.

> ⧖ HISTORY **From One Muscle Group to the Bicycle (1913 to 1981)**
>
> The bicycle entered the physiology laboratory as an instrument for imposing a known rate of work, and an ergometer built for respiration measurements during pedaling was described in 1913 [18]. That the power a person can hold falls with the time it must be held had long been established. A simple mathematical frame for it came in 1965 with the concept of critical power, defined for a single synergic muscle group. The most work that can be done grows in a straight line with the time it lasts, with a fixed energy reserve as the intercept and critical power as the slope [17, 19]. In 1981 the concept was applied to cycling, where critical power correlated highly with the anaerobic threshold [17, 20]. Measuring it stayed costly. A valid value took two to seven laboratory visits until a single all-out test of three minutes arrived [17].

7.3 Rider, Bicycle and Road Model

The plant is one rider on one bicycle on a straight road, either for a fixed 3600 s or over a fixed distance, starting from a standstill. Exertion $E \in [0, 1]$ is the only manipulated variable. The road is either set by hand as a percent grade $g_{\%} = 100 \tan\theta$ or supplied as a function of distance by one of two fixed routes, and a headwind w is a second disturbance. The quantities that appear in the equations are labeled on the drawing in Figure 7.1. The model works in SI units, except that heart rate stays in beats per minute and lactate and glucose in mmol/L, the units in which their limits are written. The recreational preset is the design case throughout.

The model integrates nineteen states, and Table 7.1 classifies the ones a controller or a student acts on. Five belong to the road and the drivetrain, being speed, distance, altitude, cumulative climbing and the applied exertion. Eleven are physiological, being oxygen uptake, heart rate, lactate, the work capacity, fatigue, core temperature, respiratory frequency, muscle and liver glycogen, circulating glucose and carbohydrate in the gut. One is an interstitial glucose estimate that is displayed and not controlled, and two accumulate the energy absorbed by the brakes and the work done at the cranks. Every parameter is adjustable and displayed, and nothing is withheld.

7.3.1 Exertion and the Power It Buys

The applied exertion follows the command with a first-order neuromuscular lag, as shown in Equation (7.1).

$$\tau_a \frac{\mathrm{d}E}{\mathrm{d}t} = u - E \tag{7.1}$$

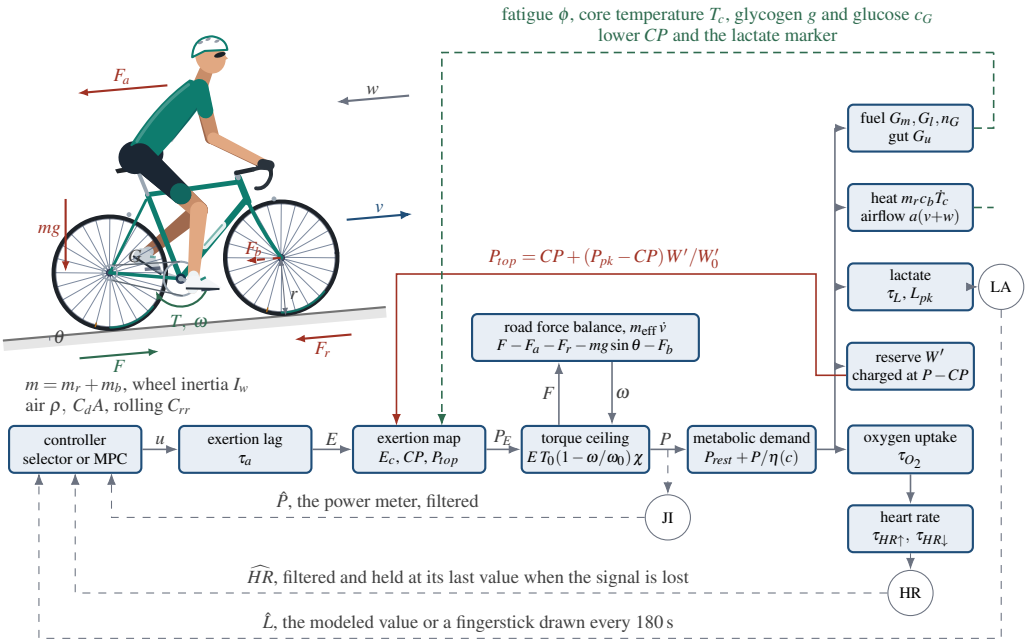


Figure 7.1 One rider on one bicycle on a road of grade angle θ. The controller sets the exertion command u, the rider applies exertion E after a neuromuscular lag τ_a, and the exertion map turns it into the power P_E the rider asks for, from critical power CP and a ceiling P_{top} that falls with the remaining work capacity W'. The torque ceiling bounds the crank torque T at crank speed ω, and the drive force F through gear G, wheel radius r and drivetrain efficiency η_d works against aerodynamic drag F_a in the apparent wind $v + w$, rolling resistance F_r, gravity mg and the brakes F_b. Crank power P sets the metabolic demand, which drives oxygen uptake, heart rate, lactate, the reserve, heat and fuel in parallel, and the slow states return to critical power and the lactate marker as one bundle. The controller reads a filtered power $\hat{P}$, a filtered heart rate $\widehat{HR}$ and a lactate value $\hat{L}$, and commands exertion alone. The rider and bicycle are the application's own drawing.

where $u \in [0, 1]$ is the command and $\tau_a = 1.0\,\text{s}$. Critical power at the cranks falls with accumulated fatigue and with the fuel state, as shown in Equation (7.2).

$$\begin{aligned} CP &= CP_0\,(1-\phi)\,\Psi_G\,\Psi_c, \qquad g = \frac{G_m + G_l}{G_{m,0} + G_{l,0}} \\ \Psi_G &= 0.80 + 0.20\,\text{clamp}\left(\tfrac{g}{0.15}, 0, 1\right) \\ \Psi_c &= 0.85 + 0.15\,\text{clamp}(c_G - 3,\, 0,\, 1) \end{aligned} \tag{7.2}$$

where CP_0 is the value for a fresh rider of the preset, ϕ is the fatigue accumulated so far, g is the share of the starting glycogen still in the stores and c_G is the blood glucose concentration. Both fuel terms stay at one until the stores drop under 15 % and glucose under 4 mmol/L, which leaves critical power almost constant through a fueled hour at moderate effort.

The power that an exertion asks for is written in two domains joined at the exertion E_c at which a

Table 7.1 Roles of the Variables for the Recreational Preset, With Values Read at 30 min of the Level Hour Under the Selector and a 215 W Power Target.

Symbol	Role	Nominal	Range
E	MV	0.702	0–1
P	CV, measured	206 W	0 to P_{top}
HR	CV, measured limit	156.8 bpm	SP 168 bpm
L	CV, estimated	3.74 mmol/L	SP 3.7, limit 4.0
s	objective, timed	16.95 km	—
$g_{\%}$	DV	0 %	−10 to +12
w	DV	0 m/s	−8 to +15
v	state, shown	34.28 km/h	≥ 0
W'	state, shown	16.0 kJ	0–16 kJ
T_c	state, shown	37.91 °C	34–41.5 °C
$G_m + G_l$	state, shown	350 g	0–425 g
$\dot{V}_{O_2}$	state, shown	37.7	3.5–48
ϕ	state, shown	0.000	0–0.40
c	output	88.8 rpm	0–150 rpm
G	discrete, automatic	3.06	1.21–4.55

$\dot{V}_{O_2}$ in mL/(kg min).

fresh rider produces exactly critical power, as shown in Equation (7.3).

$$P_E = \begin{cases} CP\,\dfrac{E}{E_c}, & E \leq E_c \\ CP + (P_{top} - CP)\,\dfrac{E - E_c}{1 - E_c}, & E > E_c \end{cases} \qquad P_{top} = CP + \max\left(0,\ P_{pk}\left(1 - \tfrac{\phi}{2}\right) - CP\right)\frac{W'}{W'_0} \tag{7.3}$$

where P_{pk} is the sprint power of a fresh rider at the optimal cadence, W' is the work capacity above critical power that remains and W'_0 is its full value. Below E_c the rider is inside the sustainable domain and exertion buys a proportional share of critical power. Above it the extra exertion buys a share of the margin between critical power and the ceiling P_{top}, and that ceiling falls linearly with the W' left, which is the three-parameter form in which the power available at any instant is proportional to the remaining capacity [4]. At an empty reserve every exertion above E_c buys critical power and no more. The actuator has then lost its whole upper range without moving, and $E_c = 0.75$ for every preset.

7.3.2 Torque, Cadence and Gearing

Power is torque times crank speed, and a bicycle is started from rest by torque alone. The torque the rider will apply is bounded by exertion times a torque-cadence line through the sprint peak, as shown

in Equation (7.4).

$$\begin{aligned} \omega &= \frac{v}{rG}, \qquad c = \frac{60\,\omega}{2\pi} \\ T_{\max} &= E\,T_0\left(1 - \frac{\omega}{\omega_0}\right)\chi \\ \chi &= \operatorname{clamp}\left(\frac{c_{\max} - c}{15\,\text{rpm}}, 0, 1\right) \\ T &= \min\left(\frac{P_E}{\omega},\ T_{\max}\right) \\ F &= \min\left(\frac{\eta_d\,T}{G\,r},\ F_{\max}\right) \end{aligned} \tag{7.4}$$

where G is the gear in wheel turns per crank turn, r is the wheel radius, η_d is the drivetrain efficiency and $F_{\max}$ is the largest drive force the tire and the rider's balance allow. The constants $\omega_0 = 2\omega_{opt}$ and $T_0 = 2P_{pk}(1-\phi/2)/\omega_{opt}$ put the apex of $T_{\max}\omega$ at the optimal cadence $c_{opt} = 120\,\text{rpm}$, where it equals the sprint power when $E = 1$. The torque-velocity and power-velocity relations of sprint cyclists peak in power at an optimal pedaling rate [21], and the model takes the torque-cadence relation as a straight line through that peak, giving $T_0 = 151.2\,\text{N m}$ and ω_0 equivalent to 240 rpm for the recreational preset. When the traction cap binds, the torque is recomputed from $F_{\max}$, and the delivered crank power is $P = T\,\omega$ in every case. At a standstill $\omega = 0$, the torque is $E\,T_0$ and the power is zero, so the bicycle starts on force and a gentle exertion gives a gentle start. The taper χ replaces a hard cutoff at $c_{\max} = 150\,\text{rpm}$, the highest cadence the rider will pedal, because a cutoff made a rider in too small a gear alternate between pedaling and coasting on successive integration steps. Pedaling stops when $E \le 0.004$ or $\chi = 0$, and the freewheel then leaves the cranks still while the wheels turn.

The drivetrain is a ladder of eleven ratios spaced geometrically from $G_{lo} = 1.21$, which is a 34-tooth chainring on a 28-tooth cog, to $G_{hi} = 4.55$, which is 50 on 11, with $G_i = G_{lo}(G_{hi}/G_{lo})^{i/10}$. Automatic shifting moves one gear per 1 s scan whenever cadence leaves the band $c_{pref} \pm 10\,\text{rpm}$ around the preferred cadence $c_{pref} = 88\,\text{rpm}$, and it selects the lowest gear below 0.3 m/s. The lowest, the middle ($\sqrt{G_{lo}G_{hi}}$) and the highest gear can also be fixed as teaching settings.

7.3.3 Road Force Balance

Road speed follows from the force balance along the road, as shown in Equation (7.5).

$$\begin{aligned} m_{\text{eff}}\frac{\mathrm{d}v}{\mathrm{d}t} &= F - F_a - F_r - m\,g\sin\theta - F_b \\ F_a &= \tfrac{1}{2}\,\rho\,C_d A\,(v+w)\,|v+w| \\ F_r &= C_{rr}\,m\,g\cos\theta \\ F_b &= k_b \max(0,\ v - v_{cap}), \qquad m_{\text{eff}} = m + I_w/r^2 \end{aligned} \tag{7.5}$$

where $m = m_r + m_b$ is the mass of the rider and the bicycle, I_w is the rotational inertia of both wheels, $\theta = \arctan(g_{\%}/100)$, w is the headwind and v_{cap} is the descending speed the rider brakes to hold. The resistance terms are the standard road-cycling model [8, 9]. A bicycle at rest with nothing pushing it forward stays at rest, because the rider has a foot down, so $\dot{v} = 0$ whenever $v = 0$ and the net force is not positive. Distance, altitude, cumulative climbing, braking energy and crank work are the integrals of v, $v\sin\theta$, $\max(0, v\sin\theta)$, $F_b v$ and P. For the recreational preset $m_{\text{eff}} = 84.75\,\text{kg}$, and 50 % exertion on the level gives 146.67 W and a steady 30.10 km/h, at which the wheel power equals drag plus rolling resistance to within 0.1 %, an equality the application's test suite checks.

7.3.4 Metabolic Demand and Gross Efficiency

Metabolic demand is the resting turnover plus crank power divided by a gross efficiency that depends on cadence, as shown in Equation (7.6).

$$\begin{aligned} P_{met} &= P_{rest} + \frac{P}{\eta(c)}, \qquad P_{rest} = \frac{3.5\,\varepsilon_{O_2}}{60}\,m_r \\ \eta(c) &= \max\left(0.10,\ \eta_0\left[1-\kappa\left(\frac{c-c_\eta}{60}\right)^2\right]\right) \end{aligned} \tag{7.6}$$

where $\varepsilon_{O_2} = 20.9\,\mathrm{J/mL}$ is the energy equivalent of oxygen, $\eta_0 = 0.235$ is the peak gross efficiency, $c_\eta = 60\,\mathrm{rpm}$ is the most economical cadence and $\kappa = 0.08$ is the fraction of efficiency lost 60 rpm away from it. The parabola is a declared teaching curve and not a fit. Energy expenditure in cycling is close to linear in work rate, with cadence a secondary factor, and a review of the evidence finds no firm energetically optimal cadence [22]. Equation (7.6) is the only route by which gearing reaches the physiology. At the preferred 88 rpm the efficiency is 0.2309, against 0.2350 at 60 rpm and 0.2162 at 120 rpm. Grade does not appear in it, so a climb at constant exertion changes the speed of the bicycle and leaves the demand on the body where it was.

7.3.5 Oxygen Uptake, Heart Rate and the Work Capacity

The physiology is the lumped athlete model that the treadmill of Chapter 6 also runs, charged here through a bicycle adapter. Oxygen uptake, expressed as the metabolic power it delivers, trails the demand by a single time constant and cannot exceed the aerobic maximum, as shown in Equation (7.7).

$$\begin{aligned} P_{aer} &= \frac{m_r\,\dot{V}_{O_2,\max}\,\varepsilon_{O_2}}{60} \\ f(x) &= \max\left(0,\ \frac{x-P_{rest}}{P_{aer}-P_{rest}}\right) \\ \tau_{O_2}\,\frac{\mathrm{d}\dot{V}}{\mathrm{d}t} &= \min(P_{met},\ P_{aer}) - \dot{V} \end{aligned} \tag{7.7}$$

where $\dot{V}$ is the delivered oxygen uptake in watts and f is the fraction of the oxygen-uptake reserve that a metabolic power represents. For the recreational preset $P_{rest} = 91.4\,\mathrm{W}$ and $P_{aer} = 1254.0\,\mathrm{W}$. The work capacity is charged in crank watts against critical power in crank watts, as shown in Equation (7.8).

$$\frac{\mathrm{d}W'}{\mathrm{d}t} = \begin{cases} -(P-CP), & P > CP \\ (CP-P)\,k_r\left(1-\dfrac{W'}{W'_0}\right), & P \le CP \end{cases} \tag{7.8}$$

where $k_r = 0.5$ is a recovery efficiency. Below critical power this is the differential form of the W' balance, with reconstitution proportional to the deficit [15], slowed by k_r. Heart rate is driven by the delivered oxygen uptake rather than by the demand, with separate onset and recovery lags and drift terms in core temperature and fatigue, as shown in Equations (7.9) and (7.10).

$$\begin{aligned} HR_{eq} &= HR_{rest} + (HR_{\max} - HR_{rest})\,f(\dot{V}) \\ &\quad + k_T \max(0, T_c - 37.8) + k_F\,\frac{\phi}{\phi_{\max}} \end{aligned} \tag{7.9}$$

$$\tau_{HR}\,\frac{\mathrm{d}HR}{\mathrm{d}t} = \mathrm{clamp}(HR_{eq},\ 35,\ 1.02\,HR_{\max}) - HR \tag{7.10}$$

where τ_{HR} is $\tau_{HR\uparrow}$ while HR_{eq} exceeds HR and $\tau_{HR\downarrow}$ otherwise. The two drift terms make heart rate climb at a constant power once the rider warms or tires.

7.3.6 Blood Lactate

Lactate is modeled in a single mixed compartment. Its steady value rises as a power of the metabolic intensity above a knee and is calibrated to cross the constraint $L_4 = 4\,\mathrm{mmol/L}$ at the rider's lactate-marker power, as shown in Equation (7.11).

$$\begin{aligned} f_T &= \frac{P_L/\eta(c_{pref})}{P_{aer} - P_{rest}} \\ f_{T,e} &= \max\big(0.40,\ f_T\,(1-\phi)\,\Psi_G\Psi_c\big) \\ x &= \max\left(0,\ \frac{f(P_{met}) - f_0}{f_{T,e} - f_0}\right) \\ L_{ss} &= L_{rest} + (L_4 - L_{rest})\,x^q\,\Lambda \\ L^* &= \begin{cases} L_{ss}, & L_{ss} \le 8 \\ 8 + \Delta_L\left(1 - \mathrm{e}^{-(L_{ss}-8)/\Delta_L}\right), & L_{ss} > 8 \end{cases} \\ \tau_L \frac{\mathrm{d}L}{\mathrm{d}t} &= L^* - L \end{aligned} \tag{7.11}$$

where P_L is the crank power at which steady lactate crosses L_4 at the preferred cadence, $f_0 = 0.35$ is the knee, $q = 3$ is the curvature, $\Lambda = 0.6 + 0.4\,\mathrm{clamp}(g/0.2, 0, 1)$ carries a depleted store, the denominator $f_{T,e} - f_0$ is floored at 0.05, and $\Delta_L = L_{pk} - 8$ with $L_{pk} = 20\,\mathrm{mmol/L}$ sets a soft ceiling. The ceiling is the one change the bicycle needs in the lactate model. A sprint reaches several times the metabolic power at maximal oxygen uptake, where a power law calibrated on submaximal work asks for hundreds of millimoles per liter, so L^* leaves the curve alone up to 8 mmol/L and bends it with a continuous slope toward L_{pk} above that. The marker $P_L = 210\,\mathrm{W}$ sits 10 W below critical power for the recreational preset, and the application's test suite checks that every preset settles on exactly 4 mmol/L at its own marker power.

7.3.7 Fatigue, Core Temperature and Breathing

Fatigue is what lowers critical power in (7.2) when lactate stays above the constraint or the core stays hot, and it recovers whenever the rider works below critical power, as shown in Equation (7.12).

$$\begin{aligned} \Gamma &= k_L\,\frac{\max(0, L - L_4)}{\max(0.5,\ L_4)} + k_{TF}\max(0, T_c - 38.5) \\ \frac{\mathrm{d}\phi}{\mathrm{d}t} &= \Gamma\left(1 - \frac{\phi}{\phi_{\max}}\right) - k_\phi\,\phi\,\mathrm{clamp}\left(1 - \frac{P}{CP}, 0, 1\right) \end{aligned} \tag{7.12}$$

Heat is balanced in a single body node whose losses are referenced to the resting state, and every crank joule leaves the body as work, as shown in Equation (7.13).

$$\begin{aligned} m_r c_b \frac{\mathrm{d}T_c}{\mathrm{d}t} &= \max(0,\ P_{met} - P) - Q_{loss} \\ Q_{loss} &= P_{rest} + G_d\,(T_c - 37.0) \\ &\quad + \min\big(G_e(T_c - 37.05),\ E_{\max}\big) \end{aligned} \tag{7.13}$$

The conductances scale with the rider and the air, and the airflow over the rider is the apparent wind, as shown in Equation (7.14).

$$\begin{aligned} a &= \mathrm{clamp}\left(a_f + \sqrt{\max(0,\ v+w)/v_{ref}},\ 0,\ 1.6\right) \\ \Phi &= \frac{m_r \dot{V}_{O_2,\max}}{75 \times 45} \\ G_d &= (g_0 + g_a a)\,\frac{m_r}{75}\,\mathrm{clamp}\left(1 + \tfrac{20-T_{air}}{25}, 0.3, 1.8\right) \\ G_e &= k_s \Phi\,(1 - 0.55\varphi)(0.45 + 0.55a) \\ E_{\max} &= E_0\,\Phi\,(1 - 0.6\varphi)(0.45 + 0.55a) \end{aligned} \tag{7.14}$$

where a_f is any extra airflow from a fan or a breeze, $v_{ref} = 8\,\mathrm{m/s}$ is the air speed that gives full convective cooling, φ is the relative humidity, and Φ scales the rider's aerobic capacity by that of a 75 kg rider at 45 mL/(kg min). In this equation a is the airflow and not the loop action of Section 1.3, and $E_{\max}$ and E_0 are the evaporative ceiling and its constant, not the exertion. The apparent wind is the single change the bicycle makes to the thermal model, and it is why a slow climb with the wind behind is the hottest riding in the application. The ride is stopped when T_c reaches 40.5 °C. Writing the loss about 37.0 °C makes a resting rider in still air at 20 °C exactly steady. Respiratory frequency follows $\tau_R \dot{f}_R = \min\big(70,\ f_{R,0} + k_R f(\dot{V}) + 2\max(0, L-2) + 3\max(0, T_c - 38)\big) - f_R$ in breaths per minute and is displayed rather than controlled.

7.3.8 Carbohydrate, Glucose and Feeding

Both fuels burn at once. The carbohydrate share rises with intensity, fat oxidation has a rate ceiling, and anything supplied above the aerobic maximum is charged to carbohydrate, so the work capacity is never free energy, as shown in Equation (7.15).

$$\begin{aligned} \sigma &= \mathrm{clamp}\big(0.20 + 0.78\,f(P_{met}),\ 0.15,\ 0.98\big) \\ P_F &= \min\left(P_{F,\max}\left[0.7 + 0.3\left(1 - \tfrac{G_m}{G_{m,0}}\right)\right],\ (1-\sigma)P_{del}\right) \\ \dot{n}_C &= \frac{P_{del} - P_F + P_{an}}{e_G}, \qquad P_{del} = \min(P_{met}, P_{aer}) \end{aligned} \tag{7.15}$$

where $P_{an} = P_{met} - P_{del}$ and $e_G = 3153\,\mathrm{J/mmol}$ is the energy in a millimole of glucose equivalent. The carbohydrate flux is drawn from muscle glycogen and from circulating glucose, the liver and gluconeogenesis defend the circulating pool, and carbohydrate eaten during the ride empties from the gut into the blood, as shown in Equation (7.16).

$$\begin{aligned} \beta &= \mathrm{clamp}\left(0.18 + 0.35\left(1 - \tfrac{G_m}{G_{m,0}}\right), 0, 0.6\right) \\ \dot{n}_w &= \max\big(0,\ \beta\dot{n}_C + 0.02\,(c_{G,0}V_G - n_G)\big) \\ \dot{n}_g &= \min(k_g,\ \dot{n}_w), \qquad \ell = \mathrm{clamp}\left(\tfrac{G_l}{0.15\,G_{l,0}}, 0, 1\right) \\ \dot{n}_H &= \mathrm{clamp}(\dot{n}_w - \dot{n}_g,\ 0,\ \dot{n}_{H,\max}\,\ell) \\ \frac{\mathrm{d}G_m}{\mathrm{d}t} &= -(1-\beta)\,\dot{n}_C M_G, \qquad \frac{\mathrm{d}G_l}{\mathrm{d}t} = -\dot{n}_H M_G \\ \frac{\mathrm{d}G_u}{\mathrm{d}t} &= -\frac{G_u}{\tau_u}, \qquad \dot{n}_r = k_u \max\left(0,\ \frac{n_G}{V_G} - c_{G,0}\right) \\ \frac{\mathrm{d}n_G}{\mathrm{d}t} &= \dot{n}_H + \dot{n}_g - \beta\dot{n}_C + \frac{G_u}{\tau_u M_G} - \dot{n}_r \end{aligned} \tag{7.16}$$

where n_G is circulating glucose in millimoles, $c_G = n_G/V_G$, $V_G = 0.25\,\mathrm{L/kg}$ times body mass is its distribution volume, $M_G = 0.180\,16\,\mathrm{g/mmol}$, $\tau_u = 600\,\mathrm{s}$ is the gastric emptying constant and $k_u = 0.12\,\mathrm{L/s}$ is an illustrative uptake above the resting concentration. The uptake term is not an insulin model. It keeps a meal from accumulating in a small circulating compartment, it is zero whenever glucose is at or below its starting value, and the glucose it removes is oxidized in place of fat, which is not a tracked store. A feeding event adds 40 g to G_u. Carbohydrate eaten during prolonged cycling holds plasma glucose up and postpones fatigue without slowing the use of muscle glycogen [23], which is the behavior (7.16) reproduces. An interstitial glucose estimate follows c_G with a 480 s lag for display. After each step the states are bounded to $W' \in [0, W'_0]$, $\phi \in [0, \phi_{\max}]$, $L \geq 0.2$, $T_c \in [34, 41.5]$, $f_R \in [5, 75]$, nonnegative stores and $c_G \geq 0.45\, c_{G,0}$.

7.3.9 Finishing Allowances

Both scored limits open over the end of the ride, as shown in Equation (7.17), because work capacity carried over the finish line was never used.

$$\begin{aligned} \zeta_t &= \mathrm{clamp}\left(\tfrac{t-(T-t_w)}{t_w}, 0, 1\right) \\ \zeta_s &= \mathrm{clamp}\left(\tfrac{s-(S-s_w)}{s_w}, 0, 1\right) \\ L_{lim} &= L_4 + (L_f - L_4)\,\zeta^2 \\ HR_{lim} &= HR_{hi} + (HR_{\max} - HR_{hi})\,\zeta^2 \end{aligned} \tag{7.17}$$

where ζ is ζ_t on a timed ride of length T and ζ_s on a route of length S. With $T = 3600\,\mathrm{s}$, $t_w = 300\,\mathrm{s}$, $s_w = 1000\,\mathrm{m}$ and $L_f = 6\,\mathrm{mmol/L}$, the lactate limit holds at 4.0 mmol/L until the last five minutes or the last kilometer, and the heart-rate limit $HR_{hi} = 0.92\,HR_{\max}$ is 170.2 bpm for the recreational preset before it opens toward the maximum. A timed ride ends at T, a route ends at $s = S$, and either ends early if core temperature reaches the stop.

7.3.10 Routes, Parameters, Presets and Numerics

Two routes are fixed by distance, so every comparison meets each hill at the same place and nothing about the terrain depends on a random stream. The rolling circuit is a 13.2 km lap of fifteen segments with grades from −5 to +8 %. Its two climbs rise at 3, 6 and 2 % over 2.1 km and at 4, 8 and 3 % over 1.5 km, and the lap climbs and descends 163 m with its high point 106.8 m above the start. The time-trial climb is 10 km in eight pitches of 3 to 7 %, gaining 518 m at a mean of 5.2 %. Each change of grade is blended over 100 m with a smoothstep, because a step in grade is a step in the gravity term.

Table 7.2 lists the constants of the recreational preset. Three presets write the rider’s mass, aerobic capacity, heart-rate range, critical power, lactate marker, work capacity, sprint power, efficiency, glycogen stores and physiological lags in one action, and every value stays editable afterward. Table 7.3 gives the three and the quantities derived from them. Each preset’s critical power sits between 78 and 86 % of the crank power at which oxygen uptake reaches its maximum, and each lactate marker sits 5 to 15 W below critical power, and the test suite fails if either leaves its band.

The framework of Section 1.6.1 integrates the model by the classical fourth-order Runge-Kutta method with a fixed step of 0.25 s, and halving that step changes the first ten minutes of scenario 8, which include two surges, by less than a meter, a convergence check the test suite repeats. The controllers execute once per simulated second, the predictive controller replans once per five seconds, and the historian keeps one sample per second. Measurement noise is drawn from a seeded stream, and the routes use no random numbers at all, so reloading a case reproduces its run to the last digit. Reset, a change of road and the four playback speeds, from real time to sixty times faster, behave as that section describes.

Table 7.2 Model Parameters of the Recreational Preset. All Are Adjustable and Displayed in the Application.

Symbol	Meaning	Value	Unit
m_r	rider mass	75	kg
$\dot{V}_{O_2,\max}$	maximal oxygen uptake	48	mL/(kg min)
HR_{rest}, $HR_{\max}$	resting, maximum heart rate	60, 185	bpm
CP_0	critical power at the cranks	220	W
P_L	lactate-marker power	210	W
W'_0	work capacity above CP	16000	J
P_{pk}	sprint power	950	W
E_c	exertion at critical power	0.75	—
τ_a	exertion lag	1.0	s
c_{opt}, $c_{\max}$	sprint cadence, highest cadence	120, 150	rpm
η_0	peak gross efficiency	0.235	—
c_η, κ	most economical cadence, curvature	60, 0.08	rpm, —
c_{pref}	preferred cadence	88	rpm
m_b	bicycle mass	8.5	kg
I_w	wheel inertia, both wheels	0.14	kg m^2
r	wheel radius	0.335	m
η_d	drivetrain efficiency	0.975	—
G_{lo}, G_{hi}	lowest and highest gear	1.21, 4.55	—
C_dA	drag area	0.32	m^2
C_{rr}	rolling-resistance coefficient	0.0045	—
$F_{\max}$	traction limit	300	N
v_{cap}, k_b	braking speed, brake gain	16, 300	m/s, N s/m
ρ	air density	1.20	kg/m^3
T_{air}, φ	air temperature, humidity	20, 0.45	°C, —
a_f, v_{ref}	extra airflow, full-cooling air speed	0, 8	—, m/s
τ_{O_2}	oxygen-uptake lag	35	s
$\tau_{HR\uparrow}$, $\tau_{HR\downarrow}$	heart-rate onset, recovery lags	40, 60	s
k_T	heart-rate drift with heat	7	bpm/°C
k_F	heart-rate drift at full fatigue	10	bpm
τ_L, L_{rest}	lactate clearance, resting lactate	230, 1.0	s, mmol/L
f_0, q	lactate knee, curvature	0.35, 3.0	—
L_{pk}	lactate soft ceiling	20	mmol/L
τ_R, $f_{R,0}$, k_R	respiration lag, rest, gain	20, 12, 38	s, 1/min
k_r	W' recovery efficiency	0.5	—
k_L	fatigue from lactate	1.7×10^{-4}	1/s
k_{TF}	fatigue from heat	4.6×10^{-5}	1/(s °C)
k_ϕ, $\phi_{\max}$	fatigue recovery, largest loss	2.0×10^{-4}, 0.40	1/s, —
c_b	specific heat of the body	3500	J/(kg K)
k_s, E_0	evaporative gain, ceiling	900, 1820	W/K, W
g_0, g_a	dry loss, still air and airflow	16, 22	W/K
T_{stop}	core temperature that stops the ride	40.5	°C
$G_{m,0}$, $G_{l,0}$	muscle and liver glycogen	350, 75	g
$c_{G,0}$, V_G	starting glucose, its volume	5.0, 0.25	mmol/L, L/kg
k_g	gluconeogenesis	0.02	mmol/s
$\dot{n}_{H,\max}$	largest hepatic output	0.18	mmol/s
$P_{F,\max}$	fat oxidation ceiling	420	W
τ_u, k_u	gastric emptying, glucose uptake	600, 0.12	s, L/s
L_4, L_f	lactate limit, at the finish	4.0, 6.0	mmol/L
HR_{hi}	heart-rate limit	$0.92\,HR_{\max}$	bpm
t_w, s_w, T	finishing window, distance, ride	300, 1000, 3600	s, m, s

7.3.11 Assumptions

Every assumption below is made for teaching purposes. The application does not prescribe training, estimate anybody's fitness or give medical guidance, and its numbers belong to three invented riders. Each assumption is listed with the situation in which it breaks down.

Exertion is a normalized effort and not a measured signal. Perceived effort has a psychophysical scale and a literature of its own [2, 5], and the two-domain map of (7.3) is a modeling choice built to reproduce the three-minute all-out test. At a fixed rating of perceived exertion real cyclists reduce

Table 7.3 The Three Rider Presets and the Actuator Gains They Imply, Fresh. The Gains Are the Slopes of (7.3) Below and Above E_c.

	Novice	Recreational	Trained
m_r (kg), $\dot{V}_{O_2,\max}$	82, 35	75, 48	68, 68
HR_{rest}, $HR_{\max}$ (bpm)	66, 182	60, 185	48, 188
CP_0, P_L (W)	150, 140	220, 210	315, 300
W_0' (kJ), P_{pk} (W)	12, 700	16, 950	22, 1250
η_0	0.215	0.235	0.245
$G_{m,0}$, $G_{l,0}$ (g)	300, 70	350, 75	430, 90
$\tau_{HR\uparrow}$, $\tau_{HR\downarrow}$, τ_L (s)	45, 70, 250	40, 60, 230	30, 50, 200
CP_0 over crank power at $\dot{V}_{O_2,\max}$	0.789	0.820	0.856
$\partial P/\partial E$ below E_c (W)	200	293	420
$\partial P/\partial E$ above E_c (W)	2200	2920	3740
ratio of the two	11.0	9.95	8.90
τ_W at $E = 1$ (s)	21.8	21.9	23.5

their power in anticipation, at a rate set by the rate of heat storage [5], and the map carries that only through the fatigue and heat terms.

The power ceiling falls linearly with W'. That is the three-parameter form [4], and it implies that a held exertion never quite exhausts W'. Riders at exhaustion do not always use the whole capacity [4], and measured reconstitution follows time constants that the single recovery efficiency k_r only approximates [15].

The torque-cadence relation is a straight line through the sprint peak. Grinding a low cadence is fatiguing in ways the model does not carry, standing changes the available torque, and the model has neither effect.

Gross efficiency is a parabola with its apex at 60 rpm. The evidence does not fix an energetically optimal cadence [22], and the curve is a declared teaching value.

The road is straight and the rider is alone. There is no cornering, no drafting, no crosswind, no bicycle handling and no change of position, so the drag area is constant.

The physiology is lumped. Lactate has a single compartment, heat has a single body node without any dehydration, heart rate responds only to intensity, temperature and fatigue, and the glucose sink above the resting concentration is illustrative rather than an insulin model.

Critical power and the lactate marker are separate but fixed. Both fall with fatigue and fuel in the same proportion, which keeps them close throughout a ride, while the literature finds them related but not interchangeable [16].

Rides repeat exactly. A scenario run twice gives identical numbers, which simplifies grading and describes no real rider, whose second hour is never the same as the first.

7.4 Dynamic Analysis

7.4.1 An Actuator Whose Gain Drains

The distinctive feature of this plant is the gain from the actuator to power. Below E_c it is $\partial P_E/\partial E = CP/E_c$, which is 293 W per unit exertion for the recreational preset and falls only slowly with fatigue and fuel. Above E_c it is $(P_{top} - CP)/(1 - E_c)$, which is 2920 W per unit for a fresh rider, ten times larger, and it falls in proportion to the W' left. Table 7.3 gives the same pair for the three presets, with a ratio of 8.90 to 11.0 between the two slopes.

The consequence of holding an exertion above E_c follows in closed form. With critical power constant and the torque ceiling inactive, (7.3) and (7.8) give $\dot{W}' = -(P_{pk} - CP)(E - E_c)W'/[(1 -$

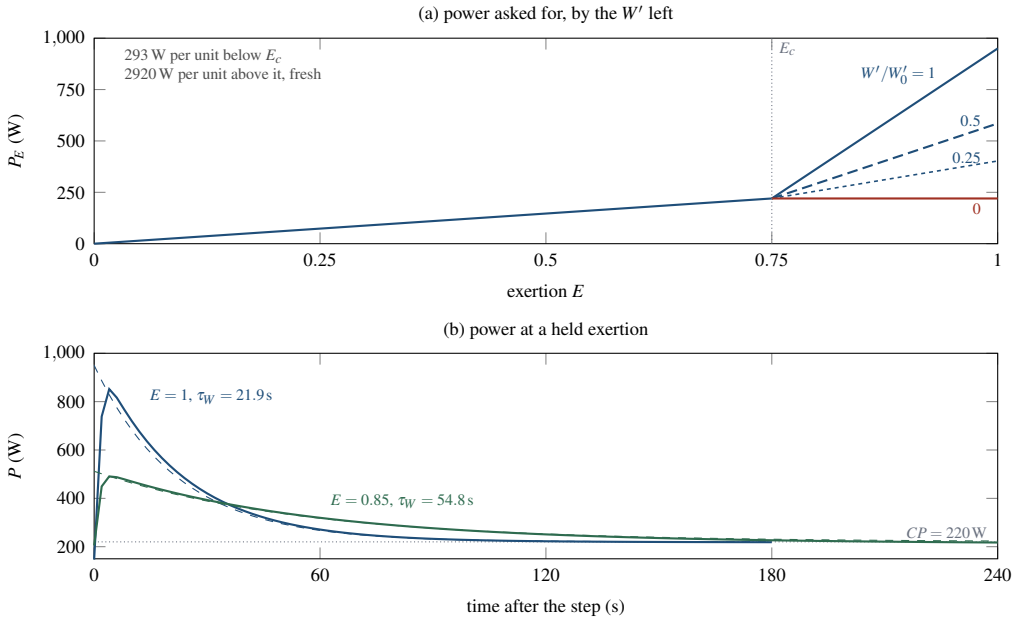


Figure 7.2 The actuator of the recreational preset. (a) Power asked for as a function of exertion for four fractions of the work capacity left, from Equation (7.3) with $CP = 220$ W. (b) Power after a step to a held exertion from full W', in scenario 3 at $E = 1$ and scenario 2 at $E = 0.85$. Solid traces are the application's trajectories and dashed traces are the closed form of Equation (7.18).

$E_c)W'_0]$, and the power settles exponentially onto critical power, as shown in Equation (7.18).

$$P(t) = CP + (P_{pk} - CP)\,\frac{E - E_c}{1 - E_c}\,\mathrm{e}^{-t/\tau_W}$$
$$\tau_W = \frac{W'_0\,(1 - E_c)}{(P_{pk} - CP)(E - E_c)} \tag{7.18}$$

For the recreational preset τ_W is 21.9 s at $E = 1$ and 54.8 s at $E = 0.85$, the time constant of Equation (1.11) in Section 1.4.6. A held exertion therefore never exhausts W' in finite time. It asymptotically spends it, and the power it buys decays onto critical power while the exertion itself does not change. That is also why the three-minute all-out test measures critical power [14]. At $E = 1$ the model's power is within 0.2 W of critical power after 180 s, with W' spent to within 5 J. Figure 7.2 compares (7.18) with the application's own trajectories. They fall short of the closed form in the first seconds, where the exertion lag τ_a delays the rise and, at full effort, the torque ceiling holds the power below 882 W at 88 rpm, and they then follow it apart from the few watts of critical power that fatigue takes during the effort.

> **TRY IT IN THE APP** **3 · Three Minutes All-Out**
>
> Load scenario 3 and press Run. The scripted step to 100 % exertion holds until nine minutes. Watch Crank power against the dashed critical power line while the exertion trace stays flat. The power is 839 W five seconds into the effort, 271 W after a minute and 219 W after three, settled onto critical power with the work capacity spent.

A power loop tuned in the sustainable domain therefore sees its loop gain rise by a factor of ten the moment its output crosses E_c and then fall as W' drains. With the gains of Section 7.5 the

Table 7.4 Apparent Time Constants After a Step in Exertion From Zero to 70 % at $t = 30$ s on the Level, Recreational Preset, Taken as the Time to 63 % of the Change Reached by 90 min, at the 0.25 s Integration Step.

Output	Initial	Final	t_{63} (s)
Crank power (W)	0	205.3	1.0
Cadence (rpm)	0	88.7	2.5
Road speed (km/h)	0	34.24	10.5
Oxygen uptake (mL/(kg min))	3.50	37.57	35.75
Respiratory frequency (1/min)	12.0	44.4	67.75
Heart rate (bpm)	60.0	156.4	81.75
Blood lactate (mmol/L)	1.00	3.67	229.5
Core temperature (°C)	37.00	37.91	328.25

dimensionless loop gain $K_c\,\partial P/\partial E$ is 0.667 below E_c for every preset and 6.64 just above it for a fresh recreational rider. Holding 245 W above the knee in scenario 15, with the meter noise off, the power standard deviation between five and ten minutes is 1.05 W, so the higher gain does not destabilize the loop. What it does do is hand the loop an actuator that runs out. In the case as loaded, with the meter noise on, the exertion needed to hold the target climbs from 76.0 % at five minutes to 81.4 % at ten, the command reaches the stop at 688 s, and the power then falls to within 4 W of critical power with the actuator saturated.

Pause. Try Exercise 7.1 before reading on.

7.4.2 Lags Between the Effort and the Limits

The time each output needs to cover 63 % of its eventual change after a step from rest to 70 % exertion on the level, a step inside the sustainable domain, is listed in Table 7.4 from one run of the application.

The ordering in Table 7.4 is the second half of the control problem. The watts follow the effort within a second and the bicycle within eleven, while heart rate is 82 s behind and blood lactate 230 s behind. A surge above E_c therefore spends W' at the rate of (7.8) for minutes before either physiological limit reports it, and an override on lactate is structurally too slow to protect W'. It can hold the average effort at a level it can measure, and it cannot stop a fast pacing loop from overspending in the meantime, which scenario 16 measures.

> **TRY IT IN THE APP** **16 · Pacing by Lactate Alone**
>
> Load scenario 16, which sets the power target out of reach, and press Run. Exertion reaches 92 % in the first minute, and the lactate loop takes over at 68 s, yet lactate keeps rising to 6.24 mmol/L. Press Reset, set Power target to 215 W and press Run again. The lactate loop now takes over at 445 s, and the hour stays compliant.

7.4.3 Terrain Enters the Objective and Not the Constraints

At constant exertion the grade reaches the force balance (7.5) and nothing else, so a climb changes the speed and leaves the watts and the body where they were. On a timed ride what matters for pacing across terrain is the speed one more watt buys, $\partial v/\partial P = 1/(\partial P/\partial v)$, where $\partial P/\partial v$ is the derivative of the resistance power $(F_a + F_r + mg\sin\theta)\,v/\eta_d$. For the recreational preset it is 53.4 W per m/s on the level at 33 km/h and 59.9 W per m/s on a 6 % climb at 11.3 km/h, so a watt buys 0.0187 m/s on the flat and 0.0167 m/s on the climb. Aerodynamic drag dominates the first and gravity the second, and the two sensitivities differ by 11 %. That is the contrast with a treadmill, where the same grade at the same belt speed raises the metabolic cost of every meter and moves the constraint itself.

The small difference has a measurable consequence. The challenge hour on the rolling circuit was ridden open loop with the exertion schedule $u = u_0 + k\,g_{\%}$ for slopes k between -0.0025 and

0.02 per percent of grade, each paired with the largest base exertion u_0 that keeps the hour compliant, found by bisection. The best constant exertion, $k = 0$ with $u_0 = 71.34\,\%$, covers 28 999 m. The best grade proportional schedule, $k = 0.00125$ per percent, covers 29 024 m, a gain of 25 m or 0.09 %. At $k = 0.01$ the hour is 102 m shorter and at $k = 0.02$ it is 692 m shorter, because the binding limit is a four-minute average of the effort that a climb lasting several minutes fills. The prediction that variable power saves time [10] rests on a fixed-distance objective and a power budget, and on a timed ride under a lactate limit neither condition holds.

7.5 Controller Design

Exertion can be set in three ways. In manual the operator's value goes straight to the rider, in the regulatory mode three proportional-integral loops compete for it through a low selector, and in the predictive mode a receding-horizon search sets it instead. The instruments are common to all three. In either automatic mode the commanded exertion may change by no more than 4 % per second, so switching between modes, or between the two automatic architectures, never produces a step in effort.

7.5.1 Instruments

The power meter reads crank power with colored noise of 4 W standard deviation and a 2 s correlation time, through a 3 s first-order filter. The heart-rate monitor reads heart rate with colored noise of 0.8 bpm and a 3 s correlation time, through a 6 s filter, as shown in Equation (7.19), and holds its last value while the signal is lost.

$$\begin{aligned} \widehat{HR}_k &= \alpha\,\widehat{HR}_{k-1} + (1-\alpha)\,(HR_k + \sigma_H n_k) \\ \alpha &= \mathrm{e}^{-\Delta t_c/\tau_f} \end{aligned} \tag{7.19}$$

where n_k is a first-order filtered Gaussian sequence rescaled to unit stationary variance and $\Delta t_c = 1\,\mathrm{s}$. Lactate is read one of two ways. The default is the modeled value, labeled on every display as an estimate because no continuous blood-lactate sensor exists. The alternative is a fingerstick, a blood sample drawn every 180 s and read 45 s later, held between readings, so that a reading is up to 225 s old when the next one arrives.

7.5.2 Velocity Form PID Algorithm

Each loop runs the shared incremental, or velocity, form algorithm of Equation (7.20) on its own error, with the output bounded to the exertion range.

$$\begin{aligned} e_k &= \mathrm{SP}_k - \mathrm{PV}_k, \qquad p_k = K_c\,e_k \\ s_k &= s_{k-1} + (p_k - p_{k-1}) \\ c_k &= u_{b,k} + s_k + I_{k-1} \\ u_k &= \mathrm{clamp}(c_k,\ 0,\ 1) \\ I_k &= I_{k-1} + \left[\frac{K_c}{\tau_I}\,e_k + \frac{u_k - c_k}{T_t}\right]\Delta t_c \end{aligned} \tag{7.20}$$

These are Equations (1.12) to (1.15) of Section 1.6.2 in the loop's own symbols, with $s_k = S_k$, $c_k = v_k$ the unclipped demand and no derivative term, and anti-windup is back-calculation on I_k alone with the tracking time $T_t = \tau_I$, so the running output is never truncated. All three loops are reverse acting, $a = +1$ in the convention of Section 1.3, because more exertion raises power, heart rate and lactate. The bounds of 0 and 1 are the only limit the back-calculation sees, because the slew limit is applied after the selector and is handled by the conditional integration described below. Derivative action is available on the filtered measurement and is zero on all three loops, because Table 7.4 shows nothing fast enough in the limits to anticipate. In manual, I_k is solved so that c_k equals the operator's exertion, and transfer into automatic starts from there.

Table 7.5 Loop Settings. Every Loop Is PI With $\tau_D = 0$, a 1 s Scan and Output Bounds of 0 and 1.

	Power	Heart rate	Lactate
K_c	$0.5/CP_0$ per W	0.004 per bpm	0.03 per mmol/L
τ_I (s)	12	150	420
bias u_b	$E_0(P_{\mathrm{SP}})$	$E_0(P_L)$	$E_0(P_L)$
default SP	200 W	168 bpm	3.7 mmol/L
margin range R_j	—	15 bpm	1.0 mmol/L

The three loops are the power loop, with $e = P_{\mathrm{SP}} - \hat{P}$ in watts, the heart-rate loop, with $e = HR_{\mathrm{SP}} - \widehat{HR}$, and the lactate loop, with $e = L_{\mathrm{SP}} - \hat{L}$. Their settings are given in Table 7.5. The power-loop gain is carried per unit of fresh critical power, $K_c = 0.5/CP_0$ per watt, so that its loop gain below E_c is $0.5/E_c = 0.667$ for every preset. The heart-rate and lactate loops have loop gains of 0.546 and 0.682 for the recreational preset and 0.49 to 0.74 across the three. The heart-rate setpoint is the limit itself, 168 bpm by default, with an optional ramp from the resting value for a ride paced on heart rate.

Every loop trims around a feedforward rather than integrating its way up from a standstill. The bias is the exertion a fresh rider needs for a power, obtained by inverting (7.3) with full stores, as shown in Equation (7.21).

$$E_0(P) = \begin{cases} E_c\, P/CP_0, & P \le CP_0 \\ E_c + (1 - E_c)\,\dfrac{P - CP_0}{P_{pk} - CP_0}, & P > CP_0 \end{cases} \tag{7.21}$$

The power loop is biased at $E_0(P_{\mathrm{SP}})$ and the two physiological loops at the nominal marker exertion $E_{nom} = E_0(P_L) = 0.716$. Both come from the configured profile, which is what a coach has on paper, and neither knows about the fatigue, the fuel or the heat that move the real rider during a ride. Closing that gap is the integral's job.

7.5.3 Low Selector With Margin-Scaled External Reset

The applied exertion is the smallest of the loop requests, and the loops that lose are reset from the exertion actually applied, as shown in Equation (7.22).

$$\begin{aligned} u^* &= \min_{j \in \mathscr{L}} u_j \\ u_k &= \mathrm{clamp}(u^*,\ u_{k-1} - r_u \Delta t_c,\ u_{k-1} + r_u \Delta t_c) \\ m_j &= \max\left(p_{j,k},\ \frac{\mathrm{SP}_j - \mathrm{PV}_j}{R_j}\right), \qquad m_P = p_{P,k} \\ I_j &\leftarrow u_k - u_{b,j} - s_{j,k} + \max(0,\ m_j) \end{aligned} \tag{7.22}$$

where $\mathscr{L}$ is the set of live loops, $r_u = 0.04/\mathrm{s}$ is the slew rate, R_j is the margin range of an override, and the reset of the last line is applied to every loop j that was not selected. The reset places each losing loop exactly its margin m_j above the applied exertion. For the power loop the margin is its proportional term, so it competes on its present error. For an override it is the larger of its proportional term and its distance from its limit scaled to exertion, so a lactate loop 1 mmol/L below its target sits a whole unit of exertion above the applied value and cannot be selected by accident, while one at its target competes on its proportional term alone and comes down onto the applied value as the limit approaches. The same structure shares the ground speed of a combine among a feed-rate loop and two overrides in Volume 4, Chapter 4, and the fuel of a jet engine among a governor and its limiters in Volume 4, Chapter 3.

Three details of (7.22) decide whether the selector works, and Section 7.7.5 measures each. The margin is scaled to the distance from the limit because a lactate error is at most a few millimoles

per liter whatever the gain, so an unscaled override that is nowhere near binding looks like the most restrictive loop. The margin is never negative, because a loser reset below the applied exertion wins the next comparison, and two overrides past their limits then take turns resetting below each other. On entry into the selector, at a reset or from manual or predictive operation, every loop has been tracking the applied exertion and the smallest increment would win, so the leader is chosen as the live loop with the smallest margin, it starts exactly from the applied exertion, and the others start at their margins above it. The same entry rule governs the return from a signal loss. The fallback is treated as a separate law, so leaving it once the signal is back is an entry, while during the loss the cap is one more candidate that the loops meet with their own requests.

Two further rules complete the selector. When the slew limit holds the applied exertion short of the selected loop's request, that loop's integral keeps the value it had before the scan, which is conditional integration against the rate limit rather than the output bound. When the heart-rate signal is lost the heart-rate loop leaves $\mathscr{L}$ and a conservative cap of $0.9\,E_{nom}$ joins it, because a controller holding a limit on a measurement it no longer has must not chase the frozen value.

7.5.4 Receding-Horizon Alternative

Every 5 s the predictive controller solves a new plan for the next 300 s, a horizon cut to the time remaining on a timed ride but never below 60 s. The plan is an exertion held piecewise constant on six blocks that begin at 0, 5, 15, 30, 50 and 75 % of the horizon. Each plan starts from the controller's own estimate, never from the plant's state: the declared preset stepped each second under the measured crank power, overwritten by the measured heart rate, lactate, speed and distance, a fingerstick through its innovation, with the unmeasured headwind inferred from power and speed. An edited parameter is thus a rider the controller was not told about, and only the ride's limits, duration and route length come from the run. The predictor steps this copy in 5 s increments. The exertion lag is dropped, the gear is the one an ideal automatic shifter would choose, and speed is integrated in five substeps with the crank power of the step held constant. From a standstill there is no cadence and so no crank power, and a force computed from power is zero, so the predictor then evaluates the drive force of (7.4) at every substep. Without that rule every plan predicted a bicycle that never moved, and a rider started under the predictive controller on a level road stayed at rest for the whole hour. The physiology takes one step of the same equations per predictor step, with the four first-order lags advanced by their exact exponential update so that no step length makes them ring, and the slow stores advanced explicitly.

This predictor is the application's own nonlinear search through a copy of the rider and the road, not the linear option of Section 1.6.3. Its objective is economic. On a timed ride the objective is Equation (7.23).

$$
\begin{aligned}
J &= \Delta s_N + s_{tail}(\mathbf{z}_N) + R(\mathbf{z}_N) \\
&\quad - 20\lambda \sum_k |u_k - u_{k-1}| - 12\,\Omega \\
R &= k_R \,\frac{W'_N}{\max(20,\ \partial P/\partial v)}\,\mathrm{clamp}\Big(\tfrac{T - t_N}{t_w}, 0, 1\Big) \\
\Omega &= 5 \sum_k \Big[40\,(L_k - L_{lim} + 0.25)_+ \\
&\qquad + 1.2\,(HR_k - HR_{lim} + 2)_+ + 400\,(T_{c,k} - 40.2)_+ \Big]
\end{aligned}
\tag{7.23}
$$

Here Δs_N is the distance the predictor covers within the horizon, $\mathbf{z}_N$ is the state it predicts at the horizon's end, $\lambda = 0.35$ is the move penalty, $k_R = 0.5$, and Ω accumulates constraint violation over the 5 s prediction steps against margins of 0.25 mmol/L, 2 bpm and 0.3 °C inside the true limits. The tail term s_{tail} estimates the distance that remains available once the horizon ends, under a base policy that holds the nominal marker exertion, rolled out to the end of the ride through a reduced model. The road is integrated in 10 s semi-implicit steps, each charged for the height it actually gains from a

Table 7.6 The Application's Own Performance Measures, Accumulated at the Logging Rate and Frozen When the Ride Ends.

Measure	Definition	Purpose
Distance or time	$s(T)$, or t at $s = S$	the objective
Average power	$\int P\,\mathrm{d}t / t$	the number a rider reads
Compliant	lactate excess below 1 and heart-rate excess below 10	pass or fail over both limits
Lactate excess	$\int \max(0, L - L_{lim})\,\mathrm{d}t$	area above a moving limit
Above HR limit	time with $HR > HR_{lim}$	duration of the excursion
W' empty	time with $W' < 0.05\,W'_0$	the actuator's lost range
Torque limited	time with $T_{\max}$ or $F_{\max}$ binding	the gear, not the rider
Peak core	$\max_t T_c$	the limit with no actuator
Glucose minimum	$\min_t c_G$	the circulating store
Glycogen left	$g(T)$	what the ride cost

5 m altitude table, and the fatigue, glycogen, glucose and core temperature states are advanced every 150 s, so the tail integrates through the knee in the fuel factor instead of sampling one side of it. The reserve term R prices the W' the base policy never spends at the speed it would buy, $W'_N/(\partial P/\partial v)$ meters, with $\partial P/\partial v$ close to 53 W per m/s on the level at 33 km/h and 60 W per m/s on a 6 % climb at 11 km/h, and ramps that value to zero over the finishing window, where it is worth nothing. On a route the objective is the negative of the predicted finishing time, written as meters at 5 m/s so that the penalties keep their scale, with the reserve ramped over the last kilometer and set to zero once the line is inside the horizon.

Plans are ranked by their constraint violation before their objective, so a plan with less violation wins however much farther the other one goes. The search starts from six seeds, being the previous plan shifted one block, the applied exertion held, the nominal marker exertion held, the smaller override request held, and two plans that hold the applied exertion for two blocks and ease by 3 and 8 % afterward. It then refines one block at a time over steps of 0.12, 0.04 and 0.012 in exertion, capped at 320 plan evaluations. When no feasible plan exists the applied exertion is limited to the nominal marker exertion, which is survivable by construction. By default the controller knows only the grade underfoot and holds it over the horizon. A teaching setting hands it the whole route profile, and every comparison that uses it says so. On the circuit hour a replan costs 57.7 plan evaluations on average and 114 at most, and over two repeated runs the mean solve time is 5.1 to 5.2 ms with a maximum of 20 ms in Node.js 26 on an Arm Cortex-X925 core shared with other work, against a replan period of 5 s.

7.6 Performance Measures

Besides the error integrals, peak errors, travel and saturation time that the framework of Section 1.6.4 keeps for each loop, the scorecard carries the ten measures of Table 7.6. They are competing costs and are never combined into a weighted score.

A ride counts as compliant when the areas above both limits stay under small thresholds. Riding on a limit on purpose then does not fail for one sample over it, and a large excursion still fails when it is brief. The integral absolute error of each loop accrues from the standing start and includes the time a loop spends unselected, so it is quoted here only between runs of identical length. Travel sums the scan-to-scan changes of the commanded exertion, $\sum_k |u_k - u_{k-1}|$, in percentage points, as the exported ride file records it. The loop scorecard's own travel figure is taken from the logged exertion after the neuromuscular lag and reads lower, 549 points against 713 for the 190 W hour of Table 7.8. Travel separates a controller that pays for its distance with continual changes of effort from one that does not, and with the power meter's noise it also separates a loop that reads a noisy measurement from one that reads a smooth one.

Table 7.7 Scenarios and Completion Requirements

#	Scenario	Behavior isolated	Required to complete
1	A standing start *Experiment*	force before power	Ride through the step to 50 % at 30 s, then hold 50 % in manual for 2 min.
2	Exertion steps *Experiment*	a step that fades	Ride all five scripted steps, holding 85 % until W' is below 10 % of W'_0.
3	Three minutes all-out *Experiment*	ceiling decays onto *CP*	Hold 100 % to the end of the effort at 9 min, then ride 6 min at 30 % in manual.
4	A steady climb *Experiment*	grade in the road only	Hold 60 % in manual for 9 min of the 6 % climb, then ride on to level road.
5	Chasing the flat speed *Demonstration*	the cost of chasing speed	Keep the scripted 95 % on the 6 % climb until W' is below 5 %.
6	Descending *Experiment*	freewheel and braking	Coast on −8 % until the brakes engage and heart rate is below 80 bpm.
7	A headwind *Experiment*	drag in the air speed	Ride 2 min at 65 % into the wind, then 80 % or more until W' is below half.
8	Repeated surges *Experiment*	slow refill of W'	Ride all four surges and recoveries, then 1 min more at 55 % in manual.
9	Gears and cadence *Experiment*	efficiency, spinning out	Ride 60 % in the highest gear, the lowest past 11 min, then 1 min on automatic.
10	Too big a gear *Demonstration*	a torque limit of the gear	Stall in the highest gear before 4 min, then restart above 8 km/h after the shift.
11	Hot climb, tailwind *Experiment*	cooling by apparent wind	Finish the hour with core temperature above 39.0 °C, then Reset, set a 3 to 4 m/s headwind and finish below 38.5 °C.
12	Under-fueled, 90 min *Experiment*	a store that lowers *CP*	Hold 66 % until blood glucose is below 2.5 mmol/L and on to 90 min.
13	The same ride, fed *Experiment*	delayed glucose	See glucose above 5.5 mmol/L after the 30 min feed, then finish 90 min at 66 %.
14	Three riders *Experiment*	effort is relative	Ride 30 min at 70 % as each preset in turn, pressing Reset between them.
15	Target above *CP* *Demonstration*	saturation on a store	Under the PID selector, saturate the request between 5 and 12 min, then drop below 90 % within 1 min of the 200 W target.
16	Lactate pacing *Demonstration*	a slow override	At 600 W, let lactate peak above 5.5 mmol/L, then Reset, set 210 to 215 W within 1 min and stay compliant to 30 min.
17	Heart-rate dropout *Experiment*	fallback and re-entry	Reach the fallback cap during the dropout, then ride 1 min under PID after it.
18	Fingerstick lactate *Experiment*	a sampled, late reading	Finish the hour on the fingerstick, then Reset, switch to the estimate within 1 min and finish again.
19	10 km time trial *Experiment*	a fixed distance	Summit at a steady 72 % with W' above 90 %, then Reset, start at 88 to 92 %, return to 72 % and summit with W' below 30 %.
C	Hour on the circuit *Challenge*	the economic problem	Over 28.8 km in the hour and compliant, on the case as given.[a]

[a] With the shipped 190 W power target the hour is 27.93 km, compliant but 0.87 km short of the target. With the suggested 215 W target it is 28.90 km, compliant and 0.10 km past it, whether the target is chosen before Run or carried through Reset.

7.7 Results

Table 7.7 lists the twenty cases in the order the application presents them, from open-loop mechanics to closed-loop pacing, with the behavior each one isolates and what completes it. Each case sets the mode, the controller, the rider, the tuning, the road and the timed events in one step and opens with a short briefing on its purpose, and credit follows the completion rules of Section 1.6.4. The challenge is judged on the case as given, so a changed model parameter or disturbance voids the attempt.

7.7.1 Open-Loop Behavior at a Held Exertion

Scenario 1 separates the time scales before any controller is involved. At 50 % exertion from a standstill the standing torque is half of T_0, the rider reaches 17.9 km/h ten seconds after setting

off and 26.8 km/h at thirty seconds, and settles at 30.1 km/h on 146.7 W. Scenario 2 then steps the exertion to 30, 55 and 70 %, which settle at 88, 161 and 205 W, and finally to 85 %, which does not settle. The fourth step buys 491 W five seconds after it lands, 320 W a minute later, 251 W after two minutes and 217 W after four, by which time W' is down to 0.2 kJ and fatigue has cut critical power to 214 W. Equation (7.18) predicts 318 W at one minute and 253 W at two for a constant critical power. Scenario 3 is the all-out test itself, at 839, 537, 271 and 219 W at 5, 20, 60 and 180 s, and afterward at 30 % exertion W' recovers to 12.3 kJ in six minutes and 14.9 kJ in eleven. Scenario 8 repeats a one-minute surge to 95 % four times with two minutes at 55 % between them. The surges peak at 736, 369, 332 and 326 W, because W' recovers only to 4.6, 3.5 and 3.4 kJ between them. The same request buys less each time, because reconstitution below critical power is slower than depletion above it.

The road scenarios separate what the grade does from what the effort does. At a constant 60 %, scenario 4 holds 176.0 W while speed falls from 32.3 to 11.3 km/h on a 6 % climb and cadence falls from 95 to 74 rpm in the lowest gear, and heart rate does not step, drifting 1.3 bpm down because 74 rpm is nearer the most economical cadence. Scenario 5 tries to hold the flat-road speed instead by going to 95 %. It buys 729 W five seconds in, 303 W at 55 s and 216 W after five minutes, and the speed is 13.8 km/h at 200 s against 11.3 km/h for the rider who stayed at 60 %, so 58 % more exertion buys 23 % more watts once W' is gone. Scenario 7 turns a 6 m/s headwind into the rider at 190.7 W, and the speed falls from 33.3 to 21.5 km/h while the aerodynamic force rises from 16.4 to 27.5 N. Scenario 6 coasts down 8 %, where the bicycle reaches the 57.6 km/h braking speed 14.5 s after the rider stops pedaling and the brakes absorb 196 W, 64.8 kJ over the descent, while heart rate falls from 142 to 61 bpm.

The gearing scenarios put an actuator limit outside the rider. Scenario 9 holds 60 % in the highest gear, where cadence falls from 95 to 56 rpm, torque rises from 17.6 to 29.9 N m for the same 176 W, efficiency rises from 22.8 to 23.5 % and heart rate settles 2 bpm lower. In the lowest gear at 32 km/h the legs would have to turn at 211 rpm, the taper holds the rider at 148 rpm, and the power falls to 73 W. Scenario 10 rolls onto 8 % at 15 km/h in the highest gear at 70 %. Within five seconds the rider is torque limited at 97 N m and delivers 184 W of the 204 W asked for, the power falls with the cadence to 84 and 21 W at 30 and 60 s, and the bicycle stops 77 s after the start. After a shift to the lowest gear the rider restarts and settles at 10.2 km/h on 205 W. No exertion setting fixes a gear that is wrong for the grade.

The slow stores then move the actuator itself. Scenario 11 climbs 6 % for an hour at 72 % in 35 °C air at 70 % humidity with a 3.5 m/s tailwind that nearly matches the rider's speed. Evaporation reaches its ceiling after 35 min, core temperature reaches 39.2 °C, thermal fatigue takes critical power from 220 to 206 W, and at the same exertion the power falls from 211 W at ten minutes to 198 W at sixty. Riding the same climb into a 3.5 m/s headwind instead, the rider finishes at 38.2 °C with the full 211 W, slower on the road and far cooler. Scenario 12 rides 90 min at 66 % from a third of the usual glycogen. The liver is down to 1.4 g of its 24 g by thirty minutes, glucose reaches its 2.25 mmol/L floor at forty, critical power falls to 173 W at sixty minutes and 160 W at ninety, and the power at the same exertion falls from 194 to 141 W. Scenario 13 adds 40 g of carbohydrate at thirty and fifty minutes. Glucose rises from 4.1 to 6.1 mmol/L in five minutes and the rider produces 176 W at sixty minutes against 153 W unfed, but by ninety minutes muscle glycogen is down to 7 g and the power is 135 W, slightly below the unfed rider's 141 W. Feeding moved the collapse and did not remove it. Scenario 14 holds 70 % for the three presets, which produce 140, 205 and 294 W at thirty minutes, a factor of 2.1, while heart rate sits at 83, 85 and 86 % of each rider's maximum. Per watt the three hearts differ by a factor of 1.6, while at the same exertion they use 74, 77 and 81 % of their heart-rate reserves, which is what makes exertion a better-conditioned actuator than power across riders.

7.7.2 Saturation and Slow Overrides

Scenario 15, captured in Figure 7.4, asks the power loop for 245 W, 25 W above critical power, with both overrides set out of reach. The loop holds a mean of 243 W between five and ten minutes, and

Table 7.8 Measured Scorecard for One Hour on the Rolling Circuit, Recreational Preset. Hindsight Is the Best Constant Exertion Found by Bisection Over Whole Hours, Which No Rider Could Know in Advance.

	PID 190 W	PID 215 W	PID 600 W	Hind-sight	MPC grade	MPC route
Distance (km)	27.93	28.90	28.80	29.00	29.05	28.97
Compliant	yes	yes	no	yes	yes	yes
Lactate excess	0	0	2292	0.96	0	0
Peak lactate	2.74	3.91	6.45	4.02	5.69	5.75
Peak HR (bpm)	150	159	171	160	173	173
W' left (kJ)	16.0	16.0	6.1	16.0	11.4	12.2
Average power (W)	188	206	205	209	210	208
Travel (points)	713	154	190	0	838	1611
Laps at (s)	—	1643	—	—	1630	1641
	—	3268	—	—	3251	3272

Lactate in mmol/L and mmol s/L. Travel is the total variation of the exertion command in percentage points. Laps are the times at which 13.2 and 26.4 km are passed, to the nearest second.

the exertion it needs rises because every minute the same exertion buys fewer watts. At 688 s the exertion command reaches its stop and the power falls to 212 W, within 4 W of a critical power that fatigue has lowered to 208 W. The integral is not allowed to wind up against the stop, so when the target drops to 200 W at twelve minutes the exertion is at 89.4 % six seconds later and 80.7 % after a minute. The loop has not failed. The actuator has.

Scenario 16 shows how slow an override is. With the power target out of reach the power loop leads from the start and exertion reaches 92 % in the first minute. The lactate loop takes over at 68 s, by which time W' is down to 7.5 kJ, and lactate, four minutes behind the effort, keeps rising until 149 s, to 6.24 mmol/L, while W' falls below 5 % at 135 s. The level hour covers 34.15 km and is not compliant, with 2646 mmol s/L above the limit and W' below 5 % for 2636 s. With a 215 W target the power loop leads until lactate reaches its 3.7 mmol/L target, the lactate loop takes over at 445 s, and the hour is 34.09 km and compliant, 57 m shorter and inside every limit. The target gives this one run whether it is chosen before Run or carried through Reset. Scenario 18 runs that 215 W hour with a fingerstick lactate. The sampled reading lets lactate reach 4.03 mmol/L, 3.5 mmol s/L over the limit and non-compliant, for 34.10 km, against 34.09 km, a peak of 3.88 mmol/L and compliance on the continuous estimate. The stale reading bought 8 m and a violation.

🕮 **Pause.** Try Exercise 7.2 before reading on.

Scenario 17 loses the heart-rate signal for five minutes at ten minutes, with the limit set to 158 bpm so that it binds, in turn with the lactate loop, from three to six minutes. At ten minutes the lactate loop holds the effort, with heart rate at 155.9 bpm. While the signal is missing the fallback cap holds exertion at 64.4 % and 189 W for the whole five minutes, below every live loop's request, and the heart rate the controller cannot see drifts from 156 to 148 bpm under the frozen 156 bpm. When the signal returns the selector is re-entered from the applied 64.4 %, and the power loop ramps it back toward its target at the slew rate, to 66.2 % after four seconds, until the lactate loop takes over again at sixteen minutes.

7.7.3 The Level Hour and the Rolling Circuit

Table 7.8 reports the scorecard for the challenge, one hour on the rolling circuit, under six ways of driving the same rider, and Figure 7.3 draws three of its rows. The predictive controller holds lactate within 0.07 mmol/L of the 3.75 mmol/L it plans against through the middle of the hour and covers 29.05 km, 151 m more than the suggested regulator with every limit respected. The regulator's hour is spent 87 % under the lactate loop and 13 % under the power loop. Both clear the challenge's 28.8 km, and the loaded 190 W target does not.

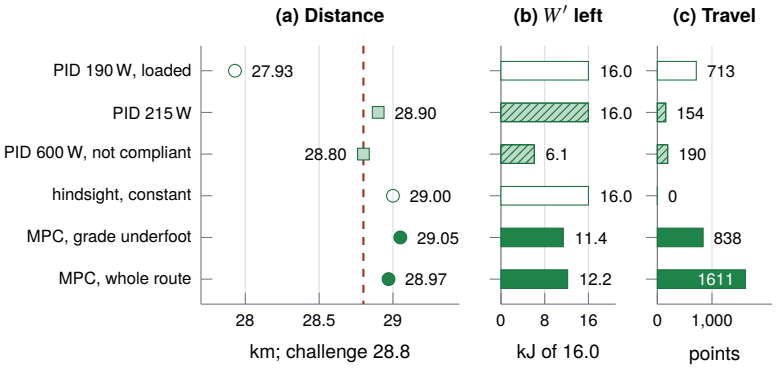


Figure 7.3 The challenge hour of Table 7.8 for its six columns: distance as dots against the challenge's 28.8 km (dashed), and the W' left at the end and the exertion travel as bars. Open marks and white bars are the loaded 190 W target and the hindsight constant exertion, squares and hatched bars the two other PID targets, and the solid ones the predictive controller. Notice that the two predictive columns go furthest while leaving work capacity unspent, and that handing the predictive controller the whole route nearly doubles its travel and shortens its hour.

Table 7.9 Where the Predictive Controller's Advantage Over the Regulator Is Accumulated on the Level Hour. Identical Rider and Limits, and Both Compliant.

Elapsed	Regulator (m)	Predictive (m)	Gap (m)
1 min	320	503	183
3 min	1454	1675	221
10 min	5502	5691	189
30 min	16951	17123	173
55 min	31232	31410	179
60 min	34088	34366	279

The columns of Table 7.8 separate four effects. The loaded 190 W target is compliant and 0.97 km short, and its travel of 713 points is noise, because the power loop reads a meter with 4 W of noise and is selected for the whole hour. The 600 W target reproduces scenario 16 on the hills, and it is 101 m shorter than the compliant 215 W hour despite breaking the limit for most of it, because lactate above the limit accumulates fatigue, and by the end of the hour critical power is down to 201 W against an unchanged 220 W for the compliant regulator. The hindsight column covers 28.999 km at a constant 71.34 %, which is 102 m more than the regulator and 48 m less than the predictive controller, so most of what separates the automatic modes is how quickly each reaches the right exertion and what each does at the ends of the hour. Handing the predictive controller the whole route makes it worse, 28.97 km against 29.05 km, and nearly doubles its exertion travel. A plan that eases before a climb to bank lactate headroom finds the headroom gone four minutes later, which is what Section 7.4 predicts for a limit that is a slow filter on effort, and the route-aware plan is 2 s ahead of the regulator at the first lap and 4 s behind at the second.

🕮 **Pause.** Try Exercise 7.3 before reading on.

On the circuit the gap between two riders at a fixed time depends on where each is on the lap, so Table 7.9 decomposes the predictive controller's advantage on the level hour instead, against the compliant 215 W regulator.

About two thirds of the 279 m exists after one minute. The predictive controller asks for 100 % exertion from the start, reaches it at the slew limit after 24 s, makes 737 W and 43.7 km/h at thirty seconds and spends 7.1 kJ of W' while lactate is still far below its limit, and from ten minutes to fifty

Table 7.10 The 10 km Climb, 518 m of Ascent, From a Standstill. Recreational Preset.

Strategy	Time	Compliant	Peak L (mmol/L)	W' left (kJ)
Steady 72 %	40:22	no	4.15	16.0
90 % for 5 min, then 72	40:20	no	6.17	6.9
85 % for 5 min, then 72	40:39	no	6.80	6.9
PID at 215 W	40:58	yes	3.83	16.0
MPC, grade underfoot	40:14	yes	5.65	10.5
MPC, whole route	40:08.8	yes	5.69	9.6

its exertion is within 1.1 points of the regulator's. The last five minutes add 100 m as the limits open. Removing the finishing allowance from the same controller gives 34.27 km on the level and 28.97 km on the circuit, so the finish is worth 94 and 82 m. Even so, the predictive controller crosses the line with 11.2 kJ of W' on the level and 11.4 kJ on the circuit, because lactate reaches the opening limit before W' runs out. The binding store at the finish is lactate headroom, not work capacity.

> ✦ DID YOU KNOW? **The Highest Peaks With No Excess**
>
> From the simulator: on the rolling circuit the predictive controller's lactate peaks at 5.69 mmol/L and its heart rate at 173 bpm, against 3.91 and 159 bpm for the regulator at 215 W, yet neither hour records lactate excess (Table 7.8). Both limits open over the last five minutes of the hour (Equation (7.17)).

7.7.4 The Time Trial

Scenario 19 changes the objective to time over a fixed distance, the 10 km climb, and Table 7.10 reports six ways of riding it.

A hard start saves 2.8 s at 90 % and loses 16.5 s at 85 %, and both starts break the lactate limit for the rest of the climb. The longer excursion of the 85 % start leaves more fatigue, with critical power at the summit at 209.9 W against 212.4 W after the 90 % start and 219.1 W after the steady ride. The steady 72 % also breaks it narrowly, at 71.4 mmol s/L, so the fastest compliant strategy by hand lies slightly below it. The predictive controller is 44.25 s faster than the regulator and compliant, and here the route profile helps by 5.0 s. On a fixed-distance objective the slow pitches carry more of the time, so a watt moved from the gentler pitches to the steeper ones is worth something, which is the mechanism behind the variable-pacing predictions [10], and the gain is small because the pitches differ by only 3 to 7 %.

7.7.5 What Each Selector Rule Is Worth

Table 7.11 removes the rules of (7.22) one at a time from a copy of the controller and measures the case each rule exists for.

The band row is the design of Chapter 6, in which a losing loop is only prevented from sitting more than a fixed band above the selection, which suits two slow loops. On the bicycle an automatic start breaks it within four scans. At the first scan the applied exertion is zero, so the band pins both overrides near 5 %, and from there they move only by their own slow increments. The power loop leads from rest, and at the fourth scan its request of 7.8 % passes the lactate loop's 5.0 %, so the lactate loop has the smallest request and is selected. From then on the power loop is itself a loser held inside the band, and the exertion rises only as fast as the lactate loop's 420 s integral moves it, to 5.3 % at twenty seconds while lactate sits at rest. The negative-margin row reproduces the ratchet the non-negative rule was written to prevent. The conditional-integration row is recorded because the rule buys nothing on the shipped cases, and without it the circuit hour is 3 m longer and the time trial 2.5 s shorter. Its purpose is to keep the selected loop from accumulating error that the rate limit

Table 7.11 The Selector With One Rule Removed at a Time. Each Row Is Measured on the Case the Rule Exists For, Against the Shipped Controller.

Rule removed	Measured effect
Margin scaled to the limit (losers held within a 0.05 band instead)	circuit hour 18.50 km against 28.90 km. The lactate loop is selected 99.9 % of the time, and exertion is 6.1 % at one minute and 39.6 % at thirty.
Margin never negative	power target 250 W and heart-rate limit 160 bpm on the circuit. Exertion falls from 76.2 to 46.2 % between 160 and 200 s with heart rate still 4.3 bpm over its limit, and the hour is 28.65 km against 28.91 km.
Leader chosen by margin on entry	return from the dropout of scenario 17. Exertion jumps from 64.4 to 74.9 % in four seconds at the slew limit, against a ramp to 66.2 % in the same time.
Conditional integration	circuit hour 28.900 km against 28.896 km, and the time trial under the power loop 2.5 s faster. The rate limit binds only in the first seconds.

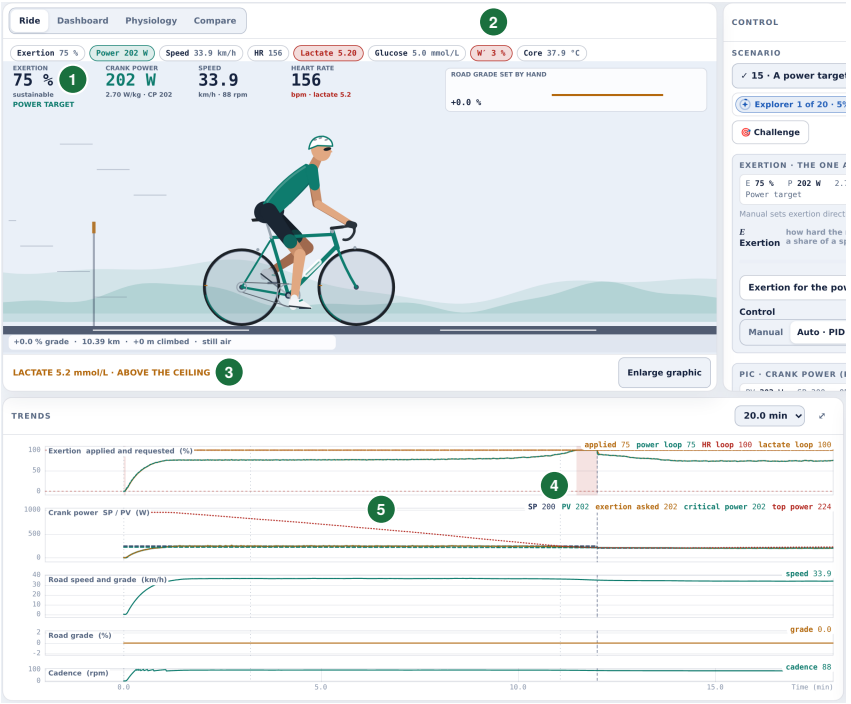


Figure 7.4 Scenario 15 at $t = 18$ min, a 245 W power target above critical power with both overrides out of reach, cropped to the Ride view and the trends. The exertion trace climbs while the power holds its target, reaches its stop at 11.5 min in the shaded saturation band, and the power then falls onto the dashed critical-power line while the dotted ceiling P_{top} decays from 950 W with the work capacity. After the target drops to 200 W at twelve minutes the exertion comes straight off the stop, and at eighteen minutes the rider holds 202 W at 75 % with W' at 3 %, a ceiling of 224 W and a critical power that fatigue has lowered to 202 W. The run records the ride as non-compliant, because with the overrides out of reach lactate peaked at 7.8 mmol/L at eleven minutes and is still 5.2 mmol/L, 2095 mmol s/L above the limit so far. Callouts: (1) the applied exertion with the power loop selected, (2) the W' left, (3) the status line on lactate, (4) the exertion command on its stop in the shaded saturation band, (5) the dotted ceiling decaying onto the dashed critical-power line.

prevented the actuator from answering, and on this plant that happens only in the first seconds of an automatic start.

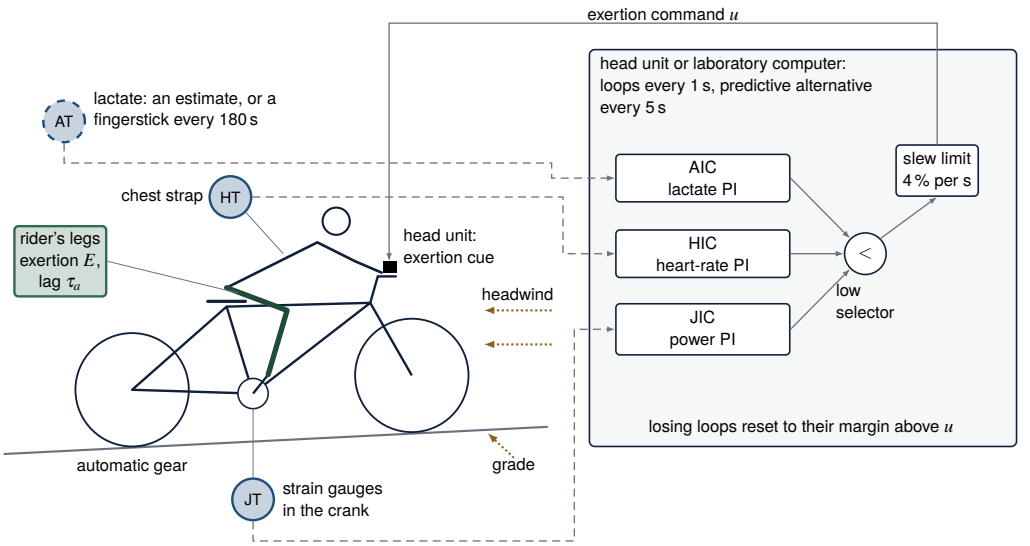


Figure 7.5 The selector loop of Section 7.5 with the equipment a rider carries in place of the schematic's symbols. Strain gauges in the crank (JT) report crank power and a chest strap (HT) heart rate, while the lactate signal (AT, dashed) is an estimate or a sampled blood reading. Three PI loops (JIC, HIC, AIC) meet in a low selector, and the slew-limited exertion command u goes to the head unit as a cue. Notice that the final element is the rider's legs, drawn solid, and that no instrument in this loop reads the work capacity that sets what the legs can deliver.

7.8 Technology in Practice

The power meter, the heart-rate monitor, the lactate reading and the exertion command of Figure 7.1 stand for the equipment a rider carries and a laboratory uses. Figure 7.5 draws the selector loop of Section 7.5 with that equipment in place of the schematic's symbols.

What Measures It

Crank power comes from strain gauges in the drivetrain, read by a head unit on the handlebar [24]. The chapter's power meter, with 4 W of colored noise and a 3 s filter, stands for those gauges and for the averaging of the two impulses of each crank revolution. Heart rate comes from a chest strap, as in Chapter 6, and the held reading of scenario 17 stands for a strap that stops reporting. Lactate has no continuous sensor. The chapter's fingerstick, drawn every 180 s and read 45 s later, stands for a blood sample, and its default lactate signal is an estimate.

> **TECHNOLOGY Sensor: The Strain-Gauge Power Meter**
>
> A cycling power meter measures pedaling force with strain gauges, most often in the crank spindle, a crank arm or a pedal spindle, and a head unit on the handlebar records the power in real time on the road. Each crank revolution delivers two impulses, one per leg, so a reported power is a mean over many of them. Gauges at different positions read differently, because the drivetrain loses some energy between them. In tests cited by a 2022 review, power meters averaged a trueness of -0.9 ± 3.2 %, and some units deviated by more than 5 %. The review recommends calibration and a day-to-day zero offset before the data are used [24].

What Moves It

The final element is the rider. The exertion command reaches the legs through the neuromuscular lag τ_a of Equation (7.1) and a slew limit of 4 % per second, which stand for a rider following a cue on a

screen. An automatic shifter picks the gear between the legs and the road, and scenario 10 shows that no exertion setting fixes a gear that is wrong for the grade.

> **TECHNOLOGY Final Element: The Rider, Cued by a Screen**
>
> Nothing on a road bicycle turns the pedals but the rider, so a pacing controller can only show a target and let the rider follow it. In a laboratory pilot a virtual coach displayed the optimal power to hold, recomputed once every 100 m and held constant over each interval so that it was comfortable to follow. The rider saw only that display, and the trial time fell by 3 % against a self-paced trial [11]. A heart-rate loop on a computer-controlled ergometer has a machine as its final element instead: the ergometer's own microcontroller adjusts the flywheel load to meet the target work rate it is sent [13].

Where the Controller Runs

On the road the platform is the head unit, a cycle computer that receives power and heart rate over a wireless link. Ten such computers recording one ergometer's power once a second agreed at constant loads and differed significantly in short all-out efforts [25]. In the laboratory a personal computer runs the loop over a serial link to the ergometer [13]. The chapter's three loops and their selector scan once a second, the rate at which those cycle computers recorded, and the predictive controller replans every 5 s, the period of the laboratory heart-rate loop. The selector, the slew limit and the fallback cap form one pacing layer above the rider, who closes the innermost loop.

> **IN PRACTICE Choosing the Sample Interval of a Heart-Rate Loop**
>
> A laboratory heart-rate loop on a cycle ergometer shows how a sample interval is chosen in practice. Its designers followed the guideline that a controller should sample 10 to 30 times faster than its closed-loop bandwidth. A bandwidth of 0.01 Hz gave a range of 0.1 to 0.3 Hz, and the loop ran every 5 s on the mean of five heart-rate samples taken once a second. Cadence was a declared test condition: one series held 70 rpm from a visual display, and another let each rider choose [13].

What Is Modernizing It

Three developments bear on this loop. The first is pacing planned on a model of the rider. An optimal pace, computed in advance by dynamic programming from individually identified models of fatigue and recovery, was read from a table every 100 m of the ride [11]. The second is an estimate of the work capacity left, shown to the rider in real time, which a differential balance model allows once its recovery time constant is known for that rider [17]. The third is a reading of the body's state without a blood sample. The pilot's authors estimated the remaining anaerobic energy open loop and expected that wearable sensors such as a sweat lactate sensor might close that loop [11]. A wearable sensor on the forehead has since tracked sweat lactate continuously through an incremental arm-crank test in 27 healthy young adults [26]. Such a measurement would turn the chapter's work capacity into a soft-sensed variable, as Section 7.10 proposes.

7.9 Limitations

No part of the model is a physiological measurement. The constants in (7.3), (7.11), (7.12) and (7.14) were set so that three invented riders behave plausibly, not fitted to data from any person or group, and while each relation follows published work, none of the parameter sets does. The application has not been validated for planning training, judging fitness or advising a rider.

Exertion is a normalized effort, and its map to power is a modeling choice made to reproduce the all-out test. A rider holding a perceived exertion also lets power fall in anticipation of the effort still to come, which the map represents only through fatigue and heat. The road has no corners, no drafting, no crosswind and no change of riding position, and the gearing is an eleven-step ladder behind one drivetrain efficiency.

The predictive controller returns a local improvement found by a bounded coordinate search over six blocks with a rolled-out tail, with no guarantee of optimality, so Table 7.8 compares two implementations rather than an architecture with its bound. On the circuit the plan that is given the route does worse than the plan that sees only the grade underfoot, and the application reports that result as it is. Giving a controller the route is a teaching privilege no real controller has, and any result obtained with it is marked as such on screen and in the exported file. Where each modeling assumption fails is set out in Section 7.3.11.

7.10 Discussion

Nothing in this structure is specific to the bicycle. Any actuator whose authority is a store that drains with use and refills slowly poses the same three problems, as a battery pack under a peak load, a hydraulic accumulator behind a press, a catalyst bed between regenerations and a heat exchanger between cleanings all do. The vault of reserves that defends a currency peg in Volume 4, Chapter 14 is a store that drains with use in the same way. A controller that reads the shortfall as a process disturbance winds up against a limit that is really a state. The correct response is to recognize that the actuator has run out, hold the integral, and let whichever loop can still act take the decision. The bicycle makes the point easily because its store has a name, a unit and a known recovery law.

Set beside the treadmill of Chapter 6, the bicycle also qualifies the value of preview. The treadmill's incline multiplies the energy cost of every meter, so an approaching hill moves the constraint and a controller that sees it early can prepare. On the bicycle at constant exertion the grade changes the speed and not the constraints, a watt buys nearly the same speed on the flat as on a 6 % climb, and the binding lactate limit averages the effort over about four minutes. Preview therefore costs 82 m on the timed circuit and buys 5.0 s on the fixed-distance climb. Knowing which of the two cases a plant belongs to settles whether a disturbance forecast is worth building before any controller is designed around one.

The equations support two methods beyond those implemented. Written with its constraints, the nineteen-state model can be posed as nonlinear model predictive control in APMonitor [27] or the Gekko Optimization Suite [28], replacing the blocked search of (7.23) with a gradient-based solution over a fully discretized horizon in which W', critical power and the lactate limit are explicit path constraints. Estimation is the more interesting extension. A moving horizon estimator on the same model could infer W' and the present critical power from the power meter and the heart-rate monitor, and so turn the authority the actuator has left, which is the most important and least observable quantity in the plant, into a soft-sensed variable instead of a model state. The predictive controller's estimate already integrates W' from the meter, but under the declared critical power, so a rider stronger than declared drains it in the estimate and not on the road.

> **OPEN CHALLENGE** **Knowing How Much Work Capacity Is Left**
>
> In the chapter the work capacity left is a model state, and the predictive controller integrates its own estimate of it from the power meter. A 2020 review notes that no technique measures it directly. Balance models estimate it from power, and a differential form lets a cyclist watch the estimate in real time, provided its recovery time constant was individually and accurately calculated beforehand. That constant varied widely between riders, refilling slowed after repeated efforts, and a model built on untrained riders underestimated the refill of elite athletes [17]. Progress would be recovery parameters identified for each rider and checked against protocols and riders unlike those that produced them, which is where the review calls for more research.

7.11 Conclusion

This chapter has presented a cycling simulator in which the manipulated variable is exertion and the power it buys is a consequence of the rider's state. The model integrates nineteen states covering the road, the drivetrain and a lumped athlete, with the road-cycling force balance, a torque ceiling from

a torque-cadence line, gross efficiency as a declared function of cadence, and an exertion map in which the power ceiling falls linearly with the work capacity left. The gain from exertion to power is 293 W per unit below the sustainable exertion and 2920 W per unit above it for a fresh recreational rider, and a held exertion above that level settles onto critical power with a time constant of 21.9 s at full effort. The limits trail the effort, with heart rate 82 s and lactate 230 s behind a step. A power loop with heart-rate and lactate overrides on a low selector covers 28.90 km of the rolling circuit in the hour and a receding-horizon controller 29.05 km, both compliant, and on the level hour two thirds of the predictive advantage is earned in the first minute. The selector depends on details that a block diagram omits, and removing the margin scaling costs 10.4 km of the hour. The terrain preview costs 82 m on the timed circuit and buys 5.0 s on the fixed-distance climb, because on a bicycle the grade enters the objective and not the constraints. A fuller application would add an indoor-trainer mode with a virtual road, drafting and position changes in the drag area, measured rider data for identification, estimation of the remaining work capacity from the power and heart-rate signals, and a gradient-based predictive controller. A rider on a bicycle is an accessible plant for learning to control an actuator whose authority is a store.

KEY POINTS

- A rider sets how hard to try, and the watts that effort buys fall when an effort above the sustainable level is held, because a finite work capacity drains.
- The gain from exertion to power is 293 W per unit below the sustainable exertion and 2920 W per unit above it for a fresh recreational rider.
- Heart rate trails the effort by 82 s and lactate by 230 s, so an override on lactate is too slow to protect the work capacity.
- On the rolling circuit the selector covers 28.90 km in the hour and the predictive controller 29.05 km, both compliant. Removing the margin scaling costs 10.4 km.

Where else this applies. One actuator shared among limits by a selector returns in Volume 4, Chapter 4 and Volume 4, Chapter 3, and a store that drains as it is used in Volume 4, Chapter 14.

References

[1] K. J. Åström and T. Hägglund, *Advanced PID Control*. Research Triangle Park, NC: ISA – The Instrumentation, Systems, and Automation Society, 2006.

[2] G. A. V. Borg, "Psychophysical bases of perceived exertion," *Medicine & Science in Sports & Exercise*, vol. 14, no. 5, pp. 377–381, 1982.

[3] A. M. Jones, A. Vanhatalo, M. Burnley, R. H. Morton, and D. C. Poole, "Critical power: Implications for determination of $\dot{V}O_2$max and exercise tolerance," *Medicine & Science in Sports & Exercise*, vol. 42, no. 10, pp. 1876–1890, 2010.

[4] R. H. Morton, "A 3-parameter critical power model," *Ergonomics*, vol. 39, no. 4, pp. 611–619, 1996.

[5] R. Tucker, T. Marle, E. V. Lambert, and T. D. Noakes, "The rate of heat storage mediates an anticipatory reduction in exercise intensity during cycling at a fixed rating of perceived exertion," *The Journal of Physiology*, vol. 574, no. 3, pp. 905–915, 2006.

[6] K. C. Cochrane, T. J. Housh, H. C. Bergstrom, N. D. Jenkins, G. Johnson, R. J. Schmidt, and J. T. Cramer, "Physiological responses during cycle ergometry at a constant perception of effort," *International Journal of Sports Medicine*, vol. 36, no. 6, pp. 466–473, 2015.

[7] M. Ellis, H. Durand, and P. D. Christofides, "A tutorial review of economic model predictive control methods," *Journal of Process Control*, vol. 24, no. 8, pp. 1156–1178, 2014.

[8] P. E. di Prampero, G. Cortili, P. Mognoni, and F. Saibene, "Equation of motion of a cyclist," *Journal of Applied Physiology*, vol. 47, no. 1, pp. 201–206, 1979.

[9] J. C. Martin, D. L. Milliken, J. E. Cobb, K. L. McFadden, and A. R. Coggan, "Validation of a mathematical model for road cycling power," *Journal of Applied Biomechanics*, vol. 14, no. 3, pp. 276–291, 1998.

[10] G. Atkinson, O. Peacock, and L. Passfield, "Variable versus constant power strategies during cycling time-trials: Prediction of time savings using an up-to-date mathematical model," *Journal of Sports Sciences*, vol. 25, no. 9, pp. 1001–1009, 2007.

[11] F. Ashtiani, V. S. M. Sreedhara, A. Vahidi, R. Hutchison, and G. Mocko, "Optimal pacing of a cyclist in a time trial based on individualized models of fatigue and recovery," *IEEE Transactions on Control Systems Technology*, vol. 31, no. 1, pp. 317–332, 2023.

[12] M. Paradiso, S. Pietrosanti, S. Scalzi, P. Tomei, and C. M. Verrelli, "Experimental heart rate regulation in cycle-ergometer exercises," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 1, pp. 135–139, 2013.

[13] K. J. Hunt and C. C. Hurni, "Robust control of heart rate for cycle ergometer exercise," *Medical & Biological Engineering & Computing*, vol. 57, no. 11, pp. 2471–2482, 2019.

[14] A. Vanhatalo, J. H. Doust, and M. Burnley, "Determination of critical power using a 3-min all-out cycling test," *Medicine & Science in Sports & Exercise*, vol. 39, no. 3, pp. 548–555, 2007.

[15] P. F. Skiba and D. C. Clarke, "The W' balance model: Mathematical and methodological considerations," *International Journal of Sports Physiology and Performance*, vol. 16, no. 11, pp. 1561–1572, 2021.

[16] P. L. Valenzuela, L. B. Alejo, A. Montalvo-Pérez, J. Gil-Cabrera, E. Talavera, A. Lucia, and D. Barranco-Gil, "Relationship between critical power and different lactate threshold markers in recreational cyclists," *Frontiers in Physiology*, vol. 12, p. 676484, 2021.

[17] A. Chorley and K. L. Lamb, "The application of critical power, the work capacity above critical power (W′), and its reconstitution: A narrative review of current evidence and implications for cycling training prescription," *Sports*, vol. 8, no. 9, p. 123, 2020.

[18] A. Krogh, "A bicycle ergometer and respiration apparatus for the experimental study of muscular work," *Skandinavisches Archiv für Physiologie*, vol. 30, no. 3, pp. 375–394, 1913.

[19] H. Monod and J. Scherrer, "The work capacity of a synergic muscular group," *Ergonomics*, vol. 8, no. 3, pp. 329–338, 1965.

[20] T. Moritani, A. Nagata, H. A. deVries, and M. Muro, "Critical power as a measure of physical work capacity and anaerobic threshold," *Ergonomics*, vol. 24, no. 5, pp. 339–350, 1981.

[21] S. Dorel, C. A. Hautier, O. Rambaud, D. Rouffet, E. Van Praagh, J.-R. Lacour, and M. Bourdin, "Torque and power-velocity relationships in cycling: Relevance to track sprint performance in world-class cyclists," *International Journal of Sports Medicine*, vol. 26, no. 9, pp. 739–746, 2005.

[22] G. Ettema and H. W. Lorås, "Efficiency in cycling: a review," *European Journal of Applied Physiology*, vol. 106, no. 1, pp. 1–14, 2009.

[23] E. F. Coyle, A. R. Coggan, M. K. Hemmert, and J. L. Ivy, "Muscle glycogen utilization during prolonged strenuous exercise when fed carbohydrate," *Journal of Applied Physiology*, vol. 61, no. 1, pp. 165–172, 1986.

[24] P. Leo, J. Spragg, T. Podlogar, J. S. Lawley, and I. Mujika, "Power profiling and the power-duration relationship in cycling: a narrative review," *European Journal of Applied Physiology*, vol. 122, no. 2, pp. 301–316, 2022.

[25] J.-A. Salas-Montoro, I. Valdivia-Fernández, A. de Rozas, J.-M. Reyes-Sánchez, M. Zabala, and J.-J. Pérez-Díaz, "Do power meter data depend on the device on which they are collected? Comparison of eleven different recordings," *Sensors*, vol. 25, no. 2, p. 295, 2025.

[26] T. Sawada, H. Okawara, K. Minami, S. Narushima, A. Shiratori, Y. Katsumata, M. Nakamura, T. Nagura, and D. Nakashima, "Validation of aerobic threshold assessment using a sweat lactate sensor during arm crank exercise by healthy adults," *Physiological Reports*, vol. 14, no. 13, p. e71002, 2026.

[27] J. D. Hedengren, R. Asgharzadeh Shishavan, K. M. Powell, and T. F. Edgar, "Nonlinear modeling, estimation and predictive control in APMonitor," *Computers & Chemical Engineering*, vol. 70, pp. 133–148, 2014.

[28] L. D. R. Beal, D. C. Hill, R. A. Martin, and J. D. Hedengren, "GEKKO optimization suite," *Processes*, vol. 6, no. 8, p. 106, 2018.

7.12 Exercises

7.1 ○ Concept check · 2 min

In scenario 2 the rider steps from 70 % to 85 % exertion at 15 min on the level and holds it. What does the crank power do over the next four minutes?

(a) It settles at a new steady value, because every exertion buys a fixed share of critical power

(b) It falls to zero once W' is empty, because the rider can no longer turn the cranks

(c) It falls from a peak toward critical power while the exertion trace stays flat

(d) It rises slowly as oxygen uptake and heart rate catch up with the new exertion

7.2 Concept check · 2 min

In scenario 16, with the power target out of reach, the lactate loop takes the selector at 68 s. Lactate still rises until 149 s, and W' falls below 5 % at 135 s. Why does the override not protect W'?

(a) Its integral wound up while it was losing, so it could not pull the exertion down

(b) The slew limit of 4 % per second holds the exertion near its peak for minutes

(c) Its margin went negative, and the reset ratcheted the exertion upward

(d) Lactate trails the effort by minutes, so W' is spent before the reading shows it

7.3 Concept check · 2 min

On the timed rolling circuit, handing the predictive controller the whole route gives 28.97 km in the hour against 29.05 km with the grade underfoot only. Which explanation does the chapter support?

(a) The grade changes the speed and not the physiology, and the binding lactate limit averages the effort over minutes

(b) The route-aware plan is solved less often than the plan that sees only the grade, so it answers each climb later

(c) The route profile makes the controller's copy of the rider less accurate, so its plans violate the limits

(d) The move penalty suppresses the extra changes of exertion that a route-aware plan needs

7.4 Theory and calculation · 6 min

Use the recreational preset, fresh: $CP = 220\,\mathrm{W}$, $E_c = 0.75$, $P_{pk} = 950\,\mathrm{W}$ and $W'_0 = 16\,000\,\mathrm{J}$.

(a) From Equation (7.3), compute the gain $\partial P_E/\partial E$ below and above E_c at full W', and their ratio. *Checkpoint:* 293 W and 2920 W per unit exertion, a ratio of 9.95.

(b) From Equation (7.18), compute τ_W for a held $E = 0.85$, and the power 60 s and 120 s after the step. *Checkpoint:* $\tau_W = 54.8\,\mathrm{s}$, 318 W and 253 W.

(c) For a held $E = 1$, compute τ_W and the time after which the closed form is within 0.2 W of critical power. Say what this means for the length of the all-out test. *Checkpoint:* $\tau_W = 21.9\,\mathrm{s}$, about 180 s.

7.5 Theory and calculation · 6 min

Take the road model of Equation (7.5) with $\rho = 1.20\,\mathrm{kg/m^3}$, $C_dA = 0.32\,\mathrm{m^2}$, $C_{rr} = 0.0045$, $m = 83.5\,\mathrm{kg}$, $\mathrm{g} = 9.81\,\mathrm{m/s^2}$, $\eta_d = 0.975$ and no wind.

(a) At a steady 30.10 km/h on the level, compute the power absorbed by drag and by rolling resistance, and the crank power that supplies both. *Checkpoint:* 112.2 W, 30.8 W, 146.7 W.

(b) Compute $\partial P/\partial v$ of the crank power $(F_a + F_r + mg\sin\theta)\,v/\eta_d$ on the level at 33 km/h and on a 6 % climb at 11.3 km/h. *Checkpoint:* 53.4 W and 59.9 W per m/s.

(c) Convert both to speed per watt and give the percentage by which the climb's value falls short of the level one. What would the same comparison show on a treadmill? *Checkpoint:* 0.0187 m/s and 0.0167 m/s per watt, about 11 %.

7.6 Theory and calculation · 6 min

Use Table 7.5, Equation (7.21) and Equation (7.22) with the fresh recreational preset ($CP_0 = 220\,\mathrm{W}$, $E_c = 0.75$, $P_L = 210\,\mathrm{W}$).

(a) Compute the power-loop gain K_c per watt and the loop gain $K_c\,\partial P/\partial E$ just below and just above E_c. *Checkpoint:* 0.00227 per W, 0.667 and 6.64.

(b) Compute the nominal marker exertion $E_{nom} = E_0(P_L)$, the fallback cap that replaces a lost heart-rate signal, and the power a fresh rider makes at that cap. *Checkpoint:* $E_{nom} = 0.716$, a cap of 64.4 %, 189 W.

(c) At 30 min of the level hour in Table 7.1, take the measurements as $HR = 156.8$ bpm and $P = 206$ W, with the 168 bpm limit and a 215 W power target. The lactate loop is selected. Compute the margins m_j at which the heart-rate loop and the power loop are reset above the applied exertion. *Checkpoint:* 0.747 for the heart-rate loop, 0.020 for the power loop.

7.7 App challenge · 8 min

Load scenario 3, Three minutes all-out, and keep the default seed and every setting it loads (Manual, 50 % exertion, the all-out effort from 6:00 to 9:00, then 30 %). Run to 20:00 at speed 20 and export the CSV file. The time column is `Time (min)`.

(a) Read column `wp_kJ` in the rows at 9.00, 15.00 and 20.00 min, within 0.1 kJ, and columns `power_W` and `cp_W` at 15.00 min, within 1 W. *Checkpoint:* 0.0, 12.3 and 14.9 kJ; 87 W against a critical power of 218 W.

(b) Below critical power Equation (7.8) is linear in W'. Write its time constant for a constant P and CP, and evaluate it with $k_r = 0.5$ and the readings of (a). *Checkpoint:* $\tau = W'_0/[k_r(CP - P)] \approx 244$ s.

(c) Predict W' six and eleven minutes after the effort from $W' = 0$ at 9:00, compare with (a), and compare the recovery with the three minutes that emptied W'. *Checkpoint:* 12.3 kJ and 14.9 kJ.

7.8 App challenge · 7 min

Load scenario 15, A power target above critical power, and keep the default seed and every setting it loads (Auto · PID, power target 245 W dropping to 200 W at 12:00, overrides out of reach). Run to 13:00 at speed 5 and export the CSV file (time column `Time (min)`).

(a) Read the time at which the exertion command first reaches 100 % and the SATURATED flag appears, within 0.1 min, and columns `power_W` and `cp_W` in the last row before 12.00 min, within 1 W. *Checkpoint:* 11.5 min; 212 W against 208 W.

(b) Read column `exertion_pct` 6 s and 60 s after the target drops (rows at 12.10 and 13.00 min), within 0.5 points. Explain from Equation (7.20) why the exertion leaves the stop at once. *Checkpoint:* 89.4 % and 80.7 %.

7.9 App challenge · 9 min

Load the challenge, One hour on the rolling circuit, and keep the default seed and every setting it loads (Auto · PID, power target 190 W). Run each hour to its end at speed 60 and read the Distance, Compliant and Lactate excess tiles. Make each change below, press Reset, then Run.

(a) Run the case as loaded. *Checkpoint:* 27.93 km, compliant.

(b) Set Power target to 215 W. *Checkpoint:* 28.90 km, compliant.

(c) Set Power target to 220 W. Explain the Compliant tile. *Checkpoint:* 28.94 km, not compliant, 14 mmol s/L.

(d) Set Control to Auto · MPC, with the grade underfoot only. *Checkpoint:* 29.05 km, compliant.

Give each distance within 0.01 km. Which runs pass the challenge, and what does (c) say about raising the power target further?

7.10 Application area · 3 min

Scenario 18 paces the level hour on a fingerstick: a blood sample drawn every 180 s and read 45 s later.

(a) How old can the lactate reading be just before a new one arrives, and how does that compare with the lag of blood lactate behind the effort in Table 7.4? *Checkpoint:* 225 s, about the 230 s lag.

(b) The chapter compares the hour on the fingerstick with the hour on the continuous estimate. State the result, and name one change an engineer could make to keep the sampled hour compliant. *Checkpoint:* 8 m more and non-compliant.

7.11 Application area · 4 min

Scenario 11 climbs 6 % for an hour at 72 % exertion in 35 °C air, first with a 3.5 m/s tailwind, at about 13.2 km/h at the end, then into a 3.5 m/s headwind, at about 12.0 km/h. No fan is fitted ($a_f = 0$).

(a) From Equation (7.14), compute the airflow a in each case and the factor $0.45 + 0.55a$ that scales evaporative cooling. *Checkpoint:* $a = 0.14$ and 0.92; factors 0.53 and 0.96.

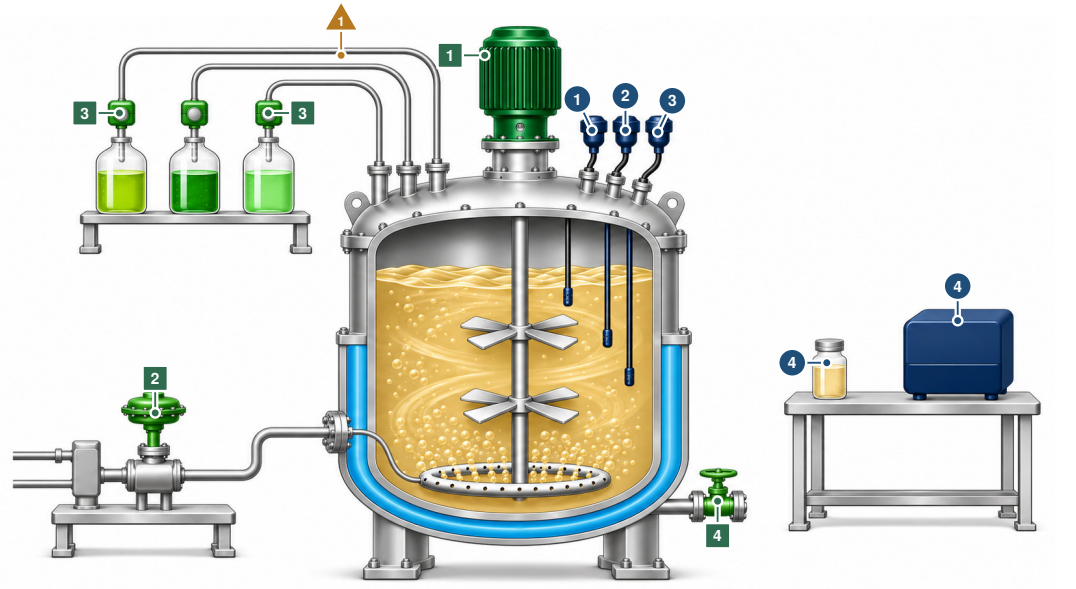
(b) The chapter reports core temperature at the end of each hour. Use the factors to explain the difference, and say what a rider on a stationary trainer indoors needs. *Checkpoint:* 39.2 °C against 38.2 °C.

7.12 Review the agent · 5 min

An AI assistant reviews scenario 15 and writes: “The loop gain of the power loop jumps from 0.667 to 6.64 above E_c, so the loop is too aggressive. It drives the exertion into saturation at 11.5 min and winds up there. Halve the power loop gain to 0.25 and the actuator no longer saturates, so the rider holds 245 W.” Find the flaw, and say how you would check it in the app.

CHAPTER 8

Feeding an Antibody Culture Whose Appetite Outruns Its Daily Sample



The controller moves 1 the agitator, which stirs the broth 2 the gas blend and its flow to the one sparger 3 three feed pumps: glucose stock, nutrient feed and base 4 the jacket valve that sets the temperature

The controller reads 1 dissolved oxygen in the broth 2 the pH of the broth 3 the temperature of the broth 4 the daily sample, drawn to the bench and read four hours later

What disturbs it 1 a weaker glucose lot hung under the same label

Notice that the three probes in the broth answer every scan, while the one sample that sees glucose, cells and product together is read four hours after it is drawn. Illustration, not to scale.

IN PLAIN TERMS **Feeding Cells That Grow Hungrier Every Day**

The main idea. Some medicines are grown by living cells in a tank the size of a small swimming pool. The cells are fed sugar, and they need more of it each day as they multiply.

The challenge. A sample of the broth is taken once a day, and the answer comes back hours later. By then the cells have eaten more, or much less if the broth has been cooled. A feed sized from that answer can starve them or leave sugar they turn into acid.

Why it matters. Starved cells make less medicine, and the acid has to be neutralized with something that stays in the tank and slows the cells down.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	The first page, the history box of Section 8.2, the fact box of Section 8.5, the two try-it boxes and the fact box of Section 8.6 and the key points
Study	70 min	The whole chapter with its pause points and the twelve exercises
Practice	25 min	The first page, Section 8.6, Section 8.7, Section 8.8 and the open challenge after it

The control question An antibody culture's appetite for glucose doubles with its cells, halves when the broth is cooled, and is measured directly only in a sample drawn once a day and reported four hours later. Which feed holds the glucose low but never at zero, and what does each correction of oxygen, pH and glucose leave behind in the broth?

Learning objectives

- Derive the band condition for a daily bolus and decide from measured daily uptakes whether any bolus target can hold the band.
- Explain why a feed-forward on the uptake measured between yesterday's samples overfeeds a culture that has just been cooled, and what a Raman trim or a model-based feed changes.
- Compute the base and the osmolality that lactate held at the pH setpoint costs, and the bound one sparge puts on the carbon dioxide removed.
- Analyze how the glucose target sets the lactate switch, and choose a target, a shift day and an osmolality cap from measured runs.
- Evaluate a predictive plan to the harvest against a classical feed on the same estimate.

Prerequisites Sections 1.4.3 (growth and uptake), 1.4.5 (band condition and bias update), 1.6.2 (velocity form) and 1.6.4 (scenarios and the challenge). Section 1.6.3 is needed only for the predictive controller.

The app https://apmonitor.com/apps/mab-bioreactor/

See also Volume 5, Chapter 4 for a model carried between a late analyzer and a later laboratory, and Volume 5, Chapter 10 for a soft sensor whose bias a delayed laboratory corrects. Volume 9, Chapter 4 for offset-free predictive control bought in the estimator, and Volume 9, Chapter 9 for control against a band rather than a setpoint.

8.1 Introduction

Therapeutic monoclonal antibodies are manufactured by cell culture processes [1], and the class has grown from its first commercial product in 1986 to forty-seven products approved in the United States or Europe by November 2014, with combined sales projected near $125 billion for 2020 [2]. A chemically defined fed-batch of Chinese hamster ovary (CHO) cells has reached 10 g/L of product on day 18 [3]. Such a culture is a feeding problem. Its glucose has to stay low but never reach zero against an appetite that grows with every doubling of the cells, while a sample taken once a day is the only measurement that sees the cells, the metabolites and the product together.

The culture of this chapter starts as 1300 L of medium seeded at half a million cells per milliliter. It holds 8.87 million by the temperature shift on day 5, and the glucose it consumes in one day rises from 0.53 g/L in its first day to 4.27 g/L between days 4 and 5, while the morning sample is reported four hours after it is drawn. What the culture corrects is the habit of feeding toward a comfortable target on a measurement a day old. A daily bolus to 4 g/L, sized from the morning sample, runs the glucose out on day 4.93 and leaves it outside the band of 1 to 6 g/L for 34.8 h. A continuous feed on the cells, built on the uptake measured the day before, overfeeds the culture to 6.09 g/L after the shift, because each cooled cell takes 0.0114 against the 0.0258 g per 10^9 cells per hour the feed assumes. A target of 4.5 g/L, a cautious margin above the floor, keeps the cells from ever switching from making lactate to taking it back, and the base that neutralizes the lactate drives the osmolality to 519 mOsm/kg. That culture is harvested on day 11.17 with 2.58 g/L of antibody, against 5.03 g/L on day 14 when a 2 g/L target is trimmed by an hourly Raman reading.

Feeding in proportion to an estimated demand that grows with the plant's own inventory predates the CHO platform, since a hybridoma fed-batch was fed on its on-line oxygen uptake through the stoichiometry of oxygen and glucose [4], and inferring what a culture does from what can be measured is now a field of bioprocess engineering of its own [5]. A cell therapy expanded in a bag, a yeast propagation fed on its off-gas and an aeration basin dosed to its biomass share the arrangement of a feed-forward on an estimate that drifts, a trim on a faster but indirect measurement, a correction from a slow laboratory, and a side effect of every correction that stays in the vessel. The lager fermenter of

Chapter 10 corrects its diacetyl model in the same way, by a laboratory result that arrives a day after its sample.

Feeding schemes such as glucose delivered on a rising pH [6] or on a Raman estimate [7] were proven on real cultures, whose glucose and lactate are known only when measured, so a student cannot watch a feed-forward go wrong while it happens or run one culture twice with one lever moved. This chapter finds out, on a simulated antibody culture whose every concentration is logged apart from what its probes, estimator and laboratory report, when a daily bolus, a continuous feed on an estimated consumption and a Raman-trimmed feed can hold glucose low but never at zero, and what each correction of oxygen, pH and glucose leaves behind in the broth. The kinetics, the bias update, the split range and the feeding strategies are established methods. So is the model predictive controller, which is measured as a fourth strategy against the shipped feeds and against the strongest classical feed given the same estimate. The culture has 16 states and 19 ledgers whose balances close in a growing volume, a capacitance probe that reads biovolume rather than cells and a controller that sees only its instruments, and the application offers ten cases and a challenge that isolate one decision each.

Section 8.2 traces the domain from the first hybridomas to cultures fed on their own measurements. Section 8.3 gives the model of the culture, its constants and its assumptions, and Section 8.4 derives four consequences of it: the band condition for a daily bolus, the carbon dioxide one sparge can carry, the osmolality each correction costs and the switch that sits inside the band. Section 8.5 sets out the instruments, the four loops, the feeding strategies, the estimator, the supervisor and the predictive controller of Section 8.5.6 as implemented. Section 8.6 reports the ten cases and the challenge, and Section 8.6.11 sets the predictive controller beside every shipped feed. Section 8.7 gives the real instruments, final elements and platform behind the model. Section 8.8 states the limitations and Section 8.9 concludes.

8.2 From Hybridomas to Cultures Fed on Their Own Measurements

The industry begins with a fused cell, and the first licensed antibody product followed in 1986 [1].

> ⧖ HISTORY **An Industry From One Fused Cell (1975 to 1985)**
>
> The industry begins with a fused cell. A report of 1975 described continuous cultures of fused cells that secreted antibody of a predefined specificity [8]. Within a decade such an antibody was treating patients. In a randomized multicenter trial of 123 people whose transplanted kidneys were being rejected, a murine monoclonal antibody reversed 94 % of the rejections against 75 % for high-dose steroids, and graft survival after one year was 62 % against 45 % [9]. Every dose of that antibody had to be grown. The feeding problem of this chapter is what followed the clinical result, and it has been a measurement problem from the start.

Feeding was a measurement problem from the start. In 1995 a hybridoma fed-batch closed its feed on the on-line oxygen uptake rate, with the glucose consumption estimated from the oxygen consumption by stoichiometry, and its metabolism moved from high lactate production and low oxygen consumption to the reverse while its growth phase lengthened [4]. By 2003 temperature had become a lever as well, since lowering a recombinant CHO culture from 37 to 33 °C suppressed its growth, kept its viability high for longer and raised the specific productivity of its erythropoietin about four times [10].

Titers then rose with the culture process. Rising demand drove efforts to raise titers in parallel with the construction of large bulk plants, and the higher titers cut production costs [1]. A chemically defined process reached 10 g/L on day 18 with peak densities above 20 million viable cells/mL and a peak oxygen uptake of 3.7 mmol/(L h), verified at the 100 L pilot scale in reactors configured like those of manufacturing [3]. Scale brought its own constraints. A 5000 L bioreactor had a lower oxygen transfer coefficient, a longer mixing time and a lower rate of carbon dioxide removal than vessels of 5 and 20 L, and raising its bottom sparge improved both transfers more efficiently than

Table 8.1 Signals of the Culture by Role, With Their Values at Inoculation

Symbol	Role	Start	Range
DO	CV of DIC-101	SP 40 %	SP 10–80 %
pH	CV of AIC-102	SP 7.00	SP 6.6–7.4
T	CV of TIC-103	SP 37 °C	SP 30–38 °C
$\hat{G}$	CV of FIC-201, the glucose estimate	SP 2.0 g/L	SP 0.5–8 g/L
N_a, e, Q_g	MVs, the DO sequence	73.1 rpm, 0, 15 L/min	40–140, 0–1, 15–100
y_C, F_b	MVs, the pH split range	10.6 %, 0	0–20 %, 0–2 L/h
T_{sup}	MV, jacket supply	37.5 °C	8–45 °C
F_{glc}	MV, glucose stock	0	0–40 L/h
F_{cf}	MV of the supervisor	1.63 L/h from day 3	0–8 % of V_0 a day
G_s	DV, true glucose lot	500 g/L	300–600, event 420
g	DV, capacitance gain drift	0	−5 to 10 % a day
—	DV, pH probe drift	0	±0.1 a day
—	DV, O_2 supply, Raman, chiller	full, in service, on	events
V	state, broth volume	1300 L	to 2000 L
X_v, X_d	states, cells	0.485, 0.015 $\times 10^6$/mL	—
G, L, P	states	6, 0.1, 0 g/L	—
Q, N, A	states	4, 25, 0.5 mM	—
C_O, C_C, B	states	DO 40 %, 2.27, 18 mM	—
Osm_x, T, T_j	states	208.1, 37.0 °C, 37.3 °C	—
d	state, cell diameter	14.0 µm	14.0–15.2 µm
X_{cap}	measured, capacitance	every scan	1.5 % noise
G_R, L_R	measured, Raman	hourly	0.25, 0.2 g/L noise
—	measured, daily sample	days 0, 1, …	4 h late
OUR, CER	measured, off-gas	logged only	2 % noise

raising its power input [11], while a large-scale CHO culture making an antibody-fusion protein reached 150 mmHg of carbon dioxide and 440 mOsm/kg [12].

The instruments came with a regulatory invitation. The process analytical technology initiative of the U.S. Food and Drug Administration encouraged manufacturers to apply innovative ideas to understand their processes, and in 2011 an in-line Raman probe was shown for the first time to predict glutamine, glutamate, glucose, lactate, ammonium and the viable and total cell densities in a mammalian cell culture bioreactor [13]. Radio-frequency impedance, the capacitance measurement, had already been described for the on-line measurement and control of viable cell density in manufacturing [14]. With a direct reading of the broth the feed could close on lactate itself. Raman models with prediction errors of 0.27 g/L for glucose and 0.20 g/L for lactate fed glucose only while lactate was below a setpoint of 4.0 or 2.5 g/L and glucose below 0.5 g/L, which extended the culture from 11 to 13 days and raised its harvest titer by 85 % [7]. A simpler loop read the pH, because cells take lactic acid back once glucose falls below about 1 mM and the pH rises, and glucose delivered on that rise reduced or removed the lactate accumulation at bench scale and at 2500 L and roughly doubled the final titers of eight cell lines [6].

Model-based feeding completes the sequence. A systematic development of predictive mathematical models for animal cell cultures was set out in 2010 [15], the shift of fed-batch cultures to lactate consumption became a subject of its own [16], and a nonlinear model predictive controller has controlled the glucose concentration of a fed-batch mammalian cell bioprocess [17]. One vessel in this simulator holds all of these pieces, beside a record of the true broth that no plant keeps.

8.3 The Culture as Modeled

One well-mixed broth grows from 1300 L at inoculation toward a working volume of 2000 L as the glucose stock, a complex feed and the base go in, as shown in Figure 8.1, and its signals are classified in Table 8.1. Time is in hours. Cells are in millions per milliliter (10^9 per liter), glucose, lactate and antibody in g/L, glutamine, ammonium, the amino acids, dissolved carbon dioxide and bicarbonate in mM, and per-cell rates per 10^9 cells per hour, so that a rate times the viable density is a rate per liter.

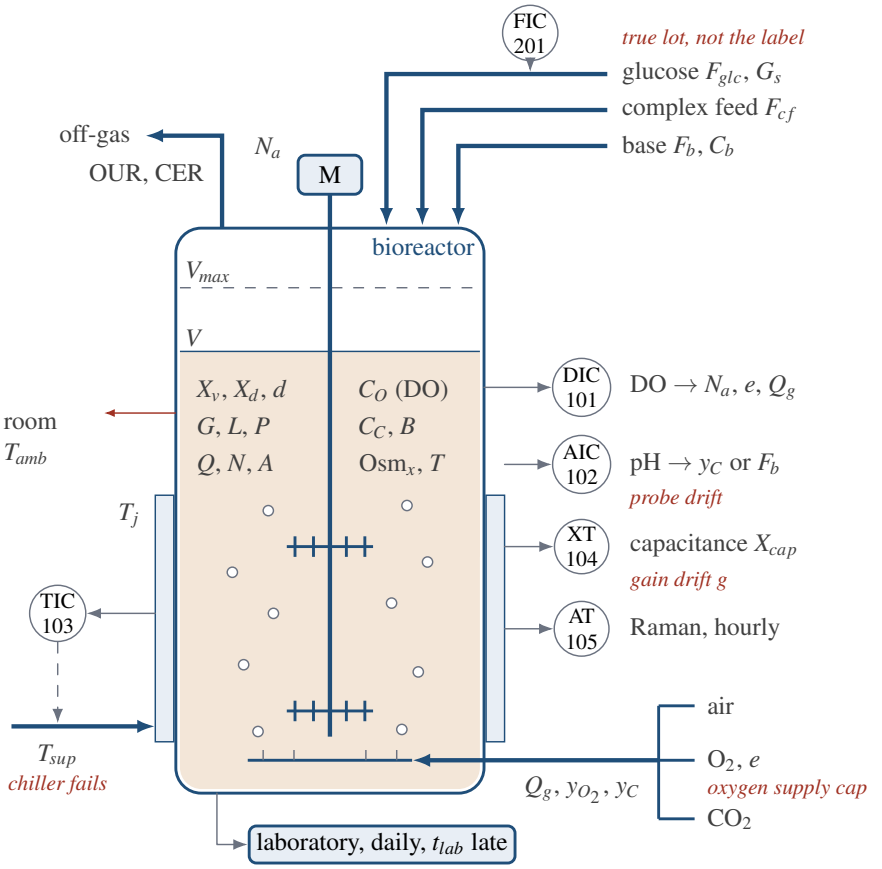


Figure 8.1 The culture, its one sparge, its feeds and what the controller reads. The broth of volume V carries viable and dead cells X_v and X_d of mean diameter d, glucose G, lactate L, antibody P, glutamine Q, amino acids N, ammonium A, dissolved oxygen C_O and carbon dioxide C_C, bicarbonate B, untracked salts Osm_x and temperature T. The agitator turns at N_a, and the sparge carries Q_g of gas whose oxygen fraction y_{O_2} follows the enrichment e and whose CO_2 fraction y_C comes from the pH loop. The feeds are the glucose stock F_{glc} at G_s, the complex feed F_{cf} and the base F_b at C_b, and the jacket holds water at T_j from a supply at T_{sup}. Red italic labels mark the disturbances. Not to scale.

8.3.1 Cells and What Slows Them

Every concentration is diluted by every addition. With the total feed $F = F_{glc} + F_{cf} + F_b$ and the dilution rate $D = F/V$, each species obeys $\mathrm{d}(VC)/\mathrm{d}t = \sum_i F_i C_{i,in} + r_C V$, which is Equation (8.1) for the volume and a concentration.

$$\frac{\mathrm{d}V}{\mathrm{d}t} = F, \quad \frac{\mathrm{d}C}{\mathrm{d}t} = \frac{1}{V}\sum_i F_i C_{i,in} + r_C - DC \tag{8.1}$$

where r_C is the net rate at which the cells make the species. The viable and dead cells follow Equation (8.2).

$$\begin{aligned} \frac{\mathrm{d}X_v}{\mathrm{d}t} &= (\mu - k_d)X_v - DX_v \\ \frac{\mathrm{d}X_d}{\mathrm{d}t} &= k_d X_v - k_{lys}X_d - DX_d \end{aligned} \tag{8.2}$$

The specific growth rate is unstructured Monod growth on glucose, glutamine and one lumped pool of amino acids, in the shape that systematic models of animal cell culture use [15], slowed by every product and condition the controller influences, as shown in Equation (8.3).

$$\begin{aligned}
\mu &= \mu_{max}\, \mathrm{e}^{\theta_\mu (T-37)}\, \frac{G}{K_G+G}\, \frac{Q}{K_Q+Q}\, \frac{N}{K_N+N} \\
&\quad \times \frac{K_L}{K_L+L}\, \frac{K_A}{K_A+A}\, f_{osm}\, f_{pH}\, f_{DO}\, f_{CO_2} \\
f_{osm} &= \frac{1}{1+\left[(\mathrm{Osm}-300)^+/K_{osm}\right]^2} \\
f_{CO_2} &= \frac{1}{1+\left[(p_{CO_2}-60)^+/K_{CO_2}\right]^2} \\
f_{pH} &= \mathrm{e}^{-[(\mathrm{pH}-7)/0.4]^2}, \qquad f_{DO} = \frac{\mathrm{DO}}{0.5+\mathrm{DO}}
\end{aligned} \tag{8.3}$$

where $(x)^+ = \max(0,x)$, T is in °C, DO in percent and $p_{CO_2} = C_C/\alpha$ in mmHg. The osmotic and carbon dioxide factors are teaching values that can be set against a case study, in which raising the osmolality from 316 to 450 mOsm/kg cut the specific growth rate by 60 % and raising p_{CO_2} from 50 to 150 mmHg at a controlled osmolality cut it by 9 % [12]. With the defaults of Table 8.2 the same changes cut it by 46 and 24 %. Death adds a base rate, starvation, ammonium, osmotic stress and a lack of oxygen, as shown in Equation (8.4).

$$\begin{aligned}
k_d &= \mathrm{e}^{\theta_d (T-37)} \Big[k_{d0} + k_{dS}\, s + k_{dA}\, \frac{a^4}{1+a^4} \\
&\quad + k_{dO}\, \frac{(\mathrm{Osm}-380)^+}{50} \Big] + 0.004\, \frac{(5-\mathrm{DO})^+}{5} \\
a &= A/K_{dA} \\
s &= \max\left(\frac{0.03}{0.03+G}, \frac{0.3}{0.3+N}, \frac{0.05}{0.05+Q} \right)
\end{aligned} \tag{8.4}$$

8.3.2 Glucose, Lactate and the Switch

Glucose comes only from the stock pump and is taken by the yield relation, growth plus a maintenance that saturates as glucose runs out, as shown in Equation (8.5).

$$\begin{aligned}
\frac{\mathrm{d}G}{\mathrm{d}t} &= \frac{F_{glc}\, G_s}{V} - q_{glc} X_v - DG \\
q_{glc} &= \frac{\mu}{Y_{xg}} + m_g\, \mathrm{e}^{\theta_m (T-37)}\, \frac{G}{0.05+G}
\end{aligned} \tag{8.5}$$

where G_s is the concentration of the stock actually on the pump, which a feed-lot disturbance can make different from the label the controller assumes. At 37 °C a cell growing at 0.022 h^{-1} takes about 0.6 ng of glucose a day and a resting one about 0.3 ng.

The cells overflow glucose to lactate while they grow fast or while glucose is plentiful, and take lactate back only when both have ended, as shown in Equation (8.6).

$$\begin{aligned}
\frac{\mathrm{d}L}{\mathrm{d}t} &= q_{lac} X_v - DL \\
q_{lac} &= Y_{lg}\, q_{glc}\, \sigma - q_{lc}\, \mathrm{e}^{\theta_m (T-37)}\, \frac{L}{K_{lc}+L}\, (1-\sigma) \\
\sigma &= 1-(1-\sigma_G)(1-\sigma_\mu) \\
\sigma_G &= \frac{1}{1+\mathrm{e}^{-(G-G_{sw})/0.25}}, \quad \sigma_\mu = \frac{1}{1+\mathrm{e}^{-(\mu-\mu_{sw})/0.003}}
\end{aligned} \tag{8.6}$$

The switch is therefore a declared condition, glucose below $G_{sw} = 2.5\,\mathrm{g/L}$ and growth below $\mu_{sw} = 0.015\,\mathrm{h^{-1}}$, standing in for a metabolic shift whose mechanism the literature treats on its own [16]. The application marks the switch at the first instant net lactate production turns into net consumption after lactate has built above 0.3 g/L, provided the glucose is at least 0.5 g/L. A culture whose glucose has run out also takes lactate back, as the only carbon left, and that flip is marked as starvation while the watch for the switch goes on.

8.3.3 Glutamine, Amino Acids, Ammonium and Antibody

The complex feed carries glutamine and the lumped amino acids, the cells take both, glutamine also decomposes in the broth, and ammonium comes from all three, as shown in Equation (8.7).

$$\begin{aligned}
\frac{\mathrm{d}Q}{\mathrm{d}t} &= \frac{F_{cf}Q_{cf}}{V} - q_{gln}X_v - k_{deg}Q - DQ \\
q_{gln} &= \frac{\mu}{Y_{xq}} + m_q\,\mathrm{e}^{\theta_m(T-37)}\frac{Q}{0.1+Q} \\
\frac{\mathrm{d}N}{\mathrm{d}t} &= \frac{F_{cf}N_{cf}}{V} - q_N X_v - DN \\
q_N &= \frac{\mu}{Y_{xN}} + m_N\,\mathrm{e}^{\theta_m(T-37)}\frac{N}{0.2+N} + 8.9\,q_P \\
\frac{\mathrm{d}A}{\mathrm{d}t} &= (Y_{aq}q_{gln} + Y_{aN}q_N)X_v + k_{deg}Q - DA
\end{aligned} \tag{8.7}$$

where $k_{deg} = k_{deg,0}\,\mathrm{e}^{0.1(T-37)}$ and the antibody consumes 8.9 mmol of amino acids per gram, about 1330 residues for a 150 kDa molecule. The antibody is made at a rate that rises as the culture is cooled, as shown in Equation (8.8).

$$\frac{\mathrm{d}P}{\mathrm{d}t} = q_P X_v - DP, \quad q_P = q_{P0}\left[1 + a_P\,(37-T)^{+}\right] f_{DO} \tag{8.8}$$

At a DO of 40 % a cell makes 26.1 pg a day at 37 °C and 46.9 pg at 33 °C, a rise of 1.8 times that is a teaching value below the fourfold rise measured for an erythropoietin line [10].

8.3.4 One Sparge: Oxygen In, Carbon Dioxide Out

Oxygen crosses from the sparge bubbles and through the surface, as shown in Equation (8.9).

$$\begin{aligned}
\frac{\mathrm{d}C_O}{\mathrm{d}t} &= k_L a\left(C^{*}_{air}\frac{y_{O_2}}{0.21} - C_O\right) + k_L a_s\,(C^{*}_{air} - C_O) \\
&\quad - q_{O_2}X_v - DC_O \\
k_L a &= k_L a_{ref}\left(\frac{N_a}{140}\right)^{\alpha_N}\left(\frac{Q_g}{100}\right)^{\beta_Q} \\
y_{O_2} &= (0.21 + 0.79\,e)(1 - y_C) \\
q_{O_2} &= q_{O_2,0}\,\mathrm{e}^{\theta_m(T-37)}\left[0.6 + 0.4\min(1, \mu/\mu_{max})\right] f_{DO}
\end{aligned} \tag{8.9}$$

where $\mathrm{DO} = 100\,C_O/C^{*}_{air}$, N_a is the agitator speed in rpm and Q_g the sparge flow in standard L/min. The sparge $k_L a$ is $0.33\,\mathrm{h^{-1}}$ at 40 rpm and 15 L/min, 3.10 at 140 rpm and $8.0\,\mathrm{h^{-1}}$ at 140 rpm and 100 L/min, and the enrichment e raises the driving force rather than the coefficient. The same bubbles carry carbon dioxide out. They rise through the broth and approach equilibrium with it, so the removal is the gas flow times how nearly the bubbles leave saturated, as shown in Equation (8.10).

$$\begin{aligned}
R_s &= 60\,000\,\dot{n}_g\,\varepsilon\,(y_{eq} - y_C), \quad \varepsilon = 1 - \mathrm{e}^{-\mathrm{NTU}} \\
\mathrm{NTU} &= \frac{k_{CO_2}\,k_L a\,V\,\alpha P}{60\,000\,\dot{n}_g}, \quad y_{eq} = \frac{C_C}{\alpha P}, \quad \dot{n}_g = \frac{Q_g}{22.414}
\end{aligned} \tag{8.10}$$

with R_s in mmol/h, $\dot{n}_g$ in mol/min, $\alpha = 0.0307\,\mathrm{mM/mmHg}$, $P = 760\,\mathrm{mmHg}$, and $k_{CO_2} = 0.89$ is the ratio of the carbon dioxide transfer coefficient to the oxygen one. The headspace adds $k_L a_s k_{CO_2} C_C$ per hour.

8.3.5 Bicarbonate, pH and Osmolality

Respiration makes carbon dioxide at the respiratory quotient, every lactate made spends a bicarbonate and turns it into dissolved carbon dioxide, every lactate taken back and every ammonium made returns one, and each millimole of sodium carbonate adds two bicarbonates while it consumes a carbon dioxide, as shown in Equation (8.11).

$$\begin{aligned} \frac{\mathrm{d}C_C}{\mathrm{d}t} &= \mathrm{RQ}\, q_{O_2} X_v - \frac{R_s}{V} - k_L a_s k_{CO_2} C_C \\ &\quad + r_L - r_A - r_B - D C_C \\ \frac{\mathrm{d}B}{\mathrm{d}t} &= -r_L + r_A + 2 r_B - DB, \quad \mathrm{pH} = 6.1 + \log_{10} \frac{B}{C_C} \end{aligned} \tag{8.11}$$

where $r_L = 11.1\, q_{lac} X_v$ is the lactate made in mM/h, r_A the ammonium made, the first two terms of $\mathrm{d}A/\mathrm{d}t$ in Equation (8.7), and $r_B = 1000\, F_b C_b / V$ the carbonate added. The medium's 18 mM of bicarbonate at pH 7.00 means 2.27 mM of dissolved carbon dioxide, 74 mmHg, at inoculation. Osmolality is the untracked salts plus every tracked solute, as shown in Equation (8.12).

$$\begin{aligned} \mathrm{Osm} &= \mathrm{Osm}_x + 5.55\, G + 11.1\, L + Q + A + N + B \\ \frac{\mathrm{dOsm}_x}{\mathrm{d}t} &= \frac{1}{V}\left[F_{cf}\left(\mathrm{Osm}_{cf} - Q_{cf} - N_{cf}\right) + 2000\, F_b C_b\right] \\ &\quad - D\,\mathrm{Osm}_x \end{aligned} \tag{8.12}$$

Nothing consumes Osm_x. It holds the medium's salts, the complex feed's salts and the base's sodium, and the only way it falls is dilution.

8.3.6 Heat, Jacket and Cell Size

The broth gains the cells' heat, the agitator's power and the room, and loses heat to the jacket, whose water is replaced from the supply, as shown in Equation (8.13).

$$\begin{aligned} \rho c_p V \frac{\mathrm{d}T}{\mathrm{d}t} &= 3.6\left[q_h X_v V + P_{ag}(N_a/140)^3 - UA(T - T_j)\right. \\ &\quad \left. - UA_{amb}(T - T_{amb})\right] - \rho c_p F\,(T - T_{amb}) \\ \frac{\mathrm{d}T_j}{\mathrm{d}t} &= \frac{T_{sup} - T_j}{\tau_j} + \frac{3.6\, UA\,(T - T_j)}{C_j} \end{aligned} \tag{8.13}$$

with every heat term in W, which the factor 3.6 turns into kJ/h, and with the feeds entering at room temperature. The capacitance probe responds to the volume of intact cells, and cells swell as their growth stops. The mean diameter follows a target set by the growth rate over a day, as shown in Equation (8.14).

$$\begin{aligned} \frac{\mathrm{d}d}{\mathrm{d}t} &= \frac{d^* - d}{\tau_d} \\ d^* &= d_{min} + (d_{max} - d_{min})\left[1 - \min(1, \mu/0.025)\right] \end{aligned} \tag{8.14}$$

8.3.7 Constants, Inoculation and Integration

Table 8.2 lists the constants. At inoculation the broth is the basal medium with the seed, 0.1 g/L of lactate and 0.5 mM of ammonium, its dissolved oxygen at the setpoint and its carbon dioxide at the value that puts the bicarbonate at the pH setpoint, and Osm_x is what makes the total the

Table 8.2 Constants of the Culture. Every Value Above the Last Rule Is a Slider of the Show Model Dialog, None Is Hidden, and the Last Rows Are Fixed in the Code. All Are Teaching Values

Symbol	Meaning	Value	Symbol	Meaning	Value
Vessel, medium and feeds			*Metabolism, with cells in* 10^9 *and rates per* 10^9 *cells per hour*		
V_0, V_{max}	working volume at seeding, most it holds	1300, 2000 L	Y_{xg}, m_g	cells per gram of glucose, maintenance	1.6 per g, 0.012 g
X_0, v_0	seed density (all cells), seed viability	0.5×10^6/mL, 97 %	Y_{lg}	lactate per glucose in overflow	0.45 g/g
G_0, Q_0, N_0	glucose, glutamine, amino acids	6 g/L, 4, 25 mM	q_{lc}, K_{lc}	lactate uptake after the switch, half point	0.0025 g, 0.4 g/L
B_0, Osm_0	bicarbonate, osmolality of the medium	18 mM, 290 mOsm/kg	G_{sw}, μ_{sw}	switch: glucose below, growth below	2.5 g/L, 0.015 h^{-1}
G_s	glucose stock as labeled	500 g/L	Y_{xq}, m_q	cells per mmol of glutamine, maintenance	2.5 per mmol, 0.002 mmol
Q_{cf}, N_{cf}	glutamine, amino acids in the complex feed	60, 330 mM	Y_{xN}, m_N	cells per mmol of amino acids, maintenance	1.0 per mmol, 0.0025 mmol
Osm_{cf}, C_b	complex feed osmolality, Na_2CO_3	900 mOsm/kg, 1 M	Y_{aq}, Y_{aN}	ammonium per glutamine, per amino acid	0.6, 0.04
Growth and death			$k_{deg,0}$	glutamine decomposition at 37 °C	0.003 h^{-1}
μ_{max}	growth rate at 37 °C	0.035 h^{-1}	$q_{O_2,0}$, RQ	oxygen uptake, respiratory quotient	0.25 mmol, 0.95
K_G, K_Q, K_N	half-saturations of growth	0.05 g/L, 0.1, 1.0 mM	q_{P0}	antibody per cell at 37 °C	0.0011 g
K_L, K_A	lactate, ammonium inhibition	15 g/L, 20 mM	*Gas, heat, cell size and laboratory*		
K_{osm}, K_{CO_2}	osmotic, CO_2 scales above 300, 60 mmHg	160, 160	k_La_{ref}	sparge k_La at 140 rpm, 100 L/min	8 h^{-1}
k_{d0}, k_{dS}	base death, death added by starvation	0.0006, 0.015 h^{-1}	α_N, β_Q	exponents on speed and gas flow	1.8, 0.5
k_{dA}, K_{dA}	ammonium death, its half point	0.006 h^{-1}, 10 mM	k_La_s, k_{CO_2}	headspace transfer, CO_2 ratio	0.15 h^{-1}, 0.89
k_{dO}	death per 50 mOsm/kg above 380	0.004 h^{-1}	C^*_{air}	dissolved O_2 at air saturation	0.20 mM
k_{lys}	lysis of dead cells	0.004 h^{-1}	UA, UA_{amb}, T_{amb}	jacket, room loss, room	600, 15 W/K, 22 °C
θ_μ, θ_m, θ_d	e-fold per kelvin: growth, uptakes, death	0.25, 0.09, 0.17 K^{-1}	τ_j, C_j	jacket turnover, jacket heat capacity	0.1 h, 420 kJ/K
a_P	antibody rate per kelvin below 37 °C	0.2 K^{-1}	P_{ag}, q_h	agitator power at 140 rpm, heat per 10^9 cells	250 W, 0.02 W
			d_{min}, d_{max}, τ_d	diameter growing, arrested, response	14.0, 15.2 µm, 24 h
			t_{lab}	sample drawn to reported	4 h
pK, α, P	CO_2 equilibrium, solubility, pressure	6.1, 0.0307, 760	—	switch widths, starvation mark	0.25 g/L, 0.003 h^{-1}, 0.5 g/L
ρc_p	broth heat capacity	4.0 kJ/(L K)	—	amino acids per gram of antibody	8.9 mmol

medium's 290 mOsm/kg, 208.1 mOsm/kg. The jacket water starts at the temperature that removes the vessel's net heat, and the controller starts at the DO demand that holds the setpoint against the seed's uptake, found by bisection, and at the carbon dioxide fraction that balances the starting carbon dioxide. Nineteen ledger states integrate what came in, what was made and what was used, so every balance closes in the growing volume, and in the default culture the worst of eleven balances closes to 1.4×10^{-12} relative after fourteen days.

The engine of Section 1.6.1 integrates the culture by classical Runge-Kutta steps of fourth order and a fixed 10 s, the controller executes once a minute, and the historian records every 5 min, 4032 rows for fourteen days. The speeds run from 0.1 to 48 simulated hours per second, so a whole culture takes 7 s at the top of the ladder. The default seed is 20261008, and Reset replays a culture as that section describes. Halving the step leaves the default culture's titer at 5.0334 g/L, its lactate switch on day 6.3168 and its highest osmolality at 393.231 mOsm/kg, and a step of 30 s gives the same titer and osmolality with the switch on day 6.3170. The application passes its 62 checks, among them the closure of every balance to 10^{-9} through whole cultures, the two conditions of the switch and the four-hour delay of the sample.

8.3.8 Assumptions

One well-mixed broth. A 5000 L vessel mixes more slowly than a bench one and can carry gradients of pH and dissolved oxygen [11], while this broth has none, so it cannot show a local excess of glucose at the feed point or a carbon dioxide gradient with height.

Unstructured kinetics with teaching constants. The constants of Table 8.2 are teaching values anchored in the CHO literature, not a fit to a cell line, and real lines differ several-fold in each. The real values of a process come from its own cell line, through a systematic development of predictive models of the kind set out for animal cell cultures [15], and from measured responses such as the growth and productivity of a fed-batch at scale [3], the specific productivity at a lower temperature [10], and the growth at elevated carbon dioxide and osmolality [12].

A declared switch. Equation (8.6) switches on two thresholds, a declared condition rather than a mechanism of the metabolic shift [16], so it cannot show a line that switches for another reason or one that never switches.

A shift without adaptation. The temperature factors act at once, so a culture at 33 °C grows at 37 % of its rate, dies at 51 %, takes up 70 % as much and makes 1.8 times the antibody per cell from the moment it arrives there. Real cells adapt over a day or two.

A simple buffer and an ideal gas phase. The pH ignores phosphate, the amino acids and the cells' own buffering, carbon dioxide hydrates instantly, the bubbles have no hydrostatic head, and foam, shear and antifoam are absent, which is why raising the gas flow costs nothing here that a plant would pay in foam.

A probe that reads size. The capacitance probe sees biovolume as cells of the fast-growing size, with a diameter that follows the growth rate, while real dielectric spectroscopy also depends on the membranes and the frequency. The Raman has a fixed noise and no calibration drift.

No product quality. Glycosylation, charge variants, aggregates and host-cell proteins are outside the model, although they often decide a real harvest more than the titer does.

8.4 What a Day, a Bubble and a Liter of Base Allow

Four consequences of Section 8.3, none of them an extra assumption, explain most of what the cases measure.

8.4.1 The Bolus and the Band

A bolus sized from a sample restores the glucose the sample had, not the broth's, which has fallen during the t_{lab} of the analysis. The broth therefore starts each interval below the target by what the cells took during the analysis and ends it below the target by what they took from the draw to the next report, and the difference, the consumption over the 24 h between two reports, has to fit inside the band. Once it does not, no target can hold a band $[G_{lo}, G_{hi}]$. Equation (8.15) states the condition, the band test of Equation (1.7) written for this culture's volume.

$$C_k = \frac{1}{\bar{V}} \int_{t_k}^{t_k+24} q_{glc} X_v V \, \mathrm{d}t \; > \; G_{hi} - G_{lo} = 5\,\mathrm{g/L} \tag{8.15}$$

with t_k the time of the k-th report, $\bar{V}$ the mean volume over the interval, and the dilution neglected. Fed continuously so that it eats what it wants, the default culture takes 0.53, 1.02, 1.83, 2.97 and 4.27 g/L in its first five days, 44 to 94 % more each day, and 2.92 g/L in the day after the shift, as shown in Figure 8.2. Kept at 37 °C until day 8 it takes 5.24, 5.84 and 5.93 g/L in days 5 to 8, and no daily bolus can hold the band.

🕮 **Pause.** Try Exercise 8.1 before reading on.

8.4.2 What the Gas Can Carry

Equation (8.10) bounds the carbon dioxide one sparge can remove by its flow, $R_s \le 60000\,\dot{n}_g (y_{eq} - y_C)$, whatever the agitation or the oxygen content. The bound is reached as ε approaches one, and

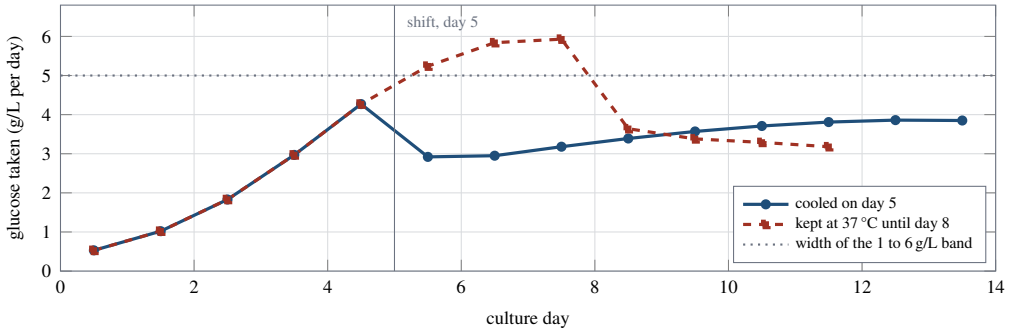


Figure 8.2 Glucose taken in each day of the culture fed continuously with the Raman trim. Cooled on day 5 (solid line, circles) the culture never takes more than 4.27 g/L in a day, and kept at 37 °C until day 8 (dashed line, squares) it takes up to 5.93 g/L, more than the band is wide (dotted line).

in the 1300 L of inoculation ε is 0.196 at 40 rpm and 15 L/min, 0.875 at 140 rpm and 15 L/min and 0.554 at 140 rpm and 100 L/min. Once the agitator is at full speed the bubbles leave the broth 87.5 % saturated, so only more gas removes more carbon dioxide, and the enrichment e, which changes neither $\dot{n}_g$ nor ε, removes none. The DO sequence therefore decides the carbon dioxide the pH loop has to live with, and through it the bicarbonate the base must supply.

🕮 **Pause.** Try Exercise 8.3 before reading on.

8.4.3 The Price of a Correction

While base is the active leg of the pH loop, holding the pH at a dissolved carbon dioxide fixed by the gas means holding the bicarbonate. Each millimole of lactate spends a bicarbonate, and half a millimole of sodium carbonate restores it while the extra carbon dioxide leaves with the gas, so a millimole per kilogram of lactate held at the setpoint costs 2 mOsm/kg, the lactate itself and one sodium, and a gram per liter costs 22 mOsm/kg. Over a whole culture the bicarbonate balance of Equation (8.11) fixes the base exactly, as half of the lactate made, less half of the ammonium made, plus half the change in the broth's bicarbonate. The default culture makes 49.3 mol of lactate net and 21.6 mol of ammonium, its bicarbonate grows from 23.4 to 50.9 mol, and it needs 27.6 mol of base, while case 7 makes 178.1 mol of lactate and 12.6 mol of ammonium, ends with 11.2 mol of bicarbonate, and needs 76.7 mol. The lactate part of the osmolality returns when the cells take lactate back, but the sodium stays, because nothing in Equation (8.12) consumes it. In the default culture Osm_x rises from 208.1 to 289.6 mOsm/kg by day 14, which is the medium's salts diluted to 143.8, the complex feed's salts at 116.3 and the sodium of the base at 29.4, and in case 7 the sodium alone is 86.8 mOsm/kg at harvest.

8.4.4 A Switch Inside the Band

After the shift the growth rate is about $0.006\,\mathrm{h}^{-1}$, far below μ_{sw}, so the glucose condition alone decides the sign of q_{lac} in Equation (8.6). The sign changes where overflow and uptake balance, and in the default culture that happens at glucose between 2.15 and 2.17 g/L, seven times from day 6.32 to day 12.52 as the trimmed feed moves the glucose across it. The useful part of the 1 to 6 g/L band is therefore 1 to about 2.15 g/L, and a target is a lever on the switch as much as on the floor.

8.5 The Controller and What It Reads

The controller executes once a minute and reads no concentration in the broth, only its instruments, its own pump commands at the stock's labeled concentration and the daily laboratory, with the settings of Table 8.3.

Table 8.3 Controller, Estimator and Supervisor Settings

Function	Setting	Default	Range
DIC-101	K_c, τ_I	0.6 % per % DO, 0.5 h	0.05–5, 0.05–5
	output 0–100 %	N_a, then e, then Q_g	or Q_g before e
AIC-102	K_c, τ_I, deadband	150 % per pH, 0.5 h, 0	5–600, 0.05–10, 0–0.2
	output ±100 %	base to 2 L/h above zero, CO_2 to 20 % below	—
	re-standardization	off	$\lambda = 1$, $\lvert b\rvert \le 0.5$
TIC-103	K_c, τ_I	6 K/K, 1.5 h	0.5–30, 0.1–10
FIC-201	strategy, target	continuous, 2.0 g/L	bolus, model-based, MPC, 0.5–8
	trim K_c, τ_I	1 L/h per g/L, 8 h	0–10, 0.5–48
MPC	κ, w_P, S, preview	3, 5, 8, on	preview off
Complex feed	rate, start	3 % of V_0 a day, day 3	0–8, 0–8
Osmolality cap	state, level	off, 390 mOsm/kg	330–460
Shift	rule, T_{low}	day 5, 33 °C at 1 K/h	or at $\hat{X}_v = 10 \times 10^6$/mL
Harvest	rule	day 14 or offline viability below 70 %	day 8–20, 40–95 %
Estimator	λ_X, Raman pull, $\hat{q}_0$	0.5 ($\lvert b\rvert \le 1.2$), 0.35, 0.02 g per 10^9 cells per hour	0–1, on or off

8.5.1 Instruments

The DO, pH and temperature probes lag the truth by 1, 0.5 and 0.5 min with 0.3 %, 0.004 and 0.02 °C of colored noise correlated over 1.2 min, and a drift disturbance walks the pH probe's offset away from the truth. The capacitance probe reads $X_{cap} = X_v (d/d_{min})^3 \, g(t) \, (1+n)$ every scan, with a gain drift g and 1.5 % of noise n correlated over 0.5 h, so it reads high as the cells swell. The Raman reports glucose and lactate every hour, 15 min after its sample, with 0.25 and 0.2 g/L of noise, close to the prediction errors of 0.27 and 0.20 g/L reported for a Raman model of a fed-batch [7]. A sample drawn at every whole day from inoculation is reported $t_{lab} = 4$ h later with the viable density, viability, glucose, lactate, ammonium, osmolality, titer, pH and p_{CO_2}, with noise of 0.04 on the logarithm of the count, 0.8 points of viability, 0.08 g/L of glucose and of lactate, 0.2 mM of ammonium, 3 mOsm/kg, 0.03 on the logarithm of the titer, 0.01 of pH and 3 mmHg. A Raman reading goes stale after three missed hours. Both the Raman and the laboratory run on the framework's shared sampled analyzer, and the off-gas uptake and evolution rates are logged with 2 % of noise and used by no loop.

> ✦ DID YOU KNOW? **A Membrane Five Nanometers Thick**
>
> What the capacitance probe senses is the thin membrane around each cell. In a study of such probes the viable cells of a culture measured 6.75 µm in radius and the dead ones 6.0 µm, around a membrane taken as 5 nm thick, about a thousandth of the cell [18].

8.5.2 Four Loops on the Shared Controller

All four loops use the framework's velocity-form law of Section 1.6.2 without derivative action and with no output bias, written for this chapter as Equation (8.16).

$$\begin{aligned} e_k &= \mathrm{SP}_k - \mathrm{PV}_k, \qquad P_k = K_c \, e_k \\ s_k &= s_{k-1} + (P_k - P_{k-1}), \qquad v_k = s_k + I_{k-1} \\ u_k &= \min\left(u_{max}, \max(u_{min}, v_k)\right) \\ I_k &= I_{k-1} + \left[\frac{K_c}{\tau_I} \, e_k + \frac{u_k - v_k}{\tau_I}\right] \Delta t_c \end{aligned} \tag{8.16}$$

where $\Delta t_c = 1$ min. As in Equations (1.13) to (1.15), the sum s_k is never clamped, back-calculation corrects only the integral at a tracking time equal to τ_I, no rate limit applies, and a loop in MANUAL tracks the operator's output u_{op} through $I_k = u_{op} - s_k$, so the return to AUTO is bumpless. All four loops are reverse acting, with action $a = +1$ in the convention of Section 1.3: more agitation and

oxygen raise the DO, more base raises the pH, a warmer jacket supply raises the temperature and more stock raises the glucose. The application's sign checks confirm each direction. DIC-101 spreads its output over three actuators, the first third taking the agitator from 40 to 140 rpm, the second enriching the sparge from air toward pure oxygen and the last raising the gas from 15 to 100 L/min, or the gas before the oxygen in the alternative sequence, and an oxygen supply disturbance caps the enrichment. The speed such an oxygen loop needs from the agitator beneath it is the subject of Volume 3, Chapter 3. AIC-102 is a split range on an error that can be deadbanded, base above zero and carbon dioxide in the sparge below it, and TIC-103 sets the jacket supply with the supervisor's shift schedule, which ramps at 1 K/h, as its setpoint.

8.5.3 Feeding by Bolus, Rate and Platform Model

The daily bolus acts at each offline result and pumps, at the full 40 L/h, the stock that brings that sample's glucose up to the target, $(G_{SP} - G_{lab})\,V / G_s$ with G_s on the label. The continuous feed pumps the estimated consumption plus a PI trim on the glucose estimate, as shown in Equation (8.17).

$$F_{glc} = \frac{\hat{q}\,\hat{X}_v\,V}{G_s} + \mathrm{PI}\big(G_{SP} - \hat{G}\big), \qquad 0 \le F_{glc} \le 40\,\mathrm{L/h} \tag{8.17}$$

where the trim's limits move every scan so that the sum stays within the pump's range and the back-calculation of Equation (8.16) respects it. The model-based feed replaces $\hat{q}$ by the platform yield relation on the growth the corrected probe sees, $q_{model} = \hat{\mu}/Y_{xg} + m_g\,\mathrm{e}^{\theta_m (T-37)}$ with the default constants whatever the plant has been set to, so it knows a cooled culture eats less, and it scales the complex feed by $\hat{X}_v / (13 \times 10^6/\mathrm{mL})$, capped at 1.4. The fourth strategy, the predictive controller of Section 8.5.6, replaces both the feed-forward and the trim.

8.5.4 The Estimator

Each offline count corrects the capacitance probe through the framework's shared bias update, the logarithmic form of Equation (1.8), applied to the reading at the hour the sample was drawn as in Equation (8.18).

$$b \leftarrow b + \lambda_X \big[\ln X_{lab} - (\ln X_{cap}(t_s) + b)\big], \qquad \hat{X}_v = X_{cap}\,\mathrm{e}^{b} \tag{8.18}$$

The per-cell uptake $\hat{q}$ is the glucose that disappeared between the last two samples, what was there plus what the pumps delivered on the label less what is left, over the viable cell-hours their counts bracket by the trapezoid rule, averaged as $0.4\,\hat{q} + 0.6\,q$ after the first interval. The growth $\hat{\mu}$ is the log slope of $\hat{X}_v$ over the last 12 h plus a 6 h average of the dilution. The glucose estimate integrates $\mathrm{d}\hat{G}/\mathrm{d}t = F_{glc} G_s / V - \hat{q}\hat{X}_v - D\hat{G}$ with G_s on the label, moves 35 % of the way to each fresh Raman reading, and is shifted by each offline result by what it was wrong by at the hour of the draw. The milk powder dryer of Volume 5, Chapter 10 corrects the bias of its soft sensor in the same way, from a sampled laboratory result that arrives late. The osmolality estimate that feeds the cap adds what the pumps deliver, Osm_{cf} for the complex feed, $4000\,C_b$ for the base and $5.55\,G_s$ for the stock, less $5.55\,\hat{q}\hat{X}_v$ for the glucose the cells use, and is shifted by each offline osmolality in the same way. It takes Osm_{cf} and C_b at their declared 900 mOsm/kg and 1 M, and both estimates begin from the declared basal medium, 6 g/L of glucose at 290 mOsm/kg, so a medium or feed altered in the parameters is a mismatch that only the samples reveal.

8.5.5 Supervisor

The supervisor shifts the temperature on the chosen day or when $\hat{X}_v$ first reaches the chosen density, runs the complex feed, and decides the harvest on the day, at the first offline viability below the limit, or whichever comes first, with a viability rule ending the culture on day 20 at the latest. The viability rule judges a decline and acts only on a result below an earlier one, so a seed counted under the limit at inoculation is not harvested on its day-0 count, and seeds of 60 to 70 % viability fed as

in case 4 reach day 14 with 4.33 to 4.58 g/L. The optional osmolality cap scales the complex feed by $\min\{1, \max[0, (\mathrm{Osm}_{cap} - \widehat{\mathrm{Osm}})/12]\}$ and lowers the glucose target in proportion, toward 1.5 g/L at the cap. At every offline result from the second on, a bolus check multiplies the consumption measured over the last interval by the growth of $\hat{X}_v$ since, clamped to 1 to 3 and applied only while the culture is still warm, and flags that no daily bolus can hold the band when the product passes the 5 g/L of Equation (8.15).

8.5.6 A Predictive Controller Over the Rest of the Batch

The last value of the feeding strategy is a model predictive controller (MPC) [19, 20]. It is written for this culture rather than taken from the framework's option of Section 1.6.3, for the reason given below. Once an hour, at minute 15 just after the Raman reports, it decides a glucose dose u in grams per liter of broth per hour and holds it until the next decision, and the pump delivers uV/G_s with G_s on the label, as every strategy does. The DO, pH and jacket loops, the complex feed, the osmolality cap, the vessel-full stop, the shift and the harvest act as under the other strategies, the trim of Equation (8.17) tracks the applied flow, and the continuous feed stays the application's default and the starting strategy of all cases but case 10.

The controller's model is the culture's own kinetics at their declared values, written on the estimator's states rather than identified, because a fed-batch never settles at an operating point that a step test could perturb. With the state $x = [G, L, \mathrm{Osm}, X, P, V, A]$ it reads as shown in Equation (8.19).

$$\begin{aligned} \frac{\mathrm{d}X}{\mathrm{d}t} &= (\mu_m - k_d - D)X, \quad \mu_m = \mu_0\, \mathrm{e}^{\theta_\mu (T - T_m)}\, \frac{f_{osm}\, f_L\, f_A}{f^0_{osm}\, f^0_L\, f^0_A} \\ \frac{\mathrm{d}G}{\mathrm{d}t} &= u - q_m X - DG + \hat{d}_G, \quad q_m = \frac{\mu_m}{Y_{xg}} + m_g\, \mathrm{e}^{\theta_m (T - 37)} \\ \frac{\mathrm{d}L}{\mathrm{d}t} &= q_{lac} X - DL + \hat{d}_L, \qquad \frac{\mathrm{d}P}{\mathrm{d}t} = q_P X - DP \end{aligned} \tag{8.19}$$

where μ_0 is the estimator's $\hat{\mu}$ moved from the mean measured temperature of its twelve hours to the measured T_m, plus the death rate of Equation (8.4) at the estimate without its starvation term, $f_L = K_L/(K_L + L)$ and $f_A = K_A/(K_A + A)$ are the inhibition factors of Equation (8.3) with f^0 their values now, q_{lac} and q_P follow Equations (8.6) and (8.8), the osmolality follows the estimator's additions model with the base the pH loop adds for the lactate made, the ammonium follows Equation (8.7) with glutamine held at 3 mM, and every feed stops at V_{max}. The temperature and the complex feed over the horizon are the recipe's (the preview), or today's with the preview Off. Starvation, glutamine and amino-acid limits, pH, DO, carbon dioxide and cell size are left out, and the floor below stands in for starvation.

Each decision simulates Equation (8.19) over the horizon under a nominal dose $u^0 = q_m X + DG - \hat{d}_G + (G_{fl} - G)/3\,\mathrm{h}$ that holds the glucose at the plan's floor, linearizes it along that path by forward differences discretized to second order over each step, and writes every output as the nominal path plus $\Theta\delta$ in the block deviations δ from u^0. Twelve steps ahead the linear responses of glucose, osmolality and titer to a move agree with the nonlinear model's within 0.98 % and the lactate's within 8.3 %. A time-invariant model frozen at the estimate, the form the shared `LinearMPC` solves, was rejected in the build, because on day 3.5 it predicted 5.17 g/L of titer where the nonlinear model under the same plan gave 2.88 g/L, so the problem is condensed by hand for the shared dense quadratic program solver. The horizon runs to the harvest day (day 20 under a viability-only rule) in $N \le 120$ steps of $\Delta = \lceil t_{rest}/120 \rceil$ hours, 112 steps of 3 h at inoculation and 1 h over the last five days, with up to 12 graduated move blocks of $1, 1, 1, 1, 2, 2, 3, 4, 6, 9, 15$ steps and the remainder, and with less than

one hour left the last plan is held. The program is Equation (8.20).

$$\begin{aligned}\min_{\delta,s}\ & -w_P\hat{P}_N + S\Big[(u_0^0+\delta_1-u_{-1})^2 + \sum_{b=2}^{n_b}(\delta_b-\delta_{b-1})^2\Big] \\ & + \sum_i \left(\rho_i s_i^2 + \sigma_i s_i\right) \\ \text{s.t. } & 1+\kappa\sigma_G - s_{G,j} \le \hat{G}_k \le 6 + s_{G,j}, \quad \hat{G}_k \ge 0.5 - s_D \\ & \hat{L}_k \le 5 + s_L, \quad \widehat{\mathrm{Osm}}_k \le \mathrm{Osm}_{max} + s_O \\ & \hat{G}_k \le G_k^0 + 1, \quad 0 \le u_k \le u_{max}, \quad s \ge 0\end{aligned} \tag{8.20}$$

where $\hat{P}_N$ is the predicted titer at the end, u_{-1} the dose applied now, G_k^0 the nominal glucose, σ_G the standard deviation of the glucose estimate, and k runs over every $\lceil N/16 \rceil$-th step with one glucose slack $s_{G,j}$ per judged step and one slack each for depletion, lactate and osmolality over the horizon. The weights are $w_P = 5$ per g/L, $S = 8$ per $(\mathrm{g}/(\mathrm{L\,h}))^2$ and $\kappa = 3$, with $(\rho,\sigma) = (10^3, 200)$ on the band, $(10^4, 2000)$ on depletion, $(100, 20)$ on the 5 g/L lactate ceiling, a teaching value since the application has no lactate limit, and $(1, 2)$ on the 400 mOsm/kg comfort line, which becomes the cap's level when the cap is on. The pump's range, a nonzero dose after the predicted full step, and the 1 g/L trust region above the nominal path, which keeps the plan where the overflow sigmoid was linearized, are hard. With 12 moves and up to 20 slacks the program has at most 32 variables and 146 rows, and it took 5 to 96 iterations in three whole cultures. A rule fixed before the grid chose the tuning from $\kappa \in \{2,3,4\}$, $S \in \{0.5, 2, 8\}$ and $w_P \in \{5, 20, 80\}$ on case 10, keeping the settings with no hour outside the band, no osmolality above 400 and no fallback, then the highest true titer, with ties within 0.01 g/L broken by the least pump travel (18 of 27 qualified, at 5.04 to 5.06 g/L).

The estimator is offset-free in the sense of an input disturbance carried beside each state [21], the structure Volume 9, Chapter 4 gives its predictive controller for a fuel that no instrument measures. Glucose and lactate each carry a disturbance $\hat{d}$ in grams per liter per hour, propagated by Equation (8.19) every scan and corrected by steady-state Kalman gains for the hourly Raman and the daily sample. A residual e is taken against the estimate at the hour the sample was drawn, t_s, and carried to now as $\hat{G} \leftarrow \hat{G} + (\ell_G + \ell_d(t - t_s))e$ and $\hat{d}_G \leftarrow \hat{d}_G + \ell_d e$. The Raman gains $(\ell_G, \ell_d) = (0.342, 0.065)$ come from process noise of 0.03^2 and 0.02^2 per hour, which puts the glucose gain at the shipped estimator's 0.35, and the daily sample's $(0.978, 0.036)$ from a disturbance noise of 0.004^2. Titer and ammonium are reset to each offline result at its draw. Offline results arrive on the hour, inside the plan's first block, so the next decision replans on them, and a result doctored 1 g/L high moves $\hat{G}$ by 0.96 g/L on arrival against -0.16 undoctored and cuts the next dose from 0.145 to 0.014 g/(L h). The disturbances also enter the prediction, held over the horizon. The standard deviation σ_G of the applied gains' error, propagated in Joseph form, sets the floor at 1.30 to 1.40 g/L with the Raman. Without it only the daily sample corrects the estimate, σ_G grows to 0.55 g/L between results and the floor with it to 2.59 g/L, which is the fallback the case teaches.

The solver's answer is kept only when it is optimal and its doses, predicted outputs, objective and slacks are all finite, correctly sized and within the pump's range. Otherwise the continuous feed takes the pump from where the trim tracked it (a 0.0145 L/h step in the forced test), the fallback is written to the historian, and the following accepted plan returns the pump through its move penalty (0.19 L/h). No plan was rejected in any measured culture. A decision costs 0.66 ms on average in Chrome, at most 9.6 ms, or 3.2 % of a core at 48 simulated hours per second, measured beside other test runs on the same machine, so no frame guard is needed. The controller reads only its instruments, the laboratory results with their draw times, the estimator's $\hat{X}_v$, $\hat{\mu}$ and osmolality, its own commands, the recipe and the shipped limits, and the application's checks catch all 31 doctored copies of it, 25 that read something more and 6 that undo a reviewed behavior.

Table 8.4 Cases, Their Changes and Their Teaching Points

#	Change from the reference line	Teaching point
1	glucose pump in MANUAL, set to 0.20 L/h on day 3	an appetite that doubles
2	daily bolus to 4 g/L from the offline sample	one decision a day, and the variant no bolus holds
3	continuous feed without the Raman	a feed-forward the shift makes wrong
4	the reference line	feed-forward plus a direct trim
5	gas before oxygen, DO setpoint 60 % from day 2.25 to 3	one sparge, two loops
6	the reference line, shift day moved	a schedule with an interior optimum
7	glucose target 4.5 g/L	a margin above a threshold
8	pH probe drifting −0.02 per day	a drift only the laboratory sees
9	harvest only below 65 % viable	a decision on a late sample
10	feeding strategy MPC	a plan to the harvest, and its margin
C	bolus to 4.5 g/L, no Raman, no cap, probe drift from day 4, weak lot on day 6	graded challenge

8.5.7 Simulation, Estimator and Controller Models

Section 8.3 is the simulation model, which only the scorecard and the truth columns of the log can see. The controller holds three internal models, the estimator's glucose integrator with an uptake measured between samples, the platform yield relation of the model-based feed, and the additions-only osmolality model, and the bolus check holds a fourth, the consumption of the last day grown by the probe. None of them knows the cell size, the true stock or the switch, and no control model is identified from plant data. The MPC adds a fifth, Equation (8.19) at the default constants with its own Kalman estimator, which carries the declared switch of Equation (8.6) at its default thresholds but neither the cell size nor the true stock, and which is also written rather than identified.

8.6 Ten Cases and a Fourteen-Day Challenge

The cases of Table 8.4 start from one reference line, the continuous feed at 2 g/L with the Raman trim, the probe correction at 0.5, oxygen before gas, the shift on day 5, the complex feed from day 3, no osmolality cap, and the harvest on day 14 or below 70 % viable, and each changes what its row states. Every number below is a run of the engine at the default seed, stepped at the physics step of 10 s exactly as the browser steps it, with each change made at inoculation before the run and the run carried to the harvest, and Table 8.5 collects the scorecard. Every scored value is a true one, which the controller never sees. A harvest decided by a noisy daily viability can move by a day with the noise, and where it does at other seeds the text says so.

8.6.1 An Appetite That Doubles

Case 1 runs the glucose pump in MANUAL at zero and sets it on day 3 to 0.20 L/h of the 500 g/L stock. The medium's 6 g/L has carried the culture to 2.62 g/L by then, and the cells take 2.35 g/L that day and 3.62 g/L the next, while the pump supplies 1.77 g/L a day at the day-4 volume of 1357 L. The glucose is 1.37 g/L on day 4 and gone on day 4.67, with 7.63 million cells/mL and before the next sample has been drawn, and it stays near zero for 223.9 h. The cells then take lactate back as the only carbon left, which the application marks as starvation, peak at 8.24 million cells/mL on day 8.98 and begin to die. The day-13 sample reads 70.4 % viable, just above the limit, so the culture runs to day 14 and is harvested at 66.3 % viability with 3.28 g/L of antibody. At eight of ten noise seeds that sample reads below 70 % and the harvest comes on day 13.17 with 3.07 g/L. With the glucose loop returned to AUTO the same culture makes 5.03 g/L.

Table 8.5 Scorecard of Each Case and Variant, From Inoculation to Harvest, at the Default Seed. Glucose Is Scored From Day 2, and Every Value Is the True One

Case and configuration	Titer (g/L)	Harvest (day)	Outside (h)	Switch (day)	Osm (mOsm/kg)	Base (mol)
1, pump 0.20 L/h from day 3	3.28	14.00	235.6	—	407	19.6
1, glucose loop in AUTO	5.03	14.00	0	6.32	393	27.6
2, bolus to 3 g/L	4.07	14.00	104.1	5.51	399	20.0
2, bolus to 4 g/L, shipped	4.14	14.00	34.8	5.85[a]	435	39.7
2, bolus to 5 g/L	3.27	12.17	6.3	5.94[a]	490	68.1
2, bolus to 6 g/L	2.73	11.17	0.7	5.09[a]	511	75.8
2, bolus to 4, shift day 8	1.61	9.17	44.3	7.72[a]	426	42.4
2, bolus to 5, shift day 8	1.72	9.17	29.1	6.83[a]	446	52.0
2, bolus to 6, shift day 8	1.74	9.17	15.0	6.78[a]	465	62.5
3, no Raman, shipped	3.43	13.17	1.3	8.34	464	52.4
3, uncorrected probe	2.76	11.17	4.0	10.43	478	62.1
3, model-based feed	4.95	14.00	0	6.49	384	29.1
4, Raman trim, shipped	5.03	14.00	0	6.32	393	27.6
4, Raman out from day 4.5	3.12	12.17	3.8	9.29	465	55.0
4, target 1.5 g/L	5.07	14.00	0	5.10	390	26.7
4, target 2.5 g/L	3.66	13.17	0	—	467	55.9
4, target 3 g/L	2.72	11.17	0	—	501	72.3
5, gas before O_2, shipped	5.28	14.00	0	6.32	382	24.0
5, O_2 before gas	5.04	14.00	0	6.32	393	27.7
6, shift on day 3	4.33	14.00	0	5.77	359	3.1
6, shift on day 4	4.89	14.00	0	5.38	370	13.6
6, shift on day 6	4.83	14.00	0	7.26	414	38.2
6, shift on day 7	4.67	14.00	0	8.22	422	43.5
6, shift on day 8	3.57	12.17	0	7.30[a]	422	45.6
6, no shift	1.92	10.17	0	7.30	418	45.7
7, target 4.5 g/L, shipped	2.58	11.17	0	—	519	76.7
7, with the osmolality cap	4.44	14.00	0	7.31[a]	399.7	46.6
8, pH probe drift, shipped	4.50	14.00	0	6.31	405	27.2
8, re-standardized	5.02	14.00	0	6.32	394	27.7
9, harvest below 65 %	7.40	18.17	40.6	6.32	395	27.6
C, shipped	4.25	14.00	35.4	5.99[a]	430	39.5
C, suggested	5.21	14.00	0	6.30	381	27.9

[a]Lactate peaked again after the switch, once the glucose rose back above the switch's threshold, and the scorecard does not credit such a switch. Outside is the hours with the glucose outside 1 to 6 g/L after day 2.

> **TRY IT IN THE APP** **1 · Manual Feeding**
>
> Load scenario 1 and press Run. The glucose pump is set by hand to 0.20 L/h on day 3 and never moves again. Watch the glucose trend alone. It reaches zero on day 4.67 and stays near it for 223.9 h, and the culture is harvested with 3.28 g/L of antibody against 5.03 g/L with the loop in AUTO.

8.6.2 One Decision a Day

Case 2 feeds a bolus to 4 g/L at each glucose result. The day-3 sample reads 2.66 g/L and receives 3.5 L of stock, the day-4 sample 1.00 g/L and 8.2 L, and the glucose runs out on day 4.93 with 8.54 million cells/mL. Over the culture it is outside the band for 34.8 h and gone for 5.6 h, and the culture makes 4.14 g/L while its osmolality peaks at 435 mOsm/kg. A higher target trades one failure for another. At 3 g/L the cells starve for 20.1 h, while at 5 and 6 g/L the glucose stays above 2.5 g/L for 74.7 and 96.6 % of the hours after day 5, against 37.8 at 4 g/L, so the cells keep making lactate, the base drives the osmolality to 490 and 511 mOsm/kg, and the cultures are harvested on days 12.17 and 11.17 with 3.27 and 2.73 g/L. The honestly infeasible variant keeps the culture warm until day 8, where Equation (8.15) fails. At targets of 5 and 6 g/L the bolus check flags it on day 5.17, from a measured 4.07 g/L and 4.05 g/L grown by the probe to a projected 5.38 and 5.46 g/L for the next day, and targets of 4, 5 and 6 g/L leave the band for 44.3, 29.1 and 15.0 h. At 4 g/L the check never fires, because its glucose had run out and the culture ate only the 3.69 g/L it was given between the day-4

and day-5 samples. All three warm cultures are harvested on day 9.17, when the day-9 sample reads below 70 % viable, and at two to five of ten noise seeds a day later.

8.6.3 A Feed-Forward That the Shift Makes Wrong

Case 3 feeds continuously on the capacitance probe without the Raman. From day 3.5 to day 5 it holds 2.14 to 2.62 g/L. Then the culture is cooled, and the true per-cell uptake falls from 0.0224 on day 5 to 0.0114 g per 10^9 cells per hour on day 6, while the feed still uses the 0.0268 and then the 0.0258 it measured between samples before the shift. Nothing the controller reads shows the glucose rising until the day-6 result arrives on day 6.17, and by then it is at 6.09 g/L, as shown in Figure 8.3. Above the switch's threshold the cells keep making lactate, the switch waits until day 8.34, lactate reaches 7.00 g/L, the base drives the osmolality to 464 mOsm/kg, and the culture is harvested on day 13.17 at 64.8 % viability with 3.43 g/L, after a day-12 sample of 71.4 %. At the nine other seeds tried the day-12 sample reads below 70 %, and the harvest comes on day 12.17 with 3.08 to 3.13 g/L.

The probe adds an error of its own. In the default culture it reads 17, 22 and 21 % high on days 7, 10 and 14 as the cells swell, and the counts bring the corrected estimate to 7 % high on day 7 and 2 % low on day 14. In case 3 an uncorrected probe raises the glucose peak to 6.28 g/L and lowers the titer to 2.76 g/L, an ideal probe gives 5.55 and 3.53 g/L, and a probe that also drifts 4 % a day uncorrected gives 7.86 g/L, no switch at all and 2.62 g/L. The model-based feed knows that a cooled culture eats less. Without the Raman it keeps the glucose at or below 3.07 g/L after day 4, switches on day 6.49 and makes 4.95 g/L, and with the probe uncorrected 3.26 and 4.44 g/L.

> **TRY IT IN THE APP** **3 · Continuous Feed on the Capacitance Probe**
>
> Load scenario 3 and run to day 6.2. The controller's glucose estimate sits at its 2 g/L target and the last offline result reads 2.45 g/L, while the true glucose has climbed to 6.09 g/L since the culture was cooled on day 5. Nothing the controller reads shows the climb until the day-6 result arrives.

Pause. Try Exercise 8.2 before reading on.

8.6.4 A Direct Reading of the Broth

Case 4 adds the Raman, whose hourly readings pull $\hat{G}$ toward the broth. It catches the cooled culture within hours, holds the glucose between 1.74 and 2.55 g/L from day 4 to the harvest, and the switch comes on day 6.32 at a lactate peak of 4.54 g/L. The culture reaches 14.78 million cells/mL and 5.03 g/L, 9.47 kg, on day 14 at 80.9 % viability and 393 mOsm/kg. With the Raman's noise turned off, the glucose stays between 1.88 and 2.54 g/L and the pump moves 26 % less. Taken out of service on day 4.5, the Raman leaves the feed to overfeed the cooled culture as in case 3, with 6.27 g/L of glucose on day 6.17, and the culture is harvested on day 12.17 with 3.12 g/L of antibody, while taken out on day 9 it costs nothing, because by then the rate measured from the samples is right. The target is a lever on the switch of Section 8.4. At 1.5 g/L the switch comes on day 5.10 and the culture makes 5.07 g/L, while at 2.5 and 3 g/L it never comes, and the cultures are harvested on day 13.17 with 3.66 g/L and on day 11.17 with 2.72 g/L.

8.6.5 One Sparge, Two Loops

Case 5 raises the gas before the oxygen and steps the DO setpoint from 40 to 60 % on day 2.25 for eighteen hours. A step of 20 points held that long would be rare in a plant, and it serves here to isolate what one sparge does to two loops. The gas goes from 30.6 L/min to 92.6 L/min by day 2.5 and to 100 L/min by day 2.57, and the extra gas strips carbon dioxide, so the pH loop puts it back. On day 2.5 it doses 3.39 % of carbon dioxide into the larger flow, 3.13 L/min, where without the step it doses 2.74 % into 38.4 L/min, 1.05 L/min, to hold the same 31.8 mmHg and pH. The pH itself barely moves, to a peak of 7.010 and a dip to 6.972 on day 3.08, so the coupling shows in the base and the osmolality instead. With the gas first the p_{CO_2} stays between 7.7 and 13.7 mmHg from day 3 to day

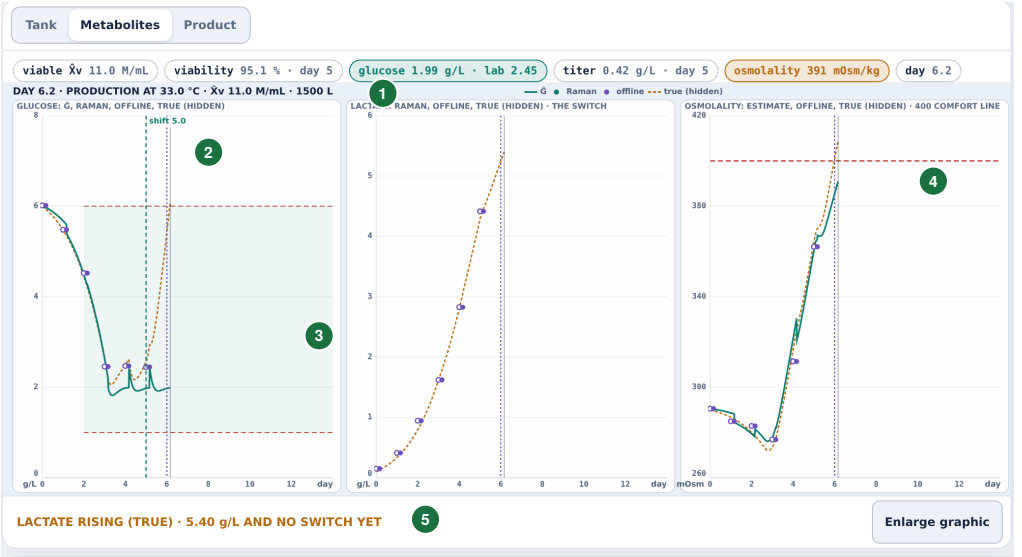


Figure 8.3 Case 3, the continuous feed on the capacitance probe without the Raman, at t = 6d 04:00 (day 6.17) in the Metabolites view, with no control changed, cropped to the plant view. Callout 1 is the badge that carries the controller's glucose estimate $\hat{G}$, 1.99 g/L against its 2.0 g/L target and the last offline result, 2.45 g/L from day 5. Callout 2 is the hidden true glucose, which has climbed to 6.09 g/L since the shift on day 5.0, and callout 3 the 1 to 6 g/L band it is about to leave, with the day-6 sample, drawn at 144 h and marked by the dashed vertical line, still at the laboratory. Callout 4 is the 400 mOsm/kg comfort line that the osmolality estimate is approaching, and callout 5 the status line, which reads lactate rising (true), 5.40 g/L and no switch yet. The badges give the corrected viable density as 11.0 million cells/mL at 1500 L, the last offline viability as 95.1 % and titer as 0.42 g/L, and the osmolality estimate as 391 mOsm/kg.

6, so 1.9 to 3.4 mM of bicarbonate holds pH 7.00 and the culture has used 23.0 mol of base by day 6. With the oxygen first the sparge never leaves 15 L/min, the p_{CO_2} runs 23.5 to 50.7 mmHg, 5.7 to 12.6 mM of bicarbonate has to be made with base, and 27.6 mol are used by day 6. Over the culture that is 382 mOsm/kg against 393 and 5.28 g/L against 5.04. With the oxygen first the same DO step moves only the enrichment, from 7 to 43 %, and the pH loop does not notice it.

8.6.6 The Shift, Early and Late

Case 6 cools the culture of case 4 on day 5, with 8.87 million cells/mL in the broth. The jacket supply falls to 28.1 °C, the broth is at 33.0 °C from day 5.18, the growth falls from 0.0223 on day 4 to 0.0059 h^{-1} on day 6 and the per-cell uptake from 0.0256 to 0.0119 g per 10^9 cells per hour. The titers for a shift on days 3, 4, 5, 6 and 7 are 4.33, 4.89, 5.03, 4.83 and 4.67 g/L, and the day-5 shift makes the most at each of ten noise seeds. Cooling early leaves too few cells, and the culture shifted on day 3 is still growing on day 14, at 15.64 million cells/mL. Cooling late leaves the warm culture making lactate, and shifted on day 7 its lactate peaks at 5.85 g/L, it needs 43.5 mol of base against 27.6, and its osmolality reaches 422 mOsm/kg, while at half of the seeds its day-13 sample already reads below 70 % and it is harvested on day 13.17 with about 4.2 g/L. Shifted on day 8 it is harvested on day 12.17 with 3.57 g/L, and never cooled on day 10.17 with 1.92 g/L. A density trigger of 8, 10 or 12 million cells/mL fires on day 4.76, 5.25 or 5.80 for 5.05, 5.01 and 4.89 g/L. It does not rescue a poor seed, because with 80 % of the seed viable the day-5 shift makes 4.77 g/L and the 10 million trigger, firing on day 5.59, makes 4.73 g/L.

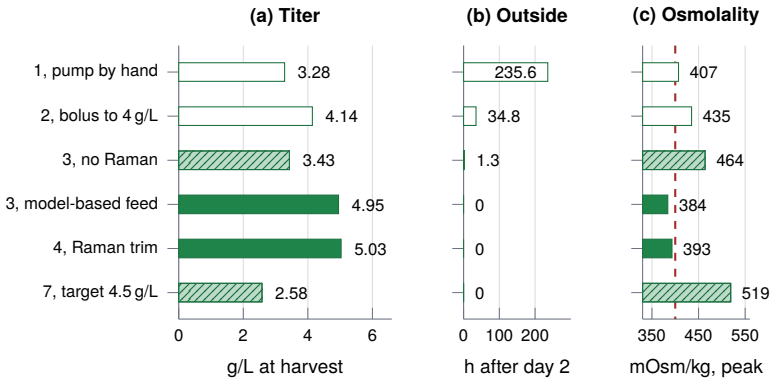


Figure 8.4 Six rows of Table 8.5 drawn as bars: the titer at harvest, the hours of true glucose outside the 1 to 6 g/L band after day 2, and the highest osmolality against the 400 mOsm/kg comfort line (dashed). White bars are the two feeds that never read the broth, hatched bars the two that read it late or not at all, and solid bars the two that read it hourly or know that a cooled culture eats less. Notice that the two tallest titers are the two lowest osmolalities, and that the target of 4.5 g/L holds the band perfectly while making the least antibody of all six.

8.6.7 A Margin Above a Threshold

Case 7 raises the target to 4.5 g/L, the setting of an operator who wants room above the floor. With the glucose at 4.27 to 5.03 g/L the cells keep overflowing it after the shift. Lactate rises from 5.57 on day 6 to 7.29 on day 8 and 8.60 g/L on day 10, ammonium from 5.00 to 7.03 mM, and the osmolality from 406 to 458 and 499 mOsm/kg, on its way to 519 mOsm/kg. Growth stops at 9.51 million cells/mL on day 6.70, the day-11 sample reports a viability below 70 %, and the culture is harvested on day 11.17 at 64.0 % with 2.58 g/L. With the osmolality cap on, the complex feed and the target come down as the estimate nears 390 mOsm/kg, the switch comes on day 7.31, the osmolality peaks at 399.7 mOsm/kg and the culture makes 4.44 g/L. Lowering the target works better, since neither 3 nor 2.5 g/L ever switches while 2 g/L switches on day 6.32. Figure 8.4 sets this case beside five others of Table 8.5.

> ✦ **DID YOU KNOW? The Comfortable Target That Costs the Most**
>
> From the simulator: a glucose target of 4.5 g/L, a cautious margin above the floor, makes 2.58 g/L of antibody, the least of any case, while a target of 2 g/L makes 5.03 g/L. The cells never stop making lactate, and the base that holds the pH drives the osmolality to 519 mOsm/kg (Table 8.5).

8.6.8 A Drift Only the Laboratory Sees

Case 8 drifts the pH probe 0.02 a day low, and the loop holds what the probe reads at 7.00. The broth goes where the probe's error takes it, 7.078 on day 4, 7.137 on day 7 while the probe reads 6.994, 7.200 on day 10 and 7.280 on day 14, and the offline results climb from 7.03 on day 2 to 7.26 on day 13 while the probe's trend stays at its setpoint. The total base hardly changes, 27.2 mol against 27.6, but the broth ends with 28.4 mM of bicarbonate against 27.1, the cells peak at 12.22 million cells/mL against 14.78, and the culture makes 4.50 g/L. Re-standardized to each offline pH through the shared bias update with a gain of one, the probe follows the laboratory, the true pH stays within 0.088 of 7.00, and the culture makes 5.02 g/L.

8.6.9 A Harvest on a Late Sample

Case 9 runs until a daily sample reports a viability below 65 %. From day 14 to day 18 the true titer is 5.03, 5.57, 6.11, 6.75 and 7.32 g/L at 80.9, 78.2, 75.6, 70.2 and 56.7 % viability. The vessel is full

Table 8.6 The Challenge With One Lever Changed at a Time

Configuration	Titer (g/L)	Outside (h)	Osm (mOsm/kg)	Viable (%)	Target
as shipped, bolus to 4.5 g/L	4.25	35.4	430	73.2	not met
bolus to 3 g/L	3.83	118.7	400	75.0	not met
bolus to 4.5 g/L with the cap	4.75	59.4	396	80.3	not met
continuous at 2 g/L, no Raman, no cap	3.41	3.6	462	65.1	not met
suggested, Raman and cap	5.21	0	381	83.3	met
suggested without the cap	5.01	0	390	81.7	met
suggested, model-based feed	5.06	0	371	84.2	met
suggested at 4.5 g/L	4.41	0	401	62.4	not met
suggested without the Raman	3.10	0	414	58.0	not met

on day 16.22 and the feeds stop, the glucose is gone on day 16.77, the day-17 sample reads 69.1 % and the day-18 sample 56.5 %, and the culture is harvested when that result arrives, on day 18.17, at 54.8 % true viability with 7.40 g/L, 14.81 kg. The limit sits at 65 % because the day-17 sample is drawn at a true 70.2 %. A 70 % limit harvests on day 17.17 at 67.7 % with 6.85 g/L, and at five of ten noise seeds a day later, while at 65 % all ten harvest on day 18.17. An 80 % limit harvests on day 15.17 at 77.8 % with 5.66 g/L, and a 75 % limit on day 17.17 at 67.7 %, more than seven points below its own limit, and at three of ten seeds each comes a day earlier.

8.6.10 The Fourteen-Day Challenge

The challenge feeds a whole culture by a daily bolus to 4.5 g/L, with no Raman and no cap, while the capacitance probe drifts 4 % a day from day 4 and a 420 g/L lot replaces the 500 g/L stock on day 6 under the same label. The target asks for at least 5.0 g/L by day 14 with the glucose between 1 and 6 g/L after day 2, the osmolality under 400 mOsm/kg, the DO above 30 %, the pH within 0.15 of 7.00 and more than 70 % viable at harvest, all scored on the truth, and a changed model parameter or disturbance voids the attempt. As shipped the glucose runs out on day 5.05 and is outside the band for 35.4 h, the osmolality peaks at 430 mOsm/kg, and the culture makes 4.25 g/L and is 73.2 % viable at harvest. Table 8.6 changes the levers one at a time. The suggested continuous feed at 2 g/L with the Raman and the cap holds 1.76 to 2.58 g/L after day 4, switches on day 6.30, keeps the pH within 0.072 and the DO above 39.2 %, and makes 5.21 g/L at 381 mOsm/kg and 83.3 %, 0.21 g/L above the target, whether the settings are revealed before the run or carried through a reset. It passes at each of ten noise seeds, with 5.16 to 5.25 g/L. The cap supplies that margin. Without it the culture passes at the default seed with 5.01 g/L at 390 mOsm/kg, only 0.01 g/L above the target, and falls short of the titer at six of the ten seeds. The Raman is necessary, since without it the feed overfeeds the cooled culture and is harvested on day 11.17 with 3.10 g/L.

8.6.11 A Plan to the Harvest, and What It Is Worth

Case 10 feeds the culture of case 4 by the MPC of Section 8.5.6. With no glucose target the plan holds the glucose between 1.09 and 2.05 g/L after day 4 on a floor of 1.30 to 1.40 g/L, where the continuous feed holds 1.74 to 2.55 g/L, so the lactate switch comes on day 5.10 at a peak of 4.18 g/L rather than on day 6.32 at 4.54 g/L. The culture makes 5.06 g/L at 81.2 % viability and 392 mOsm/kg, against 5.03 and 5.08 g/L for the continuous and model-based feeds, and the pump travels 27.8 L/h over the culture against 54.0 and 54.3 L/h. The titer the plan maximizes is its model's prediction, not a measurement. On day 7 it predicts 6.38 g/L at the harvest, a quarter above what is made, because $\hat{\mu}$ carried over the rest of the batch declines more slowly than the culture's growth, and the prediction closes to 5.38 on day 10 and 5.07 g/L on day 13. Over the first 24 h the condensed prediction of the glucose stays within 0.04 g/L of the nonlinear model under the same plan in 8 of 10 cultures (0.28 on the day-8 shift and 0.97 in case 9), while over the whole horizon the two part by up to 15.4 of glucose, 12.0 of lactate, 7.0 g/L of titer and 376.5 mOsm/kg, since the plan's deviations carry the

Table 8.7 Titer at Harvest Under Every Feeding Structure, Same Seed, Instruments and Noise. Hours of True Glucose Outside 1 to 6 g/L After Day 2 in Parentheses Where Nonzero

Culture	Bolus (shipped)	Continuous (shipped)	Model-based (shipped)	Continuous retuned[a]	Model-based on the MPC's estimate[b]	MPC	MPC, preview Off
case 10 as shipped	4.14 (34.8)	5.03	5.08	5.07	5.05	5.06	5.06
Raman out from day 4.5	4.06 (30.4)	3.12[c] (3.8)	4.95	3.95	5.06	5.01	5.01
no Raman	4.15 (35.5)	3.43[c] (1.3)	4.95	3.70[c]	5.08	4.96	4.96
shift on day 3	3.48 (0.3)	4.33	4.53	4.40	4.56	4.40	4.40
shift on day 8	1.61[c] (44.3)	3.57[c]	4.50	4.73	4.53	4.67	4.76 (1.9)
viability at harvest (%)	70.5	68.2	77.3	67.9	76.9	67.3	68.4
shift at 10×10^6 cells/mL	4.28 (52.2)	5.01	5.03	5.07	5.03	5.06	5.06
challenge, suggested levers	4.86 (55.2)	5.21	5.06	5.17	5.02	5.16	5.16
challenge, shipped levers	4.25 (35.4)	3.41[c] (3.6)	4.81	3.67[c]	4.99 (6.7)	4.88 (6.9)	4.88 (6.9)
pump travel on case 10 (L/h)	880.0	54.0	54.3	40.4	54.0	27.8	30.3

[a]The continuous feed retuned by the MPC's rule: target 1.5 g/L, gain 0.5, 8 h. The model-based feed so retuned (2 g/L, 0.5, 8 h) is within 0.01 g/L of the shipped one in every row. [b]Trimmed on the MPC's Kalman estimate, its target at the MPC's floor. [c]Harvested before day 14. In the challenge the bolus keeps 4.5 g/L and the continuous feeds 2 g/L. Every MPC run made 336 decisions, none rejected.

nonlinear model across the overflow and osmotic thresholds between the judged steps. Enforcing the trust region at every step moves no titer by more than 0.04 g/L.

Table 8.7 sets the MPC beside each shipped feed, the continuous feed retuned once by the MPC's rule (23 of 36 settings qualified) and the strongest classical feed of the MPC's signals. The MPC's clear gain is the lost Raman. Taken out on day 4.5, the Raman leaves the continuous feed to overfeed the cooled culture, which is harvested on day 12.17 with 3.12 g/L, and the retuned feed recovers only to 3.95 g/L, while the MPC keeps the glucose at 1.24 to 3.04 g/L after day 4 and makes 5.01 g/L, as shown in Figure 8.5. Its floor climbs from 1.24 toward 2.59 g/L as σ_G grows between daily results and drops back at each one. The model-based feed makes 4.95 g/L in the same culture, however, and the classical feed on the MPC's estimate makes 5.06 g/L, and with no Raman from inoculation the classical feed makes 5.08 g/L against the MPC's 4.96. The estimate, not the plan, carries the Raman-less cultures. On the day-8 shift the MPC makes 4.67 g/L but ends 67.3 % viable, under the 70 % the challenge asks, where the model-based feed makes 4.50 g/L at 77.3 %, and the retuned continuous feed makes 4.73 at 67.9 %. On the day-3 shift the model-based feeds make 4.53 and 4.56 g/L against 4.40, and at a density trigger every continuous structure lands between 5.01 and 5.07 g/L. With the challenge's suggested levers the MPC passes with 5.16 g/L, against 5.21 g/L for the suggested continuous feed. With its shipped levers, no Raman and no cap, the MPC makes 4.88 g/L with 6.9 h under 1 g/L, where the model-based feed makes 4.81 with none. The preview Off moves no titer by more than 0.09 g/L, the most on the day-8 shift, 4.76 against 4.67 g/L with 1.9 h under the band, because the hourly plan sees the shift in the measured temperature within hours. Measurably, the MPC brings a feed that survives the lost Raman where the shipped continuous feed does not (1.89 g/L more), the earliest switch of the shipped structures (day 5.10 against 5.20 to 6.32), and about half their pump travel.

8.6.12 What the Review Changed in the Plan

The MPC as first built was reviewed against every shipped feed, and three of its behaviors were changed, with the results in Table 8.8. The plan was not offset-free. Its disturbances corrected the estimate but entered neither the prediction nor the nominal dose, so under a sustained disturbance the nominal settled where $\hat{G} = G_{fl} + (3\,\mathrm{h})\hat{d}_G$. After a 300 g/L lot hung on day 6, with $\hat{d}_G$ near –0.09 g/(L h), the true glucose sat between 1.00 and 1.15 g/L from hour 178 to hour 200 under a floor of 1.24 to 1.33 g/L that the plan believed it held, and the disturbances carried in the prediction alone took that culture from 22.4 to 13.9 h under the band. The Raman's disturbance gain, 0.014 at a disturbance noise of 0.004^2, learned 63 % of the stock step in 24 h and 90 % in 42 h while the estimate ran up to 0.46 g/L above the truth, more than the margin of $3\sigma_G \approx 0.36\,\mathrm{g/L}$. The Raman

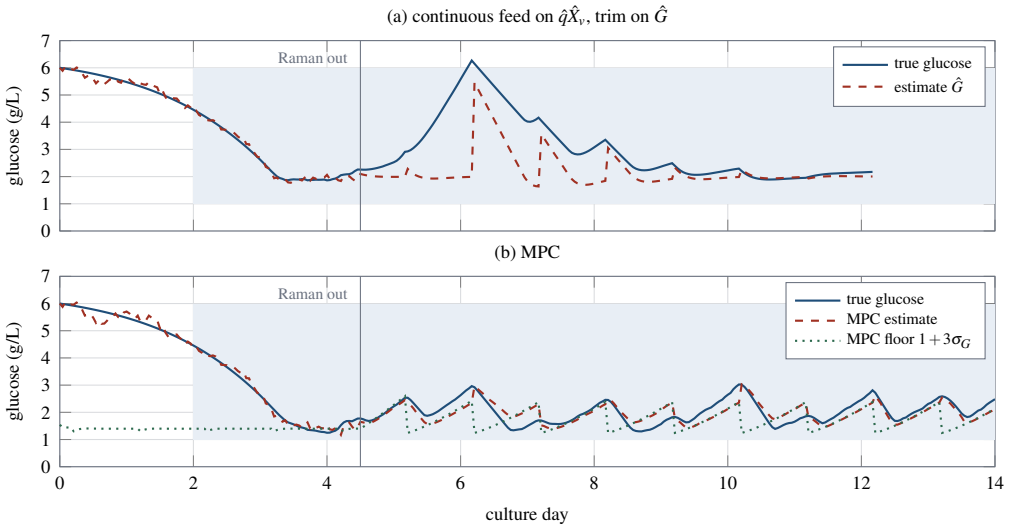


Figure 8.5 Case 10's culture with the Raman out of service from day 4.5, under the continuous feed (a) and the MPC (b), sampled hourly, with the true glucose solid, each estimate dashed and the MPC's floor dotted. The shaded box is the 1 to 6 g/L band scored after day 2. Between daily results the continuous feed's estimate sits at its 2 g/L target while the cooled culture's glucose climbs to 6.27 g/L on day 6.17, and the culture is harvested on day 12.17 with 3.12 g/L. The MPC's floor rises between daily results as its estimate ages, the true glucose stays between 1.24 and 3.04 g/L after day 4, and the culture makes 5.01 g/L.

Table 8.8 The MPC Before and After the Review: Hours of True Glucose Under 1 g/L After Day 2, Lowest True Glucose and Titer

Culture	As first built			After review		
	(h)	(g/L)	(g/L)	(h)	(g/L)	(g/L)
300 g/L lot on day 6	22.4	0.80	4.96	0	1.12	4.99
probe drift 10 % a day	2.9	0.97	5.06	0	1.05	5.10
challenge, suggested levers	4.0	0.95	5.14	0	1.14	5.16
with the drift and the lot	22.0	0.75	4.97	0	1.03	5.03
300 g/L stock, labeled so	24.9	0.83	4.97	0	1.08	5.01
challenge, shipped levers	2.3	0.98	4.87	6.9	0.94	4.88
Raman out day 2, drift, lot	17.2	0.81	4.31	30.5	0.58	4.47
no Raman, 300 g/L lot	47.1	0.50	4.92	44.9	0.49	4.93
pump travel, case 10 (L/h)			16.2			27.8

The rows below the first rule run without the Raman. No MPC row rose above 6 g/L.

gains were recomputed at 0.02^2, the shipped estimator's Raman speed, and now learn the same step 63 % in 12 h and 90 % in 13 h, with the laboratory gains and the covariance under the floor unchanged. That amendment was fixed before its single trial, with its acceptance stated in advance (no culture with the Raman leaving the band and no titer falling by more than 0.02 g/L). The pump also ran the dose on the declared 500 g/L while every shipped feed reads the label, so a 300 g/L stock correctly labeled was underdosed by 40 % and spent 24.9 h under the band. With all three changes the MPC holds the band in every culture with the Raman.

The changes have a price, reported rather than tuned away. Without the Raman the plan learns a step only from the daily sample, and a disturbance learned from the laboratory is sometimes of the wrong sign for a fresh step, which adds 4.6 and 13.3 h under the band to two of the three Raman-less

Table 8.9 Worst Cases on Case 10's Culture Under the Shipped Feeds and the MPC: Titer (g/L), Hours of Glucose Outside 1 to 6 g/L After Day 2 and Highest Osmolality (mOsm/kg), With the MPC's Lowest DO (%), Widest pH Deviation and Viability at Harvest (%)

Fault or disturbance	Continuous			Model-based			MPC					
	Titer	Out	Osm	Titer	Out	Osm	Titer	Out	Osm	DO	pH	Viable
300 g/L lot on day 6	5.01	0	382	4.94	0	360	4.99	0	378	39.2	0.081	83.6
capacitance drift 10 % a day from day 4	4.93	0	398	5.00	0	388	5.10	0	391	39.2	0.081	81.4
offline results 24 h late	4.87	0	402	5.07	0	379	5.06	0	391	39.1	0.081	81.2
seed 80 % viable	4.77	0	391	4.89	0	369	4.79	0	391	39.2	0.083	82.7
O_2 supply to 40 % on day 5	5.05	0	393	5.09	0	375	5.09	0	391	15.4	0.098	81.6
chiller failed for an hour on day 5.1	5.03	0	393	5.08	0	375	5.06	0	392	39.2	0.081	81.2
pH probe drift −0.1 a day	1.45	0	443	1.75	0	345	1.45	0	435	39.4	1.400	80.9
stock of 500 g/L labeled 300	5.03	0	394	5.04	0	377	5.07	0	393	39.2	0.083	81.2
stock of 300 g/L labeled 300	4.99	0	379	4.95	0	356	5.01	0	375	39.2	0.083	83.8
challenge, suggested levers, drift and lot	5.05	0	380	4.97	0	364	5.03	0	377	39.2	0.081	83.6
Raman out day 2, drift, 300 g/L lot	3.31	8.3	462	4.45	103.4	398	4.47	30.5	406	39.2	0.056	78.0
no Raman, 300 g/L lot	3.53	0	443	4.68	156.5	369	4.93	44.9	386	39.1	0.077	83.5
case 9, harvest below 65 % viable	7.40	40.6	395	7.68	37.2	382	7.29	36.6	395	39.2	0.123	56.8

rows of Table 8.8 and removes 2.2 h from the third. The pump travels 27.8 L/h on case 10 against 16.2, and the estimator is no longer consistent, because its σ_G of 0.08 to 0.16 g/L sits below the 0.167 g/L standard deviation of its actual error, where the first build claimed 0.115 against an actual 0.123. Raising the disturbance noise of the laboratory gains and the covariance as well, a variant tried and rejected during the build, lifts σ_G above 1 g/L between samples and sinks the Raman-less cultures, 4.98 to 2.63 g/L with the Raman out on day 4.5 and 4.95 to 2.52 with no Raman. The tuning grid was not rerun after the review.

8.6.13 Every Structure Against Every Limit

Table 8.9 runs thirteen faults and disturbances under the shipped feeds and the MPC against every limit the application checks, the 1 to 6 g/L band after day 2 and depletion, the 400 mOsm/kg comfort line, the DO at 30 and 20 %, the pH within 0.15 of 7.00, 70 % viability at harvest and the 2000 L vessel. With the Raman in service the MPC never leaves the band where the shipped feeds hold it. Its osmolality passes 400 only where the continuous feed's does too (406 against 462, 435 against 443), its lowest DO is within 0.1 points of the continuous feed's in every row (15.4 against 15.5 % when the oxygen supply falls, under the 20 % alarm for every structure that feeds continuously), and its viability is under 70 % only in case 9, which harvests on a late decline by design. Its widest pH deviation, ±0.123 at the harvest of case 9, is under the 0.15 check level. After day 16 the carbon dioxide leg of the pH loop sits at its 20 % limit under the continuous, model-based and MPC feeds alike, and the MPC's culture entered the full vessel with 0.11 g/L of lactate left to take back, against 1.89 and 1.23 g/L under the continuous and model-based feeds, which reach ±0.036 and ±0.046 then. The losses are the cultures without the Raman, 30.5 and 44.9 h under the band, against 8.3 and 0 for the continuous feed and 103.4 and 156.5 h for the model-based feed. The only depletion under the MPC is case 9's 28.9 h after the vessel fills, against 33.6 and 31.1 h. The bolus leaves the band 4.1 to 110.8 h in every row but the pH drift. No plan was rejected, and the cap, the vessel-full stop and the harvest act above the MPC unchanged.

Two older defects found during the build were fixed with no shipped or suggested run moved. A seed of 60 to 70 % viability was harvested at 4 h with no antibody on its day-0 sample, and under the decline rule it now makes 4.33 to 4.58 g/L on the continuous feed and 4.36 to 4.59 on the MPC. The glucose and osmolality estimates started at the broth's own values, so a medium set to 8 g/L or 320 mOsm/kg was known at inoculation, and both now start at the declared 6 and 290. Because 19 laboratory and Raman columns are empty until their first report, the suite's finite-column check runs under the MPC as an application test that allows them empty.

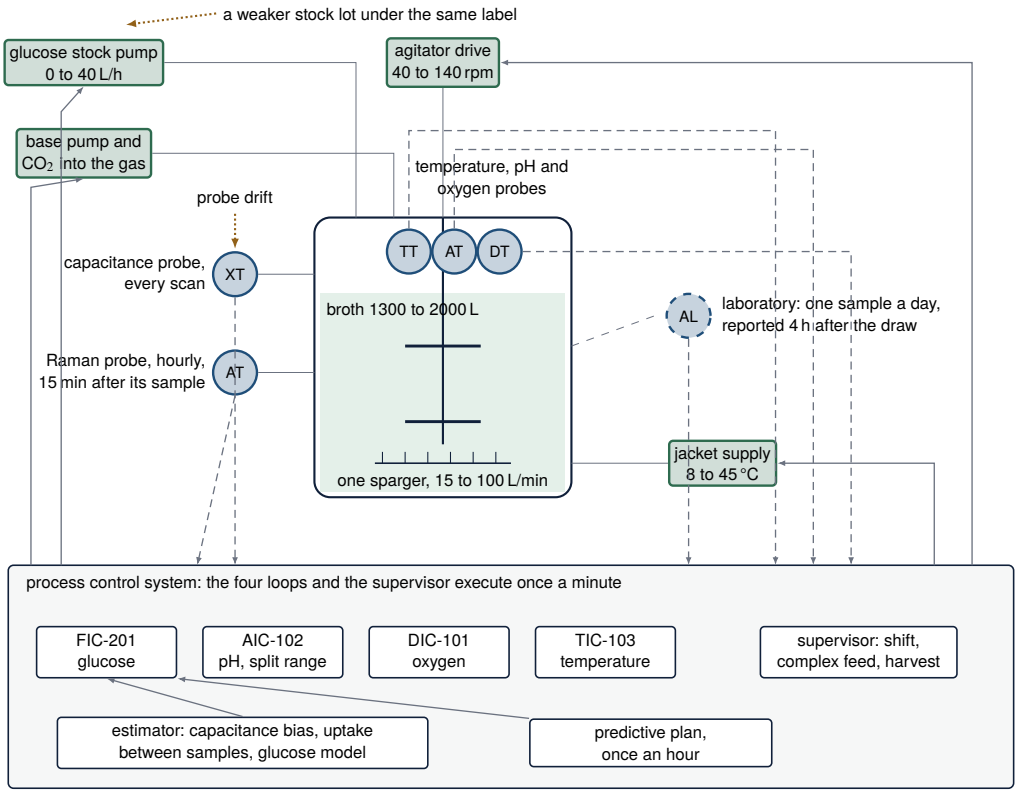


Figure 8.6 The four loops of Section 8.5 with the equipment of a culture vessel in place of the schematic's symbols. A capacitance probe (XT) and an in-line Raman probe (AT) read the broth every scan and every hour, while one sample a day goes to a laboratory (AL, dashed) and is reported four hours after the draw. The final elements are a dosing pump on the glucose stock, a base pump that shares the pH duty with carbon dioxide in the sparge gas, the agitator drive with the gas train and the jacket supply. Notice that the only instrument that sees glucose, cells, metabolites and product together is the one that answers last, and that the estimator stands between it and the feed.

8.7 Technology in Practice

The probes, the sample line and the pumps of Figure 8.1 stand for the equipment of a manufacturing culture vessel. Figure 8.6 draws the four loops of Section 8.5 with that equipment in place of the schematic's symbols.

What Measures It

Three probes stand in the broth and report every scan: a dissolved oxygen head, a glass pH electrode against a reference junction, and a temperature probe. Their lags of 1, 0.5 and 0.5 min in Section 8.5 stand for the response of a probe in a stirred broth rather than for a transmitter. Two more instruments read what the cells are doing. A capacitance probe applies a weak electric field and reports the permittivity of the broth, which follows the volume of intact cells rather than their number, while a plant otherwise learns cell physiology only from a drawn sample [18]. The in-line Raman probe reads the metabolites themselves. The chapter's hourly Raman with 0.25 g/L of noise stands for such a probe, and its four-hour laboratory for the sample that is still the reference.

TECHNOLOGY Sensor: The In-Line Raman Probe

A Raman probe stands in the broth and reads the light that molecules scatter at shifted wavelengths, so one spectrum carries glucose, lactate, cell density and product together. In a recent study a probe with a laser at 785 nm built each spectrum from 50 accumulations of 15 s, 12.5 min for one vessel, and a partial least squares model predicted glucose on a culture it had not seen with a coefficient of determination of 0.955 and a root mean square error of 0.487 g/L [22]. Such a probe fails slowly rather than suddenly. In the same cultures the spectral baseline climbed as the culture aged and stepped up at each nutrient feed, because fluorescence from the feed's aromatic amino acids masks the signal [22].

What Moves It

The glucose stock reaches the broth through a dosing pump, which the chapter ranges from 0 to 40 L/h, and the pH through a split between a carbonate base and carbon dioxide in the sparge gas. Base and gas are the two legs of AIC-102. The agitator drive and the gas train carry the oxygen demand in sequence, and the jacket supply carries the temperature shift at 1 K/h.

TECHNOLOGY Final Element: A Dosing Pump and a Split Gas

Two final elements carry the glucose and the pH of a culture vessel: a dosing pump on the stock bottle and a split between carbon dioxide in the sparge gas and a carbonate base. In one reported set of cultures the pH was held inside a deadband of 0.1 by sparging carbon dioxide and adding sodium carbonate, and the glucose pump started 38 and 39 times in fourteen days under a spectroscopic feedback strategy against 13 times under a daily bolus by hand [22]. Base is the harder of the two at scale, since base entering from the top leaves zones of high pH and low carbon dioxide behind it, while carbon dioxide that accumulates lowers growth and productivity [23].

Where the Controller Runs

The platform is the vessel's own control system, with the spectroscopic analyzer and its calibration models on a separate computer that exchanges values with it over a standard machine-to-machine protocol. In the reported installation the analyzer delivered a new prediction every 12.5 min and the feed decision followed each one, against the once-a-minute scan of the four loops in this chapter and the hourly plan of its predictive controller [22]. The supervisor of Section 8.5 is the recipe layer above those loops, and in a plant it is the layer that holds the batch record.

IN PRACTICE What Makes a Probe Part of the Control Strategy

A measurement becomes part of manufacturing through process analytical technology, which the United States regulator describes as a system for designing, analyzing and controlling manufacturing through timely measurements of critical quality and performance attributes, with the goal of ensuring final product quality. Within that framework a process end point is not a fixed time but the achievement of the desired material attribute [24]. Automatic equipment used in manufacture is routinely calibrated, inspected or checked under a written program whose records are kept (21 CFR 211.68) [25]. Engineers who have put such a loop into service report that a fallback to sampling by hand is needed for the hours when the link between analyzer and controller fails [22].

What Is Modernizing It

Three developments bear on this feed. The first is the closed analyzer loop itself. A spectroscopic strategy that held glucose at a setpoint instead of topping the culture up once a day raised the product titer by up to 35 % and lowered glycation of the product by up to 27 % [22]. The second is making a calibration survive a change of scale or medium. A 2026 benchmark of four unsupervised transfer learning methods on Raman data found two of them useful where differences between the source and the target dominate, and found no benefit where the variability inside one domain is large [26]. The third is a feed-forward taken from a model rather than from the last interval. Regression of nutrient

consumption on viable cell density held several amino acids and glucose at their targets in fed-batch and perfusion cultures [27], which is the model-based feed of Section 8.5 in a plant.

8.8 Limitations

The broth is one well-mixed volume with unstructured kinetics, a declared switch and a shift that acts at once, and every constant is a teaching value, so no number here predicts a cell line, a titer or a harvest day. The real values come from a line's own characterization and from measured responses to temperature, carbon dioxide and osmolality [3, 10, 12, 15]. The pH buffer is bicarbonate alone, the gas phase has no head and no foam, and the capacitance probe is a cube law on one mean diameter. The 1 to 6 g/L band, the 400 mOsm/kg comfort line, the harvest limits of 70 and 65 % and the challenge's targets are teaching values, while a plant takes its limits from its own process characterization and its regulatory filing. Product quality is not modeled, and the scorecard credits titer and constraints rather than the glycans and impurities a purification has to remove.

> ⚠ **OPEN CHALLENGE** **Telling How Many Cells From How Large**
>
> A capacitance probe answers with one number that mixes how many cells there are with how large they are, which is why this chapter's probe reads high as the cells swell. A sensitivity study of such probes on cell cultures found the permittivity increment sensitive to both cell size and concentration, so it suits biovolume while the cells grow, and found the single population assumption behind the usual reading invalid once a culture enters its stationary and death phases, where cells of another size and other dielectric properties appear. Membrane permittivity and internal conductivity cannot be recovered from the measured spectrum at all [18]. Progress would be a reading that separates count from size late in a culture.

The control scope is four loops on one shared law, four feeding strategies and a rule-based supervisor. The model-based feed is a feed-forward on a platform yield relation, and the MPC is a receding-horizon optimization of the kind that has been applied to glucose control [17], but it optimizes only the feed profile, and nothing in the application optimizes the shift day or the harvest. The estimator corrects the probe and the glucose from the samples but does not estimate the switch, the cell size or the true stock. The off-gas rates are logged and unused, although an on-line oxygen uptake carried the feeding of a hybridoma fed-batch as early as 1995 [4].

The MPC's model carries the probe's growth rate forward with no decline of its own, so its day-7 titer is a quarter high, and it omits starvation, the amino acids, pH, DO, carbon dioxide and cell size. Its far horizon is rough, its estimator is optimistic after the review, and its objective prices neither product quality nor medium nor time in the vessel.

8.9 Conclusion

This chapter studied a CHO fed-batch for a monoclonal antibody as a simulator in which the culture's appetite, its one sparge and its osmolality are logged beside everything its controller reads. Fed by hand, the culture runs out of glucose on day 4.67 and makes 3.28 g/L. Fed a daily bolus to 4 g/L it spends 34.8 h outside the band, and held warm until day 8 it eats up to 5.93 g/L in a day, wider than the 5 g/L band, so no bolus can hold it. Fed continuously on the capacitance probe it overfeeds the cooled culture to 6.09 g/L and makes 3.43 g/L, while the Raman trim holds 1.74 to 2.55 g/L, brings the switch on day 6.32 and makes 5.03 g/L. A target of 4.5 g/L prevents the switch and reaches 519 mOsm/kg, gas before oxygen saves base and makes 5.28 g/L against 5.04, and a probe drifting 0.02 a day leaves the broth at 7.280. The challenge improves from 4.25 g/L at 430 mOsm/kg to 5.21 g/L at 381 mOsm/kg. A model predictive controller planning to the harvest keeps 5.01 g/L when the Raman is lost, against 3.12 for the continuous feed, but a model-based feed on the same estimate makes 5.06, so the plan's own measurable gain is the earliest switch of the shipped structures and half their pump travel.

The kinetics are teaching values in one well-mixed broth. A fuller model would add a structured switch, a compartment model of mixing, foam, product quality and an optimized shift day and harvest. The cases show that the feed of a growing culture is decided by what the controller can measure, and when.

KEY POINTS

- A culture that doubles its cells doubles its appetite, so a feed sized from this morning's sample is already too small tomorrow.
- Once the glucose the cells take between two reports passes the 5 g/L width of the band, no bolus target can hold the band.
- A feed-forward on the uptake measured before a temperature shift overfeeds the cooled culture to 6.09 g/L, where an hourly direct reading holds 1.74 to 2.55 g/L.
- Every correction leaves something behind: a gram per liter of lactate held at the pH setpoint costs 22 mOsm/kg, and the sodium never leaves the vessel.

Where else this applies. A feed-forward on an estimate that a day-old laboratory result corrects returns in Chapter 10 and Volume 5, Chapter 10, and a band held on a sampled analyzer in Volume 5, Chapter 4.

References

[1] B. Kelley, "Industrialization of mAb production technology: The bioprocessing industry at a crossroads," *mAbs*, vol. 1, no. 5, pp. 443–452, 2009.

[2] D. M. Ecker, S. D. Jones, and H. L. Levine, "The therapeutic monoclonal antibody market," *mAbs*, vol. 7, no. 1, pp. 9–14, 2015.

[3] Y.-M. Huang, W. Hu, E. Rustandi, K. Chang, H. Yusuf-Makagiansar, and T. Ryll, "Maximizing productivity of CHO cell-based fed-batch culture using chemically defined media conditions and typical manufacturing equipment," *Biotechnology Progress*, vol. 26, no. 5, pp. 1400–1410, 2010.

[4] W. Zhou, J. Rehm, and W.-S. Hu, "High viable cell concentration fed-batch cultures of hybridoma cells through on-line nutrient feeding," *Biotechnology and Bioengineering*, vol. 46, no. 6, pp. 579–587, 1995.

[5] R. Luttmann, D. G. Bracewell, G. Cornelissen, K. V. Gernaey, J. Glassey, V. C. Hass, C. Kaiser, C. Preusse, G. Striedner, and C.-F. Mandenius, "Soft sensors in bioprocessing: A status report and recommendations," *Biotechnology Journal*, vol. 7, no. 8, pp. 1040–1048, 2012.

[6] M. Gagnon, G. Hiller, Y.-T. Luan, A. Kittredge, J. DeFelice, and D. Drapeau, "High-end pH-controlled delivery of glucose effectively suppresses lactate accumulation in CHO fed-batch cultures," *Biotechnology and Bioengineering*, vol. 108, no. 6, pp. 1328–1337, 2011.

[7] T. E. Matthews, B. N. Berry, J. Smelko, J. Moretto, B. Moore, and K. Wiltberger, "Closed loop control of lactate concentration in mammalian cell culture by Raman spectroscopy leads to improved cell density, viability, and biopharmaceutical protein production," *Biotechnology and Bioengineering*, vol. 113, no. 11, pp. 2416–2424, 2016.

[8] G. Köhler and C. Milstein, "Continuous cultures of fused cells secreting antibody of predefined specificity," *Nature*, vol. 256, no. 5517, pp. 495–497, 1975.

[9] Ortho Multicenter Transplant Study Group, "A randomized clinical trial of OKT3 monoclonal antibody for acute rejection of cadaveric renal transplants," *New England Journal of Medicine*, vol. 313, no. 6, pp. 337–342, 1985.

[10] S. K. Yoon, J. Y. Song, and G. M. Lee, "Effect of low culture temperature on specific productivity, transcription level, and heterogeneity of erythropoietin in Chinese hamster ovary cells," *Biotechnology and Bioengineering*, vol. 82, no. 3, pp. 289–298, 2003.

[11] Z. Xing, B. M. Kenty, Z. J. Li, and S. S. Lee, "Scale-up analysis for a CHO cell culture process in large-scale bioreactors," *Biotechnology and Bioengineering*, vol. 103, no. 4, pp. 733–746, 2009.

[12] M. M. Zhu, A. Goyal, D. L. Rank, S. K. Gupta, T. Vanden Boom, and S. S. Lee, "Effects of elevated pCO_2 and osmolality on growth of CHO cells and production of antibody-fusion protein B1: A case study," *Biotechnology Progress*, vol. 21, no. 1, pp. 70–77, 2005.

[13] N. R. Abu-Absi, B. M. Kenty, M. E. Cuellar, M. C. Borys, S. Sakhamuri, D. J. Strachan, M. C. Hausladen, and Z. J. Li, "Real time monitoring of multiple parameters in mammalian cell culture bioreactors using an in-line Raman spectroscopy probe," *Biotechnology and Bioengineering*, vol. 108, no. 5, pp. 1215–1221, 2011.

[14] J. P. Carvell and J. E. Dowd, "On-line measurements and control of viable cell density in cell culture manufacturing processes using radio-frequency impedance," *Cytotechnology*, vol. 50, no. 1–3, pp. 35–48, 2006.

[15] C. Kontoravdi, E. N. Pistikopoulos, and A. Mantalaris, "Systematic development of predictive mathematical models for animal cell cultures," *Computers & Chemical Engineering*, vol. 34, no. 8, pp. 1192–1198, 2010.

[16] B. C. Mulukutla, M. Gramer, and W.-S. Hu, "On metabolic shift to lactate consumption in fed-batch culture of mammalian cells," *Metabolic Engineering*, vol. 14, no. 2, pp. 138–149, 2012.

[17] S. Craven, J. Whelan, and B. Glennon, "Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller," *Journal of Process Control*, vol. 24, no. 4, pp. 344–357, 2014.

[18] E. Salimi, S. Absalan, J. Robitaille, J. Montes, M. Butler, D. Thomson, and G. Bridges, "Sensitivity of bulk electrical impedance spectroscopy (bio-capacitance) probes to cell and culture properties: Study on CHO cell cultures," *Biotechnology Progress*, vol. 41, no. 2, p. e3519, 2025.

[19] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert, "Constrained model predictive control: Stability and optimality," *Automatica*, vol. 36, no. 6, pp. 789–814, 2000.

[20] S. J. Qin and T. A. Badgwell, "A survey of industrial model predictive control technology," *Control Engineering Practice*, vol. 11, no. 7, pp. 733–764, 2003.

[21] G. Pannocchia and J. B. Rawlings, "Disturbance models for offset-free model-predictive control," *AIChE Journal*, vol. 49, no. 2, pp. 426–437, 2003.

[22] M. Banner, D. Ray, I. Ahanger, F. Zurlo, N. Hitchcock, R. Mistry, K. Kozakowska-McDonnell, M. Cheeks, R. Turner, J. Welsh, S. S. Farid, M. Thomas, and S. Goldrick, "Advanced glucose control strategies leveraging Raman spectroscopy for optimized mammalian cell culture manufacturing," *Biotechnology Progress*, vol. 42, no. 5, p. e88531, 2026.

[23] M. Brunner, J. Fricke, P. Kroll, and C. Herwig, "Investigation of the interactions of critical scale-up parameters (pH, pO_2 and pCO_2) on CHO batch performance and critical quality attributes," *Bioprocess and Biosystems Engineering*, vol. 40, no. 2, pp. 251–263, 2017.

[24] U.S. Food and Drug Administration, "PAT: A framework for innovative pharmaceutical development, manufacturing, and quality assurance," Guidance for Industry, Center for Drug Evaluation and Research, Center for Veterinary Medicine and Office of Regulatory Affairs, fda.gov, 2004.

[25] ——, "Automatic, mechanical, and electronic equipment," Code of Federal Regulations, Title 21, Section 211.68, revised as of April 1, 2024, U.S. Government Publishing Office, govinfo.gov, 2024.

[26] B. Hüpfl, A. Umprecht, B. Kozma, P. Ettinger, and O. Spadiut, "Limits and benefits of unsupervised transfer learning for Raman spectroscopy-based bioprocess monitoring," *Bioprocess and Biosystems Engineering*, vol. 49, no. 10, pp. 2583–2599, 2026.

[27] T. Kikuchi, S. Ohira, and H. Yamaguchi, "Viable cell density as an indicator for dynamic feeding strategy in fed-batch and perfusion CHO cell culture," *Scientific Reports*, vol. 15, no. 1, p. 44714, 2025.

8.10 Exercises

8.1 ○ Concept check · 2 min

A culture is fed by a daily bolus that brings the glucose of each morning sample up to a target when its result arrives four hours after the draw. The glucose must stay between 1 and 6 g/L. When can no choice of bolus target keep the glucose inside that band?

(a) When the glucose of a sample is already below the band's floor of 1 g/L at its draw

(b) When the stock pump cannot deliver a full bolus within the four hours of the analysis

(c) When the glucose the cells take between two reports is more than the band's width of 5 g/L

(d) When a sample's result lies more than 5 g/L below the target the bolus restores

8.2 ○ Concept check · 2 min

In case 3 the glucose is fed continuously on the capacitance probe and on the per-cell uptake measured between the last two samples, with no Raman. The culture is cooled on day 5, and by day 6.17 its true glucose has climbed to 6.09 g/L while the controller's estimate reads its 2 g/L target. What causes the climb?

(a) The capacitance probe stops responding once the cells are cooled, so the feed's cell count freezes

(b) The colder jacket slows the glucose pump, so the stock delivered falls short of what is commanded

(c) The PI trim on the glucose estimate raises the pump, because the estimate falls below its target

(d) The feed keeps the uptake measured before the shift while the cooled cells take less than half of it

8.3 Concept check · 2 min

The agitator of the culture runs at its full 140 rpm in the 1300 L of inoculation, and the bubbles leave the broth 87.5 % saturated with carbon dioxide. Which actuator of the DO sequence can still raise the carbon dioxide the sparge removes?

(a) The gas flow, because the removal is bounded by the moles of gas that pass through the broth

(b) The oxygen enrichment, because a richer gas leaves more room in each bubble for carbon dioxide

(c) The agitator, because a higher transfer coefficient brings the bubbles still closer to saturation

(d) None of them, because the pH loop sets the carbon dioxide removed through its own leg

8.4 Theory and calculation · 6 min

Fed continuously so that it eats what it wants, the default culture takes 0.53, 1.02, 1.83, 2.97 and 4.27 g/L of glucose in its first five days. Kept at 37 °C past the planned shift, it takes 5.24, 5.84 and 5.93 g/L in days 5 to 6, 6 to 7 and 7 to 8. The band is 1 to 6 g/L. Use the values as given, to two decimals.

(a) Compute the ratio of each of the first five days' uptake to the day before, and state the range as a percentage increase per day. *Checkpoint:* 1.44 to 1.92, so 44 to 92 % more each day.

(b) Apply the band condition of Equation (8.15) to the three warm days. By how much does each exceed the band's width? *Checkpoint:* By 0.24, 0.84 and 0.93 g/L: all three fail.

(c) Had the warm culture kept the smallest ratio of part (a) after day 5, what would it take in day 5 to 6? Compare with the measured value and name the terms of Equation (8.3) that explain the difference. *Checkpoint:* 6.14 g/L against 5.24 g/L measured.

8.5 Theory and calculation · 6 min

Lactate (90 g/mol) made by the cells spends bicarbonate, and the pH loop restores it with sodium carbonate while base is its active leg.

(a) Each millimole of lactate held at the pH setpoint adds itself and one sodium to the osmolality. Taking 11.1 mM per gram per liter of lactate, find the osmolality cost of 1 g/L of lactate. *Checkpoint:* About 22 mOsm/kg.

(b) Over a whole culture the base is half the lactate made, less half the ammonium made, plus half the change in the broth's bicarbonate. The default culture makes 49.3 mol of lactate net and 21.6 mol of ammonium, and its bicarbonate grows from 23.4 to 50.9 mol. Find the base. *Checkpoint:* 27.6 mol.

(c) Case 7, held at a 4.5 g/L glucose target, makes 178.1 mol of lactate and 12.6 mol of ammonium and ends with 11.2 mol of bicarbonate from the same start. Find its base, and say why the osmolality it causes stays after the cells take lactate back. *Checkpoint:* About 76.7 mol.

8.6 Theory and calculation · 7 min

Use Equations (8.9) and (8.10) with $k_L a_{ref} = 8\,\mathrm{h}^{-1}$, $\alpha_N = 1.8$, $\beta_Q = 0.5$, $k_{CO_2} = 0.89$, $\alpha = 0.0307\,\mathrm{mM/mmHg}$, $P = 760\,\mathrm{mmHg}$ and the 1300 L of inoculation.

(a) Compute the sparge k_La at 40 rpm and 15 L/min, and at 140 rpm and 15 L/min. *Checkpoint:* 0.325 and 3.10 h^{-1}.

(b) At 140 rpm and 15 L/min, compute $\dot{n}_g$, NTU and ε. *Checkpoint:* 0.669 mol/min, 2.08 and 0.875.

(c) Compute the coefficient $60\,000\,\dot{n}_g$ of the removal bound at 15 and 100 L/min, and state what it implies for the DO sequence. *Checkpoint:* About 40 200 and 268 000 mmol/h, a ratio of 6.67.

8.7 App challenge · 8 min

Open scenario 2, Daily bolus from the offline sample, at the default seed. Set the speed to 48 h/s, so a fourteen-day culture takes about 7 s. Each run stops by itself at the harvest. Read the tiles Glucose after day 2 (true), Titer at harvest (true) and Highest osmolality (true). Give hours within 0.5 h, titers within 0.02 g/L and osmolality within 2 mOsm/kg.

(a) Run as shipped, with the Glucose target at 4 g/L. *Checkpoint:* 34.8 h out, 4.14 g/L, 435 mOsm/kg.

(b) Press the reload button (Ⓡ) to restore the scenario, set the Glucose target on the FIC-201 card to 6.0 g/L before Run, and run to the harvest. *Checkpoint:* 0.7 h out, 2.73 g/L, 511 mOsm/kg.

(c) Explain why the higher target nearly clears the glucose tile and still loses a third of the antibody. *Checkpoint:* The glucose stays above the 2.5 g/L switch threshold.

8.8 App challenge · 8 min

Open scenario 4, The Raman trim, at the default seed and 48 h/s. For each run read the tiles Lactate switch (true), Titer at harvest (true), Base added and Highest osmolality (true). Give titers within 0.02 g/L, the switch day within 0.02, the base within 0.2 mol and osmolality within 2 mOsm/kg.

(a) Run as shipped, at the 2.0 g/L Glucose target, to the harvest. *Checkpoint:* Switch on day 6.32, 5.03 g/L, 27.6 mol, 393 mOsm/kg.

(b) Reload the scenario, set the Glucose target to 1.5 g/L before Run, and run to the harvest. *Checkpoint:* Day 5.10, 5.07 g/L, 26.7 mol, 390 mOsm/kg.

(c) Reload, set the Glucose target to 2.5 g/L, and run to the harvest. *Checkpoint:* No switch, 3.66 g/L, 55.9 mol, 467 mOsm/kg.

8.9 App challenge · 7 min

Open scenario 6, The temperature shift, early and late, at the default seed and 48 h/s. Read the tiles Titer at harvest (true), Viability at harvest (true) and Base added after each run. Give titers within 0.02 g/L, viability within 0.2 % and base within 0.2 mol.

(a) Run as shipped, with the Shift day on the Supervisor card at 5. *Checkpoint:* 5.03 g/L, 80.9 %, 27.6 mol.

(b) Reload, set the Shift day to 3 before Run, and run to the harvest. Then reload, set it to 8, and run to the harvest. *Checkpoint:* 4.33 g/L, 87.4 %, 3.1 mol; then 3.57 g/L, 68.2 %, 45.6 mol.

(c) Explain why the optimum lies in the middle, using the base and the viability you read. *Checkpoint:* Too few cells early, lactate, base and dying cells late.

8.10 Application area · 4 min

Case 9 harvests when a daily viability sample, drawn once a day and reported four hours later, reads below 65 %. The culture is 70.2 % viable when the day-17 sample is drawn and is harvested on day 18.17 at 54.8 % true viability with 7.40 g/L. A 75 % limit harvests on day 17.17 at 67.7 %. Why does each rule harvest well below its own limit, why does a plant care, and what would you change in the harvest rule?

8.11 Application area · 4 min

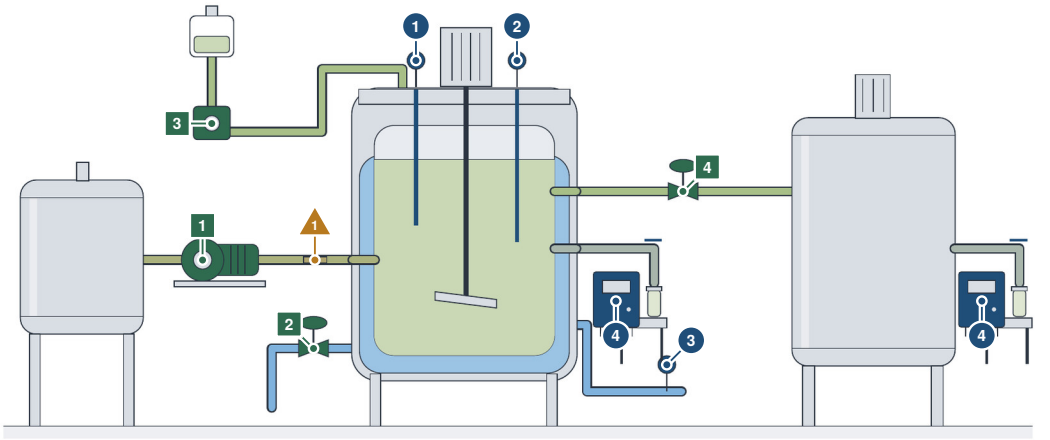
Case 5, DO and pH through one sparge, ships with the DO sequence set to Gas before O_2: agitation, then sparge flow, then oxygen enrichment. Against the order O_2 before gas, its culture reaches a highest osmolality of 382 against 393 mOsm/kg, uses 24.0 against 27.7 mol of base and makes 5.28 against 5.04 g/L (Table 8.5). Explain why the order of the DO sequence changes the osmolality. Then name two costs a plant pays for a higher sparge flow that this model leaves out, and one cost of the other order.

8.12 Review the agent · 5 min

An AI assistant reviews the trends of scenario 8, A pH probe drift seen only offline, run as shipped at the default seed. It reports: “The pH loop held its measurement at 7.00 for all fourteen days, and the culture used 27.2 mol of base against 27.6 mol for the reference culture of case 4, so the pH control is healthy. The titer of 4.50 g/L against 5.03 g/L and the cell peak of 12.22 against 14.78 million cells/mL therefore point to the seed or the medium. The next batch should start from a new vial.” Find the flaw in the reasoning, say what in the app shows it, and say how you would confirm the cause.

CHAPTER 9

Refolding Insulin Precursor When a Better Fold Can Cost the Campaign



The controller moves 1 the charge pump, which meters the dissolved protein in 2 the valve on the jacket utility 3 the pump that doses base to hold the pH 4 the valve that moves the folded lot on

The controller reads 1 the pH probe in the liquid 2 the temperature probe in the liquid 3 the temperature element on the jacket return 4 the two sample points, each read at a bench beside its vessel

What disturbs it 1 dissolved protein arriving weaker than the recipe assumes

Notice that the two benches, which alone say how much protein has folded, stand off the line, so between their reports every folded, aggregated and unfolded mass the controller uses is an estimate. Illustration, not to scale.

IN PLAIN TERMS **Folding a Protein Without Letting It Clump**

The main idea. Bacteria can be made to build the protein that turns into insulin. They pack it away as a useless lump, so the plant melts the lump and lets the protein take its proper shape in a big stirred tank.

The challenge. Each molecule either folds the right way or sticks to a neighbor and is lost. Sticking wins when the liquid is crowded, so the protein is added slowly, and then the tank is busy so long that the next batch spoils while it waits.

Why it matters. A plant is paid for the medicine it finishes in a week, not for the share of one tank that came out right.

THREE WAYS TO READ THIS CHAPTER

⊘	**Explore**	13 min	the first page, the history of the route (Section 9.2), the two fact segments, both try-it segments and the key points
🕮	**Study**	35 min	the whole chapter with its three pause points and the twelve exercises of Section 9.10
⛰	**Practice**	10 min	the first page, the ten measured cases (Section 9.6), Technology in Practice (Section 9.7), Limitations (Section 9.8) and the open challenge

The control question Unfolded insulin precursor either folds, at a rate first order in its concentration, or aggregates, at a rate second order in it, so a lot charged slowly keeps more protein than a lot charged at once. The refolding vessel is also the bottleneck of the recovery train, so the question is when a charge that folds more protein pays for the vessel hours it costs.

Learning objectives

- Compute the folded share of a bolus charge from the folding and aggregation constants, and predict how temperature and pH move it.
- Explain why the hours a lot occupies the refolding vessel, and not the yield of that lot, decide what a 72-hour campaign delivers.
- Correct a model of a vessel from an assay reported late, and stop a reaction whose product is consumed by the enzymes that made it.
- Choose a wash, a charge strategy, a refolding temperature and the two supervisory rules for a stated campaign, and defend the choice by the accepted insulin per campaign hour.
- State the conditions under which a larger harvest delivers less product, and the measurement that would settle the matter in a plant.

Prerequisites Sections 1.4.3 (competing first- and second-order kinetics and the fold they produce), 1.4.5 (a model corrected at the hour of the draw), 1.6.2 (velocity form) and 1.3 (loop action). Chapter 8 is the nearest neighbor: the same correction from a laboratory sample, on a plant faster than its samples.

The app https://apmonitor.com/apps/insulin-manufacturing/

See also Volume 8, Chapter 11 for the capture step this chapter declares as a recovery map, Volume 8, Chapter 2 for a feed metered to keep an unreacted store small, and Volume 8, Chapter 9 for a lot judged by a sampled check. Volume 5, Chapter 4 for steering on a grade that arrives hours late, and Volume 5, Chapter 6 for an upstream decision that reaches the product minutes later.

9.1 Introduction

An insulin fermentation report quotes grams of precursor per liter of broth, while the plant is paid for grams of insulin that pass their lot limits. In the simulated campaign of this chapter the cells make 3013 g of precursor-equivalent protein and a recipe modeled on a published pilot process isolates 217 g of acceptable insulin, one part in fourteen. Between the two stand a wash, a refolding vessel, an enzyme reaction that removes 53 % of every converted molecule's mass as fragments, and two purification steps. The largest loss the operator can move is in the refolding vessel. Unfolded precursor either folds, at a rate first order in its own concentration, or aggregates, at a rate second order in it [1], so a lot charged all at once loses more protein to aggregate than the same lot fed in slowly.

Two beliefs are what the simulator is designed to correct. The first is that more precursor at the harvest means more insulin at the end. A harvest at 22 h carries 3238 g, 11 % more than the 2905 g of a harvest at 16 h, and the campaign isolates 114.6 g instead of 217.3 g because the recovery train cannot take the lots before their clocks run out. The second is that the fold that keeps the most protein is the best fold. Charged as a bolus to 2.0 g/L, the first lot of 815 g folds 67.7 % and frees the vessel after 18.8 h. Fed by the scheduler under a ceiling of 0.35 g/L on the unfolded concentration it folds 84.7 %, but it holds the vessel for 39.0 h, 665 g of the harvest behind it expires at 52 h, and by the 72 h horizon the campaign has isolated nothing, against 112.7 g after the bolus.

The arrangement is general. Two reactions of different order that compete for one reactant favor the lower-order product when the reactant is kept dilute, which is why preparative refolding dilutes the denatured protein or adds it in pulses [2, 3], and aggregation of non-native protein is a multi-stage process whose kinetics set processing conditions across biotechnology products [4]. A fed-batch polymerization that meters its monomer, a crystallizer that doses its antisolvent and a fermenter fed below the rate that makes byproduct have the same shape. Volume 8, Chapter 2 meters a reagent into a batch for a different reason, to keep the heat release no faster than the feed, and the unreacted store

it builds plays the part the unfolded protein plays here. The scheduling half of the problem is also common. Production scheduling models and solution methods have found industrial use across the process industries [5], and biopharmaceutical campaigns are planned against fermentation titers that are themselves uncertain [6].

The published studies treat the two halves apart. Refolding studies optimize one vessel's yield over a fixed incubation [7], and campaign planning works at the level of batches and titers [6]. This chapter puts them on one screen. Its simulator carries the folding competition, the vessel occupancy and the hold clocks through a single 72-hour campaign, and the question it answers is when a concentration-aware charge pays for the vessel hours it costs, and how far that answer survives errors in the folding constants and the assay delay. The reduced folding law, the velocity-form proportional-integral (PI) loops and the estimator are established methods, and every comparison below is a set of measurements on one declared model: a 34-state model whose protein ledger closes to 1.8×10^{-13}, behind a measurement boundary that keeps every controller and rule away from the simulator's truth.

Section 9.2 follows insulin manufacture from extraction to refolding. Section 9.3 states the model, vessel by vessel, with its parameters and the evidence behind each one, and Section 9.4 checks the code against the identities it should satisfy and the folding law against data withheld from its fit. Section 9.5 describes the four loops, the feed scheduler, the estimator and the transfer and endpoint rules, and Section 9.6 measures ten cases, from the protein ledger to an infeasible harvest and a graded campaign. Section 9.7 describes the instruments and the platform a plant would use, Section 9.8 states what the model cannot be asked, and the exercises of Section 9.10 run the cases that carry the argument.

9.2 From Pancreatic Extract to a Folded Precursor

Insulin reached patients first as an extract of animal pancreas, and that route supplied the drug until recombinant human insulin arrived in the early 1980s [8].

> ⧖ HISTORY **From Pancreatic Extract to a Recombinant Route (1921 to 1983)**
>
> Insulin entered medicine as an extract. Pancreatic extracts that lowered blood sugar in animals had been made for two decades before 1921, but impurities and toxic reactions kept them from patients, and the Toronto work of 1921 and 1922 produced an extract from whole beef pancreas that treated people with diabetes once its purification had been worked out [9]. Extraction from animal pancreas then supplied insulin until recombinant human insulin arrived in the early 1980s [8]. Human insulin from recombinant deoxyribonucleic acid technology became the first commercial health-care product of that technology, and its development began before federal guidelines for large-scale recombinant work existed [8].

The molecule's own biosynthesis pointed to the route used now. In 1967 islet tissue was shown to make a larger single-chain protein that radioactivity passed into insulin, and the name proinsulin was proposed for it [10]. In 1971 proinsulin was converted to insulin in vitro with trypsin and carboxypeptidase B [11], the two enzymes the pilot process simulated here still uses [7]. Recombinant expression followed within a decade. Synthetic genes for the A and B chains were expressed in *E. coli* as fusions to β-galactosidase, the chains were cleaved, purified, mixed, reduced and reoxidized, and insulin was detected [12]. Production later moved from two separately made chains to a proinsulin route on an *E. coli* fermentation [13]. Insulin products commanded annual global sales above $4.5 billion by 2005 [14], and expression hosts and capacity remain a subject of review because demand is expected to outgrow current manufacturing [15].

The cost of the bacterial route is that the protein arrives as an inclusion body. Inclusion bodies are dense deposits of inactive protein that can be isolated by solid-liquid separation, solubilized and refolded in vitro, and folding protocols control the protein concentration, disulfide formation, temperature, pH and solvent to suppress unproductive side reactions [16]. Refolding of inclusion-body protein is cumbersome, often recovers poorly and accounts for the major cost of producing a

Table 9.1 Variables by Role, With Defaults and Ranges

Symbol	Role	Default	Range
$\hat{U}$	CV, estimated unfolded (FIC-304)	$U_{tgt} = 0.35$	0.05–1.5 g/L
T	CV, vessel temperature (TIC-301)	15 °C	8–30 °C
T_j	CV, jacket (TIC-302)	from TIC-301	2–40 °C
pH	CV, vessel pH (AIC-303)	10.6	9–11.5
F	MV, charge pump	—	0–400 L/h
$T_{j,SP}$, T_{sup}	MV, jacket setpoint, utility	—	2–40, −5–45 °C
u_b	MV, base pump (0.25 mol/h at 100 %)	—	0–100 %
t_{ind}, t_{harv}	decision, induction, harvest	6, 16 h	2–14, 6–30 h
V_{lot}, w	decision, lot, wash volumes	250 L, 3	100–350, 0–6
C_{tgt}	decision, charge concentration	2.0 g/L	0.5–4 g/L
c_E, T_{conv}	decision, enzyme dose, temperature	1, 15 °C	0.25–3, 8–30
rules	decision, charge, transfer, endpoint	bolus, time, time	three each
x_a	DV, acid load multiple	1	0–6
κ	DV, chiller capacity	100 %	20–100
s	DV, solubilization strength	100 %	40–120
δ_{pH}, δ_X	DV, pH and biomass probe drift	0, 0	±0.1/h, −3–6 %/h
V_f, X, S, P	states, fermenter	at 16 h: 744 L, 27.1, —, 3.90	kg/m^3
m_H, I_H, V_H	states, harvest tank	2905 g	—
P_A, I_A, V_A, w	states, recovery vessel	—	—
U_S, I_S, V_S	states, solubilized tank	—	—
VU, VN, VA, V	states, refolding (lot 1)	815 g, 408 L	—
I_R, b, T, T_j	states, refolding impurity, base	—	—
n_N, n_I, n_G, n_D	states, conversion (moles)	—	—
I_C, V_C	states, conversion impurity, volume	—	—

recombinant protein in *E. coli* [3]. The quantitative basis for controlling it came in 1991, when the yield of native protein upon refolding was explained by the rates of competing first-order folding and second-order aggregation, so that aggregation dominates at high protein concentration [1]. Dilution, pulsed addition and column-based methods followed as preparative practice [2, 3]. The pilot process that calibrates this simulator refolds a fused precursor directly, without sulfonation, at 1.5 g/L and pH 10.6 for 16 h at 15 °C, and from 200 L of broth it isolates 138.2 g of insulin on average over nine batches [7].

9.3 The Route as Modeled

The simulated route is a proposed teaching route assembled from public evidence, not a reproduction of any manufacturer's process, as shown in Figure 9.1. A fed-batch fermenter FERM-101 makes precursor in inclusion bodies. At the harvest the broth goes to a hold tank with a stability clock. Recovery vessel REC-201 takes one lot at a time through lysis, an inclusion-body wash and solubilization, and passes it to a solubilized hold tank, which frees REC-201 for the next lot. The refolding vessel is charged from that tank through the charge pump of FIC-304, its temperature is held by the cascade TIC-301 to TIC-302 and its pH by AIC-303 on a base pump. When folding is judged complete the lot is acid-precipitated into the conversion vessel, where enzymes cut the precursor to insulin and then slowly over-process it. A declared capture and polish map and a gate on purity and host-cell protein turn each lot into accepted or rejected bulk. Three vessels hold one lot each, so the schedule is a decision rather than a queue.

The variables of Table 9.1 are classified by role. Thirty-four states are integrated, 28 physical and 6 ledgers that accumulate production and losses so that conservation can be checked exactly. The physics is in SI units (seconds, cubic meters, kilograms, moles), and the interface displays hours, liters, grams and g/L, which equals kg/m^3.

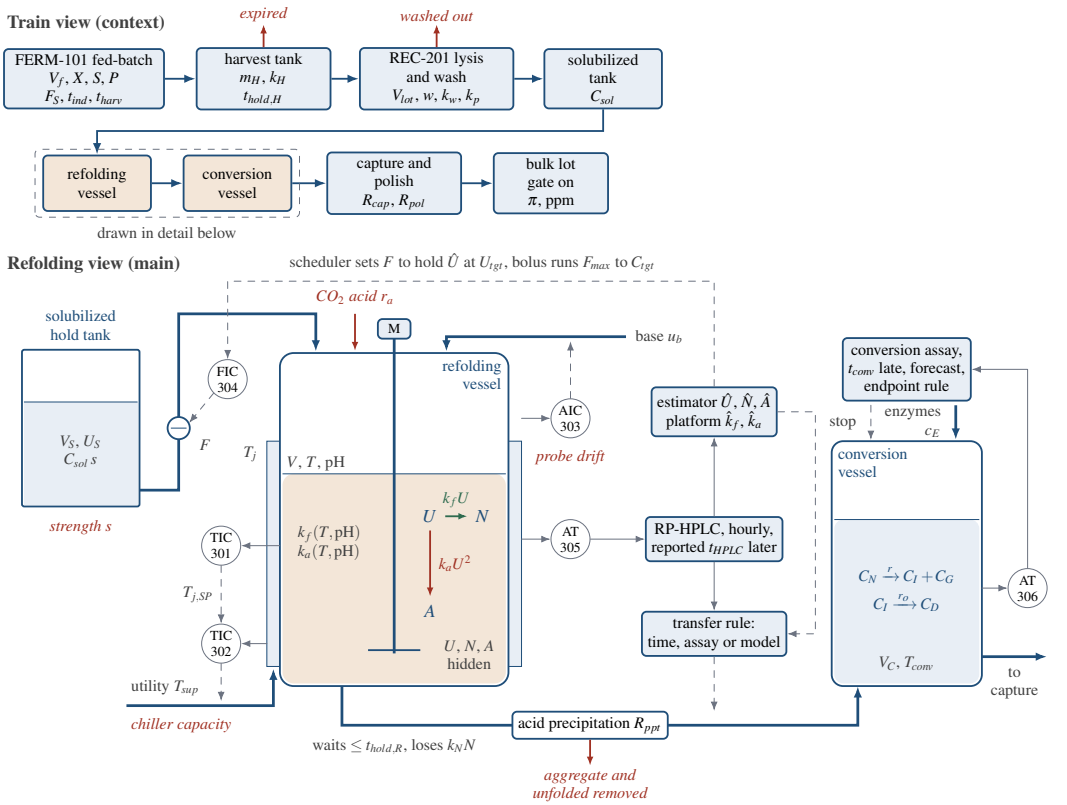


Figure 9.1 Insulin recovery and refolding as modeled. The upper strip is the Train view, from fermenter to bulk lot, with the two vessels drawn in detail below. The lower panel follows the Refolding view: the solubilized hold tank at V_S and concentration $C_{sol}s$ feeds the charge F, the refolding vessel of volume V holds the hidden states U, N and A with folding $k_f U$ and aggregation $k_a U^2$, TIC-301 and TIC-302 cascade from T to the jacket T_j and the utility T_{sup}, and AIC-303 doses base u_b against the acid load r_a. The RP-HPLC assay at AT-305 reports t_{HPLC} late to the estimator of $\hat{U}$, $\hat{N}$ and $\hat{A}$, which sets the scheduler's charge and the transfer rule. The conversion vessel holds C_N, C_I, the fragments C_G and the over-processed C_D, with its own delayed assay and endpoint rule. Disturbances are in red italics. Process model only, not to scale.

9.3.1 The Fermenter That Supplies the Load

The fermenter runs on one substrate with a declared oxygen and cooling capacity, and induction switches on a production rate that decays while most of the growth stops, as shown in Equation (9.1).

$$
\begin{aligned}
\frac{\mathrm{d}V_f}{\mathrm{d}t} &= F_S, \qquad \frac{\mathrm{d}X}{\mathrm{d}t} = \left(\mu - k_d - \frac{F_S}{V_f}\right)X \\
\frac{\mathrm{d}S}{\mathrm{d}t} &= \frac{F_S}{V_f}(S_f - S) - \left(\frac{\mu}{Y_{X/S}} + m_S + \frac{q_P}{Y_{P/S}}\right)X \\
\frac{\mathrm{d}P}{\mathrm{d}t} &= q_P X - \left(k_{P,loss} + \frac{F_S}{V_f}\right)P \\
\mu &= \mu_{max}\frac{S}{K_S + S} f_O f_b^{\iota}, \qquad f_O = \min\left(1, \frac{X_{cap}}{X}\right) \\
q_P &= \iota\, q_{P,max}\, \mathrm{e}^{-(t - t_{ind})/\tau_q}\frac{S}{K_S + S} f_O
\end{aligned}
\tag{9.1}
$$

where ι is 0 before induction and 1 after it, and the feed stops at the working volume $V_{f,max}$. The production $q_P X V_f$ is integrated as the ledger of everything made. Every constant in Equation (9.1) is a teaching assumption. The one anchored number is the load, 3.87 g of precursor-equivalent protein per liter of broth, derived from the pilot's Table 3 as 539.8 g refolded at a 69.7 % step yield from 200 L [7]. That is protein charged to refolding per liter of broth, and the model places it at the harvest, 3.90 g/L in 744 L at 16 h, so after its own hold and wash losses a 250 L lot charges 3.26 g per liter of broth into refolding. The burden $f_b = 0.05$ leaves 5 % of the growth rate after induction, which gives case 2 an interior best hour. The pilot's own culture kept growing after its temperature-shift induction (its Fig. 1), so the value is a strong-burden teaching choice. With $f_b = 0.1$ the best induction hour tried is still near 6 h, and with 0.2 or more the earliest hour tried delivers the most.

9.3.2 Harvest Hold, Wash and Solubilization

At t_{harv} the broth moves to the hold tank with $m_H = PV_f$ of precursor and $I_H = f_{HCP} X V_f$ of host-cell protein. The tank loses precursor at $\dot{m}_H = -k_H m_H$, and whatever remains $t_{hold,H}$ after the harvest is discarded. A lot of $\min(V_{lot}, V_{A,max}, V_H)$ is charged to REC-201 when it and the solubilized tank are free, lysed for t_{lys}, and washed at constant volume, with w the wash volumes passed, as shown in Equation (9.2).

$$\frac{\mathrm{d}P_A}{\mathrm{d}w} = -k_p P_A, \qquad \frac{\mathrm{d}I_A}{\mathrm{d}w} = -k_w I_A, \qquad \frac{\mathrm{d}w}{\mathrm{d}t} = \frac{1}{t_w} \tag{9.2}$$

Three wash volumes leave 2.7 % of the host-cell protein and 84 % of the precursor. Solubilization then sets the lot's volume to $V_A = P_A/(C_{sol} s)$, with s the strength the solubilization reaches, holds it for t_{sol}, and transfers it to the solubilized tank.

9.3.3 The Refolding Competition

The refolding vessel is first charged with buffer to $V_0 = \max\left(0.02, \min(V_{R,max}, m_S/C_{tgt}) - V_S\right)$, so that the lot of mass m_S in volume V_S reaches C_{tgt} when it is all in. The charge then enters at F with concentration $c_U = U_S/V_S$, and the precursor-equivalent masses of unfolded (U), correctly folded (N) and aggregated (A) protein follow Equation (9.3).

$$\begin{aligned} \frac{\mathrm{d}(VU)}{\mathrm{d}t} &= F c_U - V\left(k_f U + k_a U^2\right) \\ \frac{\mathrm{d}(VN)}{\mathrm{d}t} &= V k_f U - h k_N V N, \qquad \frac{\mathrm{d}(VA)}{\mathrm{d}t} = V k_a U^2 \\ \frac{\mathrm{d}V}{\mathrm{d}t} &= F, \qquad \frac{\mathrm{d}V_S}{\mathrm{d}t} = -F, \qquad \frac{\mathrm{d}U_S}{\mathrm{d}t} = -F c_U \end{aligned} \tag{9.3}$$

where $k_a U^2$ is the mass disappearance rate of unfolded protein, so no factor of two appears, and $h = 1$ only while a folded lot waits for the conversion vessel. With the feed off, $V(U + N + A)$ is constant, and a bolus at constant volume folds the fraction

$$\frac{N_\infty}{U_0} = \frac{\ln(1+\alpha)}{\alpha}, \qquad \alpha = \frac{k_a U_0}{k_f} \tag{9.4}$$

which depends on the charge only through α. Feeding the same lot slowly keeps U below U_0 and beats this bound, at the cost of vessel hours. Temperature and pH act on the two constants as shown in Equation (9.5).

$$\begin{aligned} k_f &= k_{f,15}\, \mathrm{e}^{\theta_f (T-15)} \exp\left[-\left(\frac{\mathrm{pH} - \mathrm{pH}^*}{\sigma_{pH}} \right)^2 \right] \\ k_a &= k_{a,15}\, \mathrm{e}^{\theta_a (T-15)}\, \mathrm{e}^{\lambda_{pH} \max(0,\ \mathrm{pH}^* - \mathrm{pH})} \end{aligned} \tag{9.5}$$

with T in °C. Folding doubles per 10 K ($Q_{10} = 2.0$) and aggregation nearly quadruples ($Q_{10} = 3.86$), so a warmer vessel finishes sooner and keeps less. At 1.5 g/L the bolus has $\alpha = 0.72$ and folds 75.3 % at 15 °C, and $\alpha = 1.39$ with 62.7 % at 25 °C.

Pause. Try Exercise 9.1 before reading on.

> **DID YOU KNOW? A Fed Charge in a Commercial Tank**
>
> A commercial refolding step for a two-chain immunotoxin is run in jacketed stainless steel tanks of 950 L nominal volume, and at the 5 L bench scale, with every other condition held, feeding the solubilized protein in place of adding it at once raised the volumetric productivity of the refold by 34 % [17].

9.3.4 Vessel Temperature and pH

The vessel's energy balance carries the jacket, the agitator, the room and the incoming charge, and the jacket relaxes toward the applied utility temperature, as shown in Equation (9.6).

$$\begin{aligned} \frac{\mathrm{d}T}{\mathrm{d}t} &= \frac{UA(T_j - T) + P_{mix} - UA_{amb}(T - T_{amb})}{C_{ves} + \rho c_p V} \\ &\quad + \frac{\rho c_p F (T_{feed} - T)}{C_{ves} + \rho c_p V} \\ C_j \frac{\mathrm{d}T_j}{\mathrm{d}t} &= \frac{C_j}{\tau_j}\left(T_{sup}^{app} - T_j\right) - UA(T_j - T) \\ T_{sup}^{app} &= \max\left(T_{sup},\ 45 - 50\kappa\right) \end{aligned} \tag{9.6}$$

where P_{mix} acts only while there is liquid and the empty vessel keeps its own heat capacity C_{ves}. The chiller event of the challenge sets $\kappa = 0.6$, so the supply can go no colder than 15 °C. The pH is a linear titration of the net base b about the buffer capacity, as shown in Equation (9.7).

$$\mathrm{pH} = \mathrm{pH}^{*} + \frac{b}{\beta V}, \qquad \frac{\mathrm{d}b}{\mathrm{d}t} = q_{b,max}\frac{u_b}{100} - x_a r_a V \tag{9.7}$$

clipped to 6–13, where r_a is the acid load of carbon dioxide absorbed by the alkaline buffer and $q_{b,max} = 0.25\,\mathrm{mol/h}$ of 1 M base.

9.3.5 Precipitation, Conversion and Isolation

At transfer, acid precipitation keeps R_{ppt} of the folded protein, the fraction c_A of the aggregate and $c_{I,ppt}$ of the host-cell protein, and removes the unfolded protein. The conversion vessel counts molecules, as shown in Equation (9.8).

$$\begin{aligned} r &= \frac{k_{cat} c_E f_T C_N}{K_M + C_N}, \qquad r_o = \frac{\rho\, k_{cat} c_E f_T C_I}{K_{M,o} + C_I} \\ \frac{\mathrm{d}n_N}{\mathrm{d}t} &= -rV_C, \qquad \frac{\mathrm{d}n_I}{\mathrm{d}t} = (r - r_o)V_C \\ \frac{\mathrm{d}n_G}{\mathrm{d}t} &= rV_C, \qquad \frac{\mathrm{d}n_D}{\mathrm{d}t} = r_o V_C \end{aligned} \tag{9.8}$$

where $C = n/V_C$, $f_T = \mathrm{e}^{\theta_E (T_{conv} - 15)}$, c_E is the dose as a multiple of nominal, and $k_{cat} c_{E,0}$ is a maximum rate in $\mathrm{mol\,m^{-3}\,s^{-1}}$. One precursor molecule gives one insulin molecule and its cleavage fragments n_G, which carry $MW_P - MW_I = 6.59\,\mathrm{kg/mol}$ of every 12.4 kg/mol, and the over-processed species n_D keeps insulin's mass so that the mass ledger closes exactly. Reaction water is not counted. When the endpoint rule stops the reaction, the declared map gives the lot of Equation (9.9).

$$\begin{aligned} m_{out} &= MW_I\, n_I\, R_{cap} R_{pol} \\ m_{rel} &= MW_P\, n_N + MW_I\, n_D + A_C \\ \pi &= \frac{m_{out}}{m_{out} + c_{rel} m_{rel} + c_{HCP} I_C} \\ \mathrm{ppm} &= 10^6 \frac{c_{HCP} I_C}{m_{out}} \end{aligned} \tag{9.9}$$

A lot is accepted when $\pi \geq 95.0\,\%$ and the host-cell protein is at most 1000 ppm. The carried-over related substances and host-cell protein are booked as purification losses, and accepted mass is insulin only. These limits are teaching values. Real acceptance criteria come from the product's specification, the pharmacopeia and regulatory guidance such as ICH Q6B [18], and a simulated pass is not batch release. The scorecard's lead tile is $\dot{m} = m_{acc}/t_{camp}$, accepted insulin divided by the whole 72 h horizon, so stopping early cannot flatter it, and unprocessed, expired and in-train material are reported beside it.

9.3.6 Instruments

Only the instruments reach a controller. The vessel and jacket temperatures pass first-order lags of 60 and 30 s with 0.05 K of noise, the pH probe a 120 s lag, 0.008 of noise and a drift δ_{pH}, the fermenter's optical probe 4 % noise and a gain drift, and the vessel weights 0.2 %. Two sampled analyzers of the shared framework (Section 1.6) carry the delayed assays. The RP-HPLC assay of the refolding vessel is drawn every hour and reported 0.75 h later with the total protein ($\sigma = 0.03\,\mathrm{g/L}$) and the folded, aggregated and unfolded shares (σ 0.012, 0.010, 0.012). The conversion assay is drawn every 30 min and reported 30 min later with insulin, precursor and related substances (σ 0.012, 0.010, 0.012 g/L). Each reading holds until the next report, and both analyzers are reset when their vessel is charged.

9.3.7 Parameters, Evidence and Numerics

Every parameter is in Table 9.2 with its source label. A value marked S is read in a source, F is fitted to a source trend, D is derived by arithmetic on source numbers, and T is a teaching assumption. Every parameter is adjustable in the application's parameter panel. Inside the engine of Section 1.6.1, states advance by fourth-order Runge-Kutta steps of $\Delta t = 5\,\mathrm{s}$, the controller scan is 30 s, and one row a minute is logged. The fastest dynamics are the jacket turnover (300 s) and the probes, while folding and conversion run over hours. Every irreversible event (induction, harvest, expiry, each lot transfer, the endpoint and isolation) happens once at a step boundary after the integration, and the right-hand side is pure. Each case starts with the fermenter at $V_{f,0} = 0.60\,\mathrm{m^3}$, $X_0 = 6\,\mathrm{kg/m^3}$ and $S_0 = 1.5\,\mathrm{kg/m^3}$, every downstream vessel empty, and the refolding vessel and jacket at the temperature setpoint.

9.3.8 Assumptions

The refolding vessel is one well-mixed liquid. A real charge enters at a point where the local unfolded concentration exceeds the vessel average, which is where a second-order term matters most, so the model can show only the feed rate among the three levers a plant uses (charge point, agitator and rate). The competition has one unfolded species, one correctly folded species and an irreversible aggregate, with no disulfide intermediates or recoverable misfolded monomer. The concentration dependence is the declared second-order form, not a fit (Section 9.4). Conversion is one effective enzyme reaction with one over-processing path, while the pilot uses two enzymes in sequence that make an intermediate this model does not carry. The pH model is a linear titration valid for a few tenths of a unit about the optimum. Everything upstream of the harvest is a teaching model of no particular strain, and purification is a declared recovery and selectivity map, because preparative chromatography is the subject of Volume 8, Chapter 11, where the capacity that decides when to take a column off is itself unmeasured.

9.4 Code Verification and Agreement With the Pilot Study

Code verification and agreement with the literature are kept apart here. The first asks whether the program solves the stated equations, and the second whether the stated equations reproduce a published trend. No claim is made about what a reader learns from either. The identities of the reduced model were checked independently of the application's own tests on six disturbed runs and are listed in Table 9.3. The ledger relates everything the cells made to everything in a vessel, lost, expired or isolated, and it closes to 1.8×10^{-13} relative, or 1.8×10^{-10} in the overloaded case where both hold limits expire.

Table 9.2 Parameters and Their Evidence. S Source, F Fitted, D Derived, T Teaching Assumption

Symbol	Meaning	Value	Unit	Ev.
$k_{f,15}$	folding constant, 15 °C, pH 10.6	4.61×10^{-5}	s^{-1}	F
$k_{a,15}$	aggregation constant at 1.5 g/L	2.21×10^{-5}	$m^3/(kg\ s)$	F
θ_f	folding temperature coefficient	0.0693	K^{-1}	T
θ_a	aggregation temperature coefficient	0.1351	K^{-1}	F
pH^*, σ_{pH}	folding optimum, width	10.6, 1.0	—	S, T
λ_{pH}	aggregation rise below pH^*	1.2	per pH	T
C_{sol}	solubilized concentration	10	kg/m^3	T
$V_{R,max}$, $V_{C,max}$	refolding, conversion volumes	1.0, 1.2	m^3	T
$k_{cat}c_{E,0}$	maximum conversion rate	4.61×10^{-5}	$mol/(m^3\ s)$	F
ρ	over-processing relative rate	0.15	—	T
K_M, $K_{M,o}$	Michaelis constants	0.5, 0.5	mol/m^3	T
θ_E	enzyme temperature coefficient	0.0693	K^{-1}	T
MW_P, MW_I	precursor (112 residues), insulin	12.4, 5.808	kg/mol	D, S
R_{ppt}	folded protein kept at precipitation	0.993	—	S
c_A, $c_{I,ppt}$	aggregate, host protein carried	0.05, 0.40	—	T
R_{cap}, R_{pol}	capture, polish recovery	0.90, 0.85	—	T
c_{rel}, c_{HCP}	carried into the bulk	0.040, 0.005	—	T
k_w, k_p	wash removal, precursor loss	1.2, 0.06	per volume	T
t_{lys}, t_w, t_{sol}, t_{turn}	lysis, per wash, solubilize, turnaround	0.5, 0.5, 1, 0.5	h	T
$V_{A,max}$	recovery vessel volume	0.35	m^3	T
k_H, k_N	losses in harvest hold, while waiting	0.010, 0.012	h^{-1}	T
$t_{hold,H}$, $t_{hold,R}$	harvest, folded-lot hold limits	36, 10	h	T
β, r_a	buffer capacity, acid load	2.0, 0.09	mol/m^3 per pH, per h	T
UA, UA_{amb}	jacket, room conductance	2000, 120	W/K	T
C_j, C_{ves}, τ_j	jacket, shell capacity, turnover	0.209, 0.30, 300	MJ/K, s	T
P_{mix}, T_{amb}, T_{feed}	agitator, room, feed	300, 21, 15	W, °C	T
μ_{max}, K_S	growth, half-saturation	0.52, 0.05	h^{-1}, kg/m^3	T
$Y_{X/S}$, $Y_{P/S}$, m_S	yields, maintenance	0.42, 0.12, 0.035	–, –, h^{-1}	T
$q_{P,max}$, τ_q, f_b	production, decay, burden	0.0444, 5, 0.05	h^{-1}, h, –	T
X_{cap}, k_d, $k_{P,loss}$	capacity, death, loss	38, 0.004, 0.006	kg/m^3, h^{-1}	T
S_f, f_{HCP}, $V_{f,max}$	feed, host protein, volume	600, 0.085, 1.0	kg/m^3, –, m^3	T
P_H	precursor at harvest, per broth	3.87	kg/m^3	D
π_{min}, HCP_{max}	lot limits	95.0, 1000	%, ppm	T
t_{HPLC}, t_{conv}	assay delays	0.75, 0.5	h	T

Table 9.3 Code Verification and Agreement With the Pilot Study

Check	Result
Code verification	
protein ledger, six disturbed campaigns	$\leq 1.8 \times 10^{-13}$ (1.8×10^{-10} case 9)
$V(U+N+A)$ with the feed off, 20 h	$< 10^{-12}$ relative
$\dot{V} = F$, vessel mass equals fed less misfolded	$< 10^{-12}$ relative
$k_a = 0$: $U = U_0 e^{-k_f t}$	$< 10^{-9}$ relative
bolus against Equation (9.4), 15, 20, 25 °C	$< 3 \times 10^{-4}$ in the fraction
moles and fragment mass in conversion	$< 10^{-12}$ relative
wash against e^{-kw}, every lot	3.8×10^{-13}
Δt halved, challenge (suggested)	5.565 and 5.565 g/h
playback in 60 and 3600 s chunks	no difference
Fitted, so code verification of the fit	
bolus, 1.5 g/L, 15 and 25 °C	75.31, 62.67 % (source 75.3, 62.8)
conversion, fixed 16 h, 15 °C	56.6 % molar (source 56.6)
Withheld from the fit	
Fig. 4A, eight digitized points	within 0.11 of the 20 h value
Table 2 at 1.5–2.0 g/L	71.2, 66.4 against 68.0–74.2, 57.7–66.0
Table 2 at 0.5–1.0 g/L	83.9, 76.9 against 53.6–66.0 (opposite)

The pilot study gives three numbers that fix the fitted constants. Refolding at 15 and 25 °C yielded 75.3 and 62.8 %, with the concentration of that pair not stated [7]. The fit assumes the article's optimum of 1.5 g/L, so $k_{a,15}$ scales with that assumption and would be 0.75 times as large at 2.0 g/L.

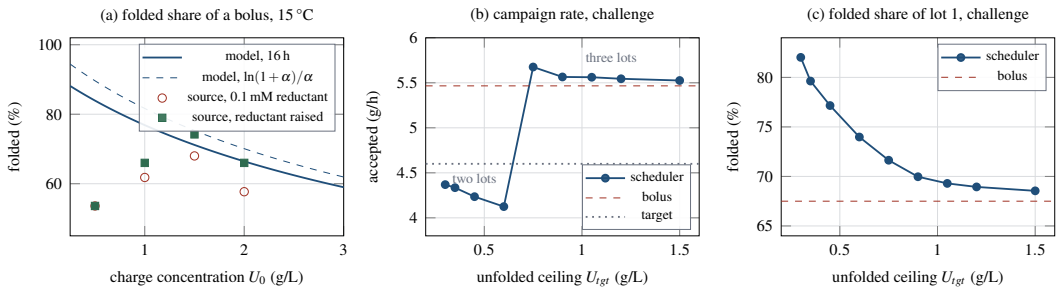


Figure 9.2 The folding law against its source and the campaign it drives. (a) The model's 16 h bolus yield at 15 °C (solid) and its closed form (dashed) against the two concentration series of the pilot study's Table 2, which were withheld from the fit (open circles at constant reductant, filled squares with the reductant raised). (b) Accepted insulin per campaign hour in the challenge with the suggested settings and the scheduler's ceiling U_{tgt} varied (solid, filled circles), against the bolus (dashed) and the target (dotted). (c) The first lot's folded share in the same runs, with the bolus again dashed. Below 0.75 g/L each lot folds better but only two lots fit. At 0.5 g/L the two source series share one point.

The constant $k_{f,15}$ is fitted to the statement that the 15 °C reaction was complete after 15 h [7], and the model reaches 95 % of its final folded mass in 16.5 h. The pilot's enzyme step converted 56.6 % of the refolded precursor molecules (25.8 % by mass, scaled by its residue ratio of 51 to 112), and $k_{cat}c_{E,0}$ is fitted so that a 16 h reaction converts the same, given $\rho = 0.15$. One end-point yield cannot fix both constants, so ρ is a teaching value, and with it the insulin peaks at 72.4 % of the molecules after 7.49 h.

Two comparisons were withheld from the fit. The article's Fig. 4A, sampled every 1.5 h for 20 h and digitized to about 10 g, reaches 95 % of its 20 h value at about 12 h at 15 °C, against 14.0 h in the model, and the eight digitized fractions agree with the model within 0.11. The source curve is nearly linear for 9 h where the first-order model is front-loaded. The second check is the concentration series of the article's Table 2, as shown in Figure 9.2a. In both of its series, one at constant reductant and one with the reductant raised with the protein, the yield rises from 0.5 g/L to 1.5 g/L and falls at 2.0 g/L. The model's 16 h yield falls all the way, from 83.9 to 66.4 %. It agrees with the raised-reductant series at 1.5 and 2.0 g/L (71.2 against 74.2, 66.4 against 66.0) and runs opposite to both series below 1.5 g/L. The second-order law is the plan's declared reduced model and the mechanism of the cases, so it is kept, and every number the simulator gives outside 1.5–2 g/L and 10–30 °C is an extrapolation of a teaching model. These comparisons are agreement with one study's data, not validation of the model for a plant.

9.5 The Controllers and What They Read

The four loops are the velocity-form PID of Section 1.6.2 with the derivative off, executed every 30 s on the measurement packet. With $e_k = \mathrm{SP}_k - \mathrm{PV}_k$, the law is Equations (1.12) to (1.15) written for this chapter, the running sum S_k of Chapter 1 appearing here as s_k and its integral I_k carrying the anti-windup term.

$$\begin{aligned} s_k &= s_{k-1} + K_c(e_k - e_{k-1}), \qquad \hat{u}_k = u_0 + s_k + I_{k-1} \\ u_k &= \min\left(u_{max},\ \max(u_{min},\ \hat{u}_k)\right) \\ I_k &= I_{k-1} + \left[\frac{K_c}{\tau_I}e_k + \frac{u_k - \hat{u}_k}{\tau_I}\right]\Delta t_c \end{aligned} \tag{9.10}$$

where u_0 is the output at which the loop is initialized and $\Delta t_c = 30\,\mathrm{s}$, with back-calculation anti-windup at a tracking time equal to τ_I and the clamp on the emitted output only. In Manual each

loop tracks the operator's output, so transfers are bumpless. TIC-301 sets the jacket setpoint within 2–40 °C at $K_c = 4\,\mathrm{K/K}$ and $\tau_I = 30\,\mathrm{min}$. TIC-302 sets the requested utility supply within −5–45 °C at $K_c = 3\,\mathrm{K/K}$ and $\tau_I = 5\,\mathrm{min}$, and its card shows the requested supply beside the applied one of Equation (9.6) when the chiller is short. AIC-303 drives the base pump within 0–100 % at $K_c = 150\,\%$ per pH and $\tau_I = 30\,\mathrm{min}$. All three are reverse acting, with action $a = +1$ in the convention of Section 1.3, because a warmer jacket warms the vessel, a warmer utility supply warms the jacket and more base raises the pH. TIC-301 and TIC-302 are the outer and inner halves of one cascade, so the inner loop is tuned six times faster than the outer one.

9.5.1 Three Ways to Charge

FIC-304 charges the solubilized lot by one of three strategies. A bolus runs the pump at $F_{max} = 400\,\mathrm{L/h}$ until the tank is empty, and the first lot of case 4 is in after 0.23 h. A fixed rate runs at F_{set}, 60 L/h by default. The scheduler holds the estimated unfolded concentration at its ceiling with a feedforward and a PI trim, as shown in Equation (9.11).

$$\begin{aligned} F_{FF} &= \frac{V\left(\hat{k}_f U_{tgt} + \hat{k}_a U_{tgt}^2\right)}{C_{sol}} \\ F^* &= \max\left(0,\; F_{FF} + \mathrm{PI}(U_{tgt} - \hat{U})\right) \\ F &= \min\left(F^*,\; F_{max},\; \frac{V_{fin} - V}{\Delta t},\; \frac{V_S}{\Delta t}\right) \end{aligned} \tag{9.11}$$

where $\hat{k}_f$ and $\hat{k}_a$ are Equation (9.5) evaluated with the platform's declared constants at the measured temperature and pH, the trim has $K_c = 100\,\mathrm{L/h}$ per g/L and $\tau_I = 30\,\mathrm{min}$, and its output is bounded to $[-F_{FF},\; F_{max} - F_{FF}]$. The volume V_{fin} is the one the charge will reach. The scheduler is a feedforward on a model with feedback on an estimate, and none of the receding-horizon machinery of Section 1.6.3 is used: no objective is minimized and nothing is predicted beyond the present rate. An advanced option stretches the charge so that the lot is ready when the conversion vessel is expected to be free. It never engages in a shipped case or in the challenge, because the conversion vessel is always free first.

> **TRY IT IN THE APP** **4 · A Bolus Charge Against a Fed One**
>
> Load the case and press run. On the Refolding view, watch the competition panel while the first lot goes in as a bolus: folding and aggregation both start at once, and the aggregation rate falls away only as the unfolded protein is used up. The chapter reports 67.7 % of this lot folded.

9.5.2 The Estimator

The estimator integrates Equation (9.3) on the commanded feed with the platform's constants, and corrects it with each RP-HPLC report against the value it had when that sample was drawn, which is the sampled correction of Section 1.4.5 with a fixed weight in place of the logarithmic bias of Equation (1.8), as shown in Equation (9.12).

$$\begin{aligned} \frac{\mathrm{d}(\hat{V}\hat{U})}{\mathrm{d}t} &= F C_{sol} - \hat{V}\left(\hat{k}_f \hat{U} + \hat{k}_a \hat{U}^2\right) \\ \hat{U} &\leftarrow \hat{U} + g\left(U_{assay}(t_s) - \hat{U}(t_s)\right), \qquad g = 0.6 \\ \hat{N}_{plat} &= \hat{V}\hat{U}\,\frac{\ln(1+\hat{\alpha})}{\hat{\alpha}}, \qquad \hat{\alpha} = \frac{\hat{k}_a \hat{U}}{\hat{k}_f} \end{aligned} \tag{9.12}$$

where t_s is the sample time, C_{sol} is the design value (so a weak solubilization is unknown to the estimator), $\hat{N}$ and $\hat{A}$ are corrected the same way, and the assay's masses use the vessel volume measured at t_s. The estimate history is stored every 300 s. A conversion estimate on Equation (9.8) is corrected the same way by each conversion report, and after each report it is integrated forward in

Table 9.4 The Ten Cases

#	Case and departure	Teaching point
1	recipe, Train view, 72 h	precursor mass against insulin
2	induction time moved	titer against usable product
3	one wash volume	host-cell protein against loss
4	transfer on the estimate, Refolding view	bolus against fed charge
5	both model rules, temperature moved	faster lots, less folded
6	acid load ×3, base pump in Manual at 0	pH, folding and aggregation
7	endpoint rule moved	a peak behind a delayed assay
8	harvest at 22 h	a harvest the train cannot take
9	22 h harvest of a faster feed, $t_{hold,H} = 10$ h, 350 L lots	infeasible, logged sweep
10	challenge: 1.5 wash volumes, bolus to 2.5 g/L, acid ×4 at 30 h, chiller 60 % at 46 h	the whole campaign

60 s steps for up to 30 h to forecast the time and height of the insulin peak. The estimator keeps the declared constants when a case perturbs the plant's. Chapter 8 corrects a culture's glucose the same way, from one sample a day, and Volume 5, Chapter 4 steers a cook whose grade arrives hours after the move that set it.

9.5.3 Transfer and Endpoint Rules

The transfer rule empties the refolding vessel after a fixed hold $t_{fold} = 16$ h from the start of the charge, when two successive RP-HPLC folded shares differ by less than 0.01 after 2 h, or when the estimate says the charge can make less than $\varepsilon = 3\,\%$ of $\hat{N}$ more ($\hat{N}_{plat} \le \varepsilon\hat{N}$, after one report and 1 h). The endpoint rule stops the conversion after a fixed $t_{end} = 16$ h, after two successive reports more than 2 % below the lot's best insulin report, or when the forecast puts the peak within 0.5 h. The pilot's methods describe refolding by dilution held for 16 h and an enzyme reaction at 15 °C for 16 h [7], which the bolus and the two time rules follow. The application keeps three models apart: the plant of Section 9.3, the identified behavior a learner reads from the trends, and the controller's model, which is the platform's declared constants inside Equations (9.11) and (9.12) and knows nothing of a perturbed plant.

> **TRY IT IN THE APP** **7 · The Conversion Endpoint Behind a Delayed Assay**
>
> Run this case once with each of the three endpoint rules. The fixed time stops at 16 h with 57.3 % of the molecules converted, the assay rule at 9.01 h with 71.7 and the forecast at 7.51 h with 72.5, against a true peak of 72.5 % at 7.64 h.

9.6 Ten Cases, Measured

The cases are the application's scenarios in the sense of Section 1.6.4, and they are listed in Table 9.4. All start from a recipe modeled on the pilot's, a 6 h induction, a 16 h harvest, 250 L lots washed three times, a bolus to 2.0 g/L at 15 °C and pH 10.6, and fixed 16 h holds, and differ from it only where their rows say so. Every number in this section comes from the implemented model run without a display, advanced by the same step sequence the browser uses, with seed 20261020 and every setting typed in before the clock starts. Loading order does not matter, since each case produced identical logs when opened first, when selected after another case and when rerun from Reset.

9.6.1 The Competition in One Vessel

The opening experiment charges the first lot too quickly, as shown in Figure 9.4. At 21.5 h, 2.5 h after a bolus put 815 g into 408 L, the vessel holds 416 g unfolded, 232 g folded and 167 g aggregated, and aggregation runs at 35 g/h against folding at 69 g/h. The lot ends at 67.7 % folded in case 4, within 2.4 points of the closed form for its $\alpha = 0.96$ (70.1 %). More buffer is the other lever. The same lot charged as a bolus into 815 L for 1.0 g/L folds 78.3 %. The scheduler at 0.35 g/L gains 17.0 points

Table 9.5 Measured Outcomes of Cases 1 to 9

#	Measured
1	3013 g made, 217 g accepted in two lots, 3.02 g/h. Of the made: precipitation 24.3 %, fragments 18.5, wash 14.8, still in the train 14.6, cells and holds 10.4, purification 10.1.
2	precursor at harvest, induction at 3, 6, 7, 10 h: 2166, 2905, 2933, 2399 g. Accepted at 4, 6, 10 h: 188, 217, 179 g.
3	lot 1 at 0, 1, 2, 3, 5 wash volumes: 131.6, 124.3, 117.1, 110.3, 97.8 g at 8349, 2789, 892, 285, 29 ppm. Washed out at 1, 3, 5: 158, 445, 696 g.
4	lot 1, bolus to 2.0 g/L: 67.7 % folded, 30.0 aggregated, 18.8 h. Scheduler at 0.35, 0.6, 0.9 g/L: 84.7, 79.0, 72.1 % in 39.0, 27.8, 18.8 h. At 0.35, 665 g expire and nothing is accepted (bolus 112.7 g).
5	lot 1 at 15, 20, 25, 30 °C: 67.7, 62.1, 55.1, 48.7 % in 18.8, 14.8, 9.5, 6.8 h. Campaign 3.91, 5.04, 4.53, 4.04 g/h.
6	pH falls 0.13 h^{-1} to 8.44, lot folds 46.2 % with 47.9 aggregated, against 65.4 and 30.8 in Auto.
7	lot 1, fixed, assay, forecast: 57.3, 71.7, 72.5 % of molecules at 16.00, 9.01, 7.51 h, 110.3, 137.9, 139.4 g. True peak 72.5 % at 7.64 h. Campaign 3.02, 3.69, 3.82 g/h.
8	harvest at 16, 19, 22 h: 2905, 3131, 3238 g precursor, 217.3, 226.0, 114.6 g accepted, 0, 59, 136 g expired.
9	5961 g harvested, 1618 g expired, 308.1 g in one lot (4.28 g/h). Best of 48 runs 542.5 g with 539 g lost, none clears the tank.

Table 9.6 The Challenge With the Plant Perturbed and the Controller's Constants Kept (Accepted Insulin per Campaign Hour, g/h)

Plant	Scheduler 0.9	Bolus	Verdict
declared	5.56	5.47	both pass
$k_{a,15}$ −20 / +20 %	5.89 / 5.34	5.79 / 5.19	both pass
$k_{f,15}$ −20 %	3.68	5.01	scheduler fails
$k_{f,15}$ +20 %	5.94	5.74	both pass
θ_a −20 %	5.77	5.66	both pass
$k_{cat}c_{E,0}$ −20 %	5.55	5.45	both pass
ρ +50 %	4.98	4.91	both pass
c_{rel} 0.050 / 0.070	5.56 / 0.00	5.47 / 0.00	fail at 0.070
RP-HPLC delay 1.5 h	5.59	5.47	both pass
conversion assay delay 1 h	5.61	5.36	both pass

over the bolus, and its cost appears in the campaign rather than in the vessel. In case 5 temperature trades the same way. The first lot keeps less at every step warmer, but at 20 °C a third lot fits and the campaign makes 5.04 g/h against 3.91 at 15 °C, while at 30 °C the lots are too poor and the rate falls back to 4.04 g/h. With the pH loop in Manual in case 6, the ratio k_aU/k_f at 1 g/L rises from 0.57 at pH 10.47 to 5.6 at pH 9.52 and about 670 at 8.44, and the lot loses 19.2 points of folded protein.

🕮 **Pause.** Try Exercise 9.2 before reading on.

9.6.2 The Wash and the Endpoint

Host-cell protein decides the wash. One wash volume leaves 2789 ppm and rejects a lot of 124.3 g, two pass at 892 ppm, and each further volume costs 6 to 7 g of insulin per lot. The endpoint is the peak of a reaction that consumes its product. The fixed 16 h reaction lands 15.2 points below the peak and leaves the lot at 95.76 % purity, 0.76 points above the limit. The assay rule stops 1.4 h after the true peak, and at half the dose the assay's own scatter stops the reaction 5.8 points short of the forecast's 72.5 %.

> **+ DID YOU KNOW? Twice the Reaction Time, a Fifth Less Insulin**
>
> From the simulator: in case 7 the conversion stopped by the forecast rule ends after 7.51 h with 139.4 g of insulin, while the fixed 16 h reaction ends with 110.3 g (Table 9.5). More than twice the reaction time leaves a fifth less product, because the enzymes go on to cut what they have made.

9.6.3 Titer, Recipe and Concentration-Aware Scheduling

The plan's comparison is run inside the challenge, with its acid upset at 30 h and chiller loss at 46 h, and the stage ledgers are listed in Table 9.7. The fixed recipe, with only the wash corrected to three

Table 9.7 Campaign Ledgers by Policy in the Challenge (Acid Upset at 30 h, Chiller Loss at 46 h, 72 h Horizon, Default Seed). Losses as Percent of the Precursor Made

Policy	Made (g)	Wash (%)	Precip. (%)	Fragm. (%)	Purif. (%)	In train (%)	Expired (g)	Lot 1 folded (%, h)	Buffer (L)	Lots	Accepted (g)	Rate, verdict (g/h)
shipped (1.5 wash volumes, bolus to 2.5 g/L)	3013	7.7	29.9	19.2	10.6	15.1	0	62.3, 16.0	744	0 of 2	0	0.00, fails
fixed recipe (shipped, 3 wash volumes)	3013	14.8	27.2	17.5	9.6	13.7	0	62.3, 16.0	677	2	208	2.88, fails
titer: recipe, harvest at 22 h	3459	13.5	17.4	8.0	4.4	35.3	136	62.2, 16.0	708	1	109	1.51, fails
titer: suggested, harvest at 22 h	3459	14.1	20.0	15.6	9.5	17.7	0	68.5, 12.8	1400	2	296	4.11, fails
titer: suggested, induction at 7 h	3027	14.7	22.0	24.4	14.9	0.0	0	70.0, 14.8	1280	3	405	5.62, passes
suggested settings, bolus	3013	14.8	23.5	23.9	14.7	0.0	0	67.5, 14.8	1283	3	394	5.47, passes
suggested settings, scheduler at 0.9 g/L	3013	14.6	21.8	24.3	14.8	0.0	0	70.0, 14.8	1267	3	401	5.56, passes
suggested settings, scheduler at 0.35 g/L	3013	9.9	9.8	18.8	11.2	0.0	665	79.6, 20.4	857	2	312	4.34, fails

volumes, folds each lot to about 62 % in a 16 h hold, converts 58 % of the molecules, and finishes two lots for 2.88 g/h. Maximizing titer by a later harvest is the wrong lever. Harvesting at 22 h raises the precursor made from 3013 to 3459 g, but the recipe then finishes one lot (1.51 g/h) with 1222 g still in the train, and even the suggested downstream settings finish only two (4.11 g/h). Maximizing titer by the induction hour is harmless here. Induction at 7 h instead of 6 adds 28 g at the harvest and, with the suggested settings, raises the rate from 5.56 to 5.62 g/h, because the 16 h harvest leaves the train time for three lots. The difference between the two is time. Extra precursor helps only when it arrives without delaying the first lot.

🕮 **Pause.** Try Exercise 9.3 before reading on.

Concentration-aware scheduling wins narrowly. The suggested settings (three wash volumes, 1.5 g/L charge, 20 °C for both vessels, both model rules and the scheduler at 0.9 g/L) isolate 401 g in three lots, 5.56 g/h, against 5.47 g/h for a bolus with the same other settings. The scheduler's first lot folds 70.0 % against 67.5 in the same 14.8 h, so at a loose ceiling the better fold costs no vessel time. Over six paired seeds the scheduler makes 5.55–5.61 g/h and the bolus 5.43–5.47 g/h, a margin of 0.10–0.16 g/h. The ceiling decides which side of a cliff the campaign is on, as shown in Figure 9.2b and c. Every ceiling from 0.75 to 1.5 g/L fits three lots and passes, and the lowest of them is best, 5.68 g/h with a first lot folded 71.6 %, and 5.64–5.68 g/h over the six seeds. At 0.6 g/L the first lot folds 74.0 % but only two lots fit and the rate drops to 4.12 g/h, and at 0.35 g/L the campaign also loses 665 g to the harvest clock. The suggested set is a passing configuration, not the optimum. The eight policies are drawn side by side in Figure 9.3.

9.6.4 Robustness of the Comparison

The scheduler's advantage rests on the controller's folding constants, as shown in Table 9.6. The plant was perturbed while the estimator kept the declared constants. Slower aggregation, faster folding, a weaker temperature dependence of aggregation and slower assays all leave the ranking intact. A plant that folds 20 % slower than the controller assumes reverses it. The scheduler's model rule then waits longer, its third lot is ready too late, and it fails at 3.68 g/h, while the bolus passes at 5.01 g/h. The 0.75 g/L ceiling fails the same way (3.77 g/h). The purity gate is the most sensitive declared number. Recipe lots sit at 95.76 % and flip to rejected at $c_{rel} = 0.048$, 19 % above the declared 0.040, so cases 1, 3 and 8 depend on it, while the suggested settings pass from 0.030 to 0.060 and fail at 0.070, as does the bolus variant at 0.070.

9.6.5 An Infeasible Harvest and the Challenge

Case 9 delivers 5961 g of precursor into a tank with a 10 h hold limit. As shipped, two 350 L lots leave the tank at 22.0 and 25.5 h, the clock runs out at 32 h and 1618 g (27 %) is discarded. A logged sweep of 48 runs over lot size, charge strategy, refolding temperature and conversion temperature finds no setting that clears the tank. The least loss is 539 g, with three lots and 542.5 g isolated at 7.54 g/h, and the run that isolates the most, 601.1 g, still discards 1618 g. The case is scored for recognizing the limit and getting more through it, not for an impossible target. The challenge's target is 4.6 g/h

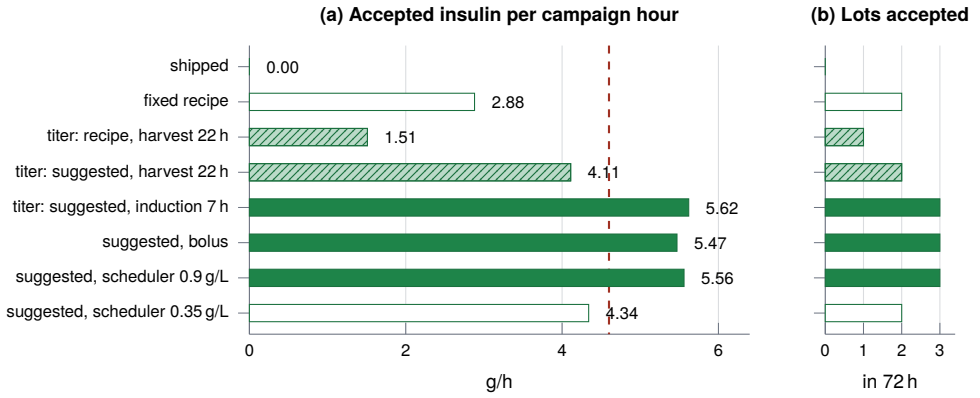


Figure 9.3 The eight policies of Table 9.7 drawn as bars: accepted insulin per campaign hour against the challenge target of 4.6 g/h (dashed line), and the lots the 72 h horizon accepted. Solid bars pass, hatched bars chase titer and white bars fail. Notice that the three passing policies are the three that finish three lots, that the richest harvest of all (two hatched bars) buys at most two, and that the lowest ceiling tried, 0.35 g/L, folds its lots best and still ends below the target.

with every lot accepted, nothing expired, and the vessel within 1.5 °C and 0.3 of its setpoints. As shipped both lots are rejected on host-cell protein at 1654 and 1685 ppm and the chiller loss leaves a 15 °C vessel 1.6 °C high. The suggested settings pass at all six seeds and, at the default seed, hold the vessel within 0.04 °C and 0.12 of pH.

The model runs 850 000 to 875 000 steps per second headless on the build host, so a 72 h campaign of 51 900 steps takes 0.06 s, and in a browser it plays in 36 s at the top speed of 7200 times real time. Halving the step to 2.5 s leaves the challenge's rates unchanged to 0.001 g/h and the first lot's folded share to 0.01 point.

9.7 Technology in Practice

The simulated train carries four loops, two sampled analyzers and two rule sets. Figure 9.5 draws them with the devices a plant would install.

What Measures It

Temperature, pH and vessel weight are read continuously, and the composition is not. A glass electrode gives millivolts per pH unit and a resistance thermometer or a thermocouple in a thermowell gives the vessel and jacket temperatures, which is what the 60 s, 30 s and 120 s lags of Section 9.3 stand for. The share of the protein that has folded has no probe at all. It is drawn as a sample and separated, and the chapter's RP-HPLC analyzer is that instrument, reporting one result per cycle and holding it until the next.

> **TECHNOLOGY Sensor: At-Line Liquid Chromatography**
>
> The folded, aggregated and unfolded shares of a refolding vessel are read by reversed-phase liquid chromatography on a drawn sample. The separation is denaturing, so it tolerates the chaotropic salts of this stream, which interfere with the measuring principle of near-infrared and Raman probes, and at-line means a measurement made close to the process stream. A method developed for inclusion-body work runs in 8.1 min at 78 °C and 1.2 mL/min, short enough to steer a step while it proceeds, where the gel assay it replaces is too slow for in-process control [19]. The chapter's analyzer stands for this device: one sample an hour, the report 0.75 h later, and nothing in between.

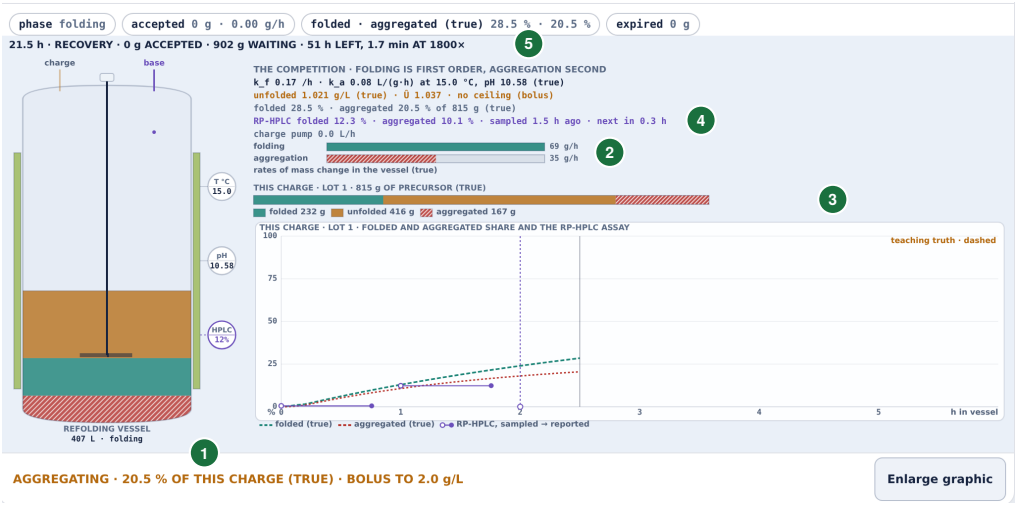


Figure 9.4 A bolus charge aggregating, case 4 as shipped at $t = 21.5\,\mathrm{h}$ on the Refolding view, cropped to the plant view. The first lot of 815 g went in as a bolus at 19.0 h to 2.0 g/L in 408 L, and 2.5 h later it is 28.5 % folded, 20.5 % aggregated and 51.0 % unfolded (232, 167 and 416 g, teaching truth). Callout 1 is the vessel with its temperature, pH and assay badges. Callout 2 is the competition panel, which gives $k_f = 0.17\,\mathrm{h}^{-1}$ and $k_a = 0.08\,\mathrm{L/(g\,h)}$ at 15.0 °C and pH 10.58, the true unfolded 1.021 g/L against the estimate 1.037 g/L, and folding at 69 g/h against aggregation at 35 g/h. Callout 3 is the charge drawn as folded, unfolded and aggregated mass, and callout 4 the latest RP-HPLC report, from the sample drawn at 20.0 h, which reads 12.3 % folded and 10.1 % aggregated with the next report due in 0.3 h. Callout 5 is the status line: 902 g of harvest waits behind the lot.

What Moves It

Four final elements act. The charge pump meters the solubilized lot, a base pump doses the titrant for the pH loop, a utility valve sets the jacket supply through the cascade, and a transfer valve empties the vessel into the conversion stage once a rule fires. Only the first two run continuously, and the chapter treats all four as ideal: a commanded flow becomes that flow. A plant pays for the difference in calibration work.

> **TECHNOLOGY** **Final Element: The Charge Pump**
>
> The solubilized lot is metered into the refolding vessel by a positive-displacement pump. A peristaltic head is common in sterile service, because the liquid touches only a length of replaceable tubing and the delivered volume follows the rollers, so the rate is set by speed alone. The dose is referred to weight. In a study of 149 847 pharmaceutical dosing measurements the pump was calibrated against a balance before every run, fitting a fresh length of tubing shifted that calibration, and one fifth of the recorded sequences drifted within a single calibration period [20]. A worn tube shows on a trend as a vessel weight that climbs more slowly than the commanded charge, which is the one failure the chapter's ideal F cannot produce.

Where the Controller Runs

The loops of a recovery train are hosted on a distributed control system, an architecture that puts a supervisory level over local controllers and is used for pharmaceutical processing among other industries [21]. Supervisory and lower loops run continuously with cycle times from milliseconds to minutes, and loops are nested so that the setpoint of one is the process variable of another [21], which is the cascade TIC-301 to TIC-302 of Section 9.5. The chapter's controller scan is 30 s, slow for a jacket of 300 s turnover and fast for a reaction of hours. The scheduler, the estimator and the transfer

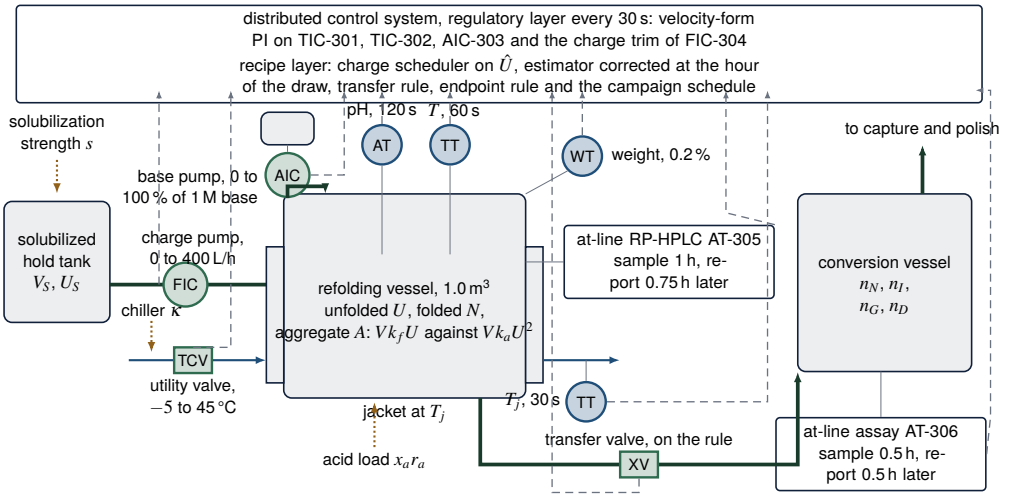


Figure 9.5 The chapter's loops with the devices a plant would install. Solid green lines carry protein, thin blue lines the jacket utility, dashed gray lines the signals and dotted amber arrows the disturbances. Notice that the two composition measurements are the only instruments that are not continuous: each at-line chromatograph reports one result per cycle, after the delay printed beside it.

and endpoint rules belong one layer above the loops, in the recipe or phase logic that sequences a batch, where the delayed assay arrives and the decision to charge, transfer or stop is taken.

> **IN PRACTICE** **Dilute Early, and Quench at the Peak**
>
> Two rules of thumb govern the two reactions in service. Proinsulin is diluted to under 1 mg/mL in the refolding buffer to prevent cross-folding between molecules, and the enzyme reaction that follows is quenched at the maximal insulin point, because the des-threonine species grows with reaction time [22]. Both calls are made on chromatograms of drawn aliquots, which is why each step is run to a measured end and not to a clock. United States drug manufacturing rules require written in-process controls and the testing of in-process material at the completion of significant phases or after long storage [23].

What Is Modernizing It

Three lines of work would change the measurement boundary this chapter draws. The first is in-process chromatography: an at-line method short enough to follow solubilization as it runs turns an end-of-step check into a control signal, and the same monitoring showed that solubilization time alone moves the refolding yield that follows [19]. The second is in-situ optical monitoring. Intrinsic tryptophan and tyrosine fluorescence tracks refolding in real time down to 0.01 g/L, with the shift of the emission peak following the folding intermediates and the intensity following the extent of aggregation, and both correlations have been written into a mechanistic model as an observation function [24]. The third is the control layer itself: state estimation that fills in what no probe reads, and model-based control, including receding-horizon control, of the feed rate of a fed charge [25]. Each one shortens the 0.75 h the simulated estimator must bridge.

9.8 Limitations

The upstream fermenter is a teaching model whose kinetics and induction burden have no source, and only the load it delivers is anchored. The folding law agrees with its source's concentration series only at 1.5–2.0 g/L, the over-processing ratio is a teaching value, purification is a declared map whose carried-over fraction decides every recipe lot within 19 %, and the well-mixed vessel

cannot represent the charge point, where aggregation is worst. The controllers are four PI loops, a feedforward scheduler with a trim, and rule-based decisions. Nothing optimizes the schedule, the ceiling or the harvest, the best ceiling of 0.75 g/L is a sweep result, and the scheduler's margin of 0.10–0.16 g/h disappears if the plant folds 20 % slower than its constants. Quality is scored on simulated attributes, and every limit, hold time and horizon is a teaching value.

OPEN CHALLENGE **Refolding Still Has No Platform Recipe**

Refolding remains the bottleneck of inclusion-body processing. A review of monitoring and control lists the open drawbacks as the large buffer volumes and tank capacities, uneven mixing in those volumes, analytics that struggle at the low protein concentrations dilution forces, kinetics that run into unusable aggregate and generally low yields, with no generic platform procedure and a lack of robust control strategies [25]. Progress would look like the route that review proposes: measure the folding kinetics in real time, estimate from them the states that no probe reads, and close the loop on the feed rate of a fed charge.

9.9 Conclusion

This chapter has worked an insulin recovery train in which the folding competition, the vessel schedule and the hold clocks share one 72-hour campaign, over a protein ledger that closes to 1.8×10^{-13}. A fed charge folds 17 points more of a lot than a bolus and can still leave the campaign with nothing accepted, because the vessel it occupies is the bottleneck. A harvest 11 % richer in precursor delivers half the insulin when it delays the train, while a richer harvest at the same hour helps slightly. In the challenge, concentration-aware charging beats a bolus by 0.10–0.16 g/h over six seeds and the best ceiling tried reaches 5.68 g/h, but both scheduler settings fail when the plant folds 20 % slower than the controller assumes. The refolding constants reproduce the pilot's temperature pair by construction and its concentration series only at 1.5–2.0 g/L, and the upstream kinetics are teaching values. A fuller treatment would add an optimized schedule, a two-enzyme conversion and refolding data at fixed reductant. What the cases show is that the yield of one vessel and the productivity of a campaign are different objectives, and that the second one is what a plant is paid for.

KEY POINTS

- A vessel that keeps more of its protein can still lose the campaign, because everything behind it waits.
- Unfolded precursor folds at a rate first order in its concentration and aggregates at a rate second order in it, so a lot charged slowly keeps more protein.
- The refolding vessel is the bottleneck, and the campaign is scored by accepted insulin per hour, not by one lot's yield.
- A harvest at 22 h carries 11 % more precursor than one at 16 h and the campaign isolates about half as much insulin.
- The scheduler's margin over a bolus disappears when the plant folds 20 % slower than the controller assumes.

Where else this applies. Chapter 8 corrects a culture from a sample reported hours late, Volume 5, Chapter 4 steers a cook whose grade arrives after the move that set it, and Volume 8, Chapter 2 meters a feed to keep an unreacted store small.

References

[1] T. Kiefhaber, R. Rudolph, H.-H. Kohler, and J. Buchner, "Protein aggregation in vitro and in vivo: A quantitative model of the kinetic competition between folding and aggregation," *Nature Biotechnology*, vol. 9, no. 9, pp. 825–829, 1991.

[2] A. P. J. Middelberg, "Preparative protein refolding," *Trends in Biotechnology*, vol. 20, no. 10, pp. 437–443, 2002.

[3] S. M. Singh and A. K. Panda, "Solubilization and refolding of bacterial inclusion body proteins," *Journal of Bioscience and Bioengineering*, vol. 99, no. 4, pp. 303–310, 2005.

[4] C. J. Roberts, "Non-native protein aggregation kinetics," *Biotechnology and Bioengineering*, vol. 98, no. 5, pp. 927–938, 2007.

[5] I. Harjunkoski, C. T. Maravelias, P. Bongers, P. M. Castro, S. Engell, I. E. Grossmann, J. Hooker, C. Méndez, G. Sand, and J. Wassick, "Scope for industrial applications of production scheduling models and solution methods," *Computers & Chemical Engineering*, vol. 62, pp. 161–193, 2014.

[6] K. Lakhdar, S. S. Farid, N. J. Titchener-Hooker, and L. G. Papageorgiou, "Medium term planning of biopharmaceutical manufacture with uncertain fermentation titers," *Biotechnology Progress*, vol. 22, no. 6, pp. 1630–1636, 2006.

[7] C.-K. Kim, S.-B. Lee, and Y.-J. Son, "Large-scale refolding and enzyme reaction of human preproinsulin for production of human insulin," *Journal of Microbiology and Biotechnology*, vol. 25, no. 10, pp. 1742–1750, 2015.

[8] I. S. Johnson, "Human insulin from recombinant DNA technology," *Science*, vol. 219, no. 4585, pp. 632–637, 1983.

[9] L. Rosenfeld, "Insulin: Discovery and controversy," *Clinical Chemistry*, vol. 48, no. 12, pp. 2270–2288, 2002.

[10] D. F. Steiner, D. Cunningham, L. Spigelman, and B. Aten, "Insulin biosynthesis: Evidence for a precursor," *Science*, vol. 157, no. 3789, pp. 697–700, 1967.

[11] W. Kemmler, J. D. Peterson, and D. F. Steiner, "Studies on the conversion of proinsulin to insulin. I. conversion in vitro with trypsin and carboxypeptidase B," *Journal of Biological Chemistry*, vol. 246, no. 22, pp. 6786–6791, 1971.

[12] D. V. Goeddel, D. G. Kleid, F. Bolivar, H. L. Heyneker, D. G. Yansura, R. Crea, T. Hirose, A. Kraszewski, K. Itakura, and A. D. Riggs, "Expression in Escherichia coli of chemically synthesized genes for human insulin," *Proceedings of the National Academy of Sciences*, vol. 76, no. 1, pp. 106–110, 1979.

[13] R. E. Chance and B. H. Frank, "Research, development, production, and safety of biosynthetic human insulin," *Diabetes Care*, vol. 16, no. Supplement 3, pp. 133–142, 1993.

[14] G. Walsh, "Therapeutic insulins and their large-scale manufacture," *Applied Microbiology and Biotechnology*, vol. 67, no. 2, pp. 151–159, 2005.

[15] N. A. Baeshen, M. N. Baeshen, A. Sheikh, R. S. Bora, M. M. M. Ahmed, H. A. I. Ramadan, K. S. Saini, and E. M. Redwan, "Cell factories for insulin production," *Microbial Cell Factories*, vol. 13, p. 141, 2014.

[16] R. Rudolph and H. Lilie, "In vitro folding of inclusion body proteins," *The FASEB Journal*, vol. 10, no. 1, pp. 49–56, 1996.

[17] T. Linke, M. T. Aspelund, C. Thompson, G. Xi, A. Fulton, M. Wendeler, T. M. Pabst, X. Wang, W. K. Wang, K. Ram, and A. K. Hunter, "Development and scale-up of a commercial fed batch refolding process for an anti-CD22 two chain immunotoxin," *Biotechnology Progress*, vol. 30, no. 6, pp. 1380–1389, 2014.

[18] European Medicines Agency, "ICH Q6B specifications: Test procedures and acceptance criteria for biotechnological/biological products," Scientific guideline CPMP/ICH/365/96, effective 1 September 1999, 1999. [Online]. Available: https://www.ema.europa.eu/en/scientific-guidelines/ich-q6b-specifications-test-procedures-acceptance-criteria-biotechnological-biological-products

[19] J. Ebner, D. Humer, R. Klausser, V. Rubus, R. Pell, O. Spadiut, and J. Kopp, "At-line reversed phase liquid chromatography for in-process monitoring of inclusion body solubilization," *Bioengineering*, vol. 8, no. 6, p. 78, 2021.

[20] D. Privitera, A. Mecocci, and S. Bartolini, "Extensive dataset for peristaltic pump accuracy enhancement in pharmaceutical environments," *Scientific Data*, vol. 12, no. 1, p. 1618, 2025.

[21] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule, and M. Thompson, "Guide to operational technology (OT) security," National Institute of Standards and Technology, Tech. Rep. NIST Special Publication 800-82r3, 2023.

[22] Y. Y. Siew and W. Zhang, "Downstream processing of recombinant human insulin and its analogues production from *e. coli* inclusion bodies," *Bioresources and Bioprocessing*, vol. 8, no. 1, p. 65, 2021.

[23] U.S. Food and Drug Administration, "21 CFR 211.110: Sampling and testing of in-process materials and drug products," Code of Federal Regulations, title 21, revised as of April 1, 2024, 2024. [Online]. Available: https://www.govinfo.gov/content/pkg/CFR-2024-title21-vol4/xml/CFR-2024-title21-vol4-sec211-110.xml

[24] C. L. Igwe, D. F. Müller, F. Gisperg, J. N. Pauk, M. Kierein, M. Elshazly, R. Klausser, J. Kopp, O. Spadiut, and E. Přáda Brichtová, "Online monitoring of protein refolding in inclusion body processing using intrinsic fluorescence," *Analytical and Bioanalytical Chemistry*, vol. 416, no. 12, pp. 3019–3032, 2024.

[25] J. N. Pauk, J. Raju Palanisamy, J. Kager, K. Koczka, G. Berghammer, C. Herwig, and L. Veiter, "Advances in monitoring and control of refolding kinetics combining PAT and modeling," *Applied Microbiology and Biotechnology*, vol. 105, no. 6, pp. 2243–2260, 2021.

9.10 Exercises

9.1 Concept check · 2 min

A lot of solubilized precursor is charged into the refolding vessel as a bolus at 15 °C and pH 10.6. The same lot is then charged into twice as much buffer, so that the unfolded concentration once the charge is in is halved. What happens to the share of the lot that ends up correctly folded?

(a) It is unchanged, because folding is first order and the lot's mass is the same

(b) It rises, because $\alpha = k_a U_0 / k_f$ is halved and $\ln(1+\alpha)/\alpha$ falls as α rises

(c) It falls, because the folding rate $k_f U$ is halved while the hold is the same length

(d) It doubles, because aggregation is second order in the unfolded concentration

9.2 Concept check · 2 min

Folding doubles for each 10 K ($Q_{10} = 2.0$) and aggregation nearly quadruples ($Q_{10} = 3.86$). In case 5 the whole campaign delivers the most insulin with the refolding vessel at 20 °C, not at 15 and not at 30 °C. Why is the best temperature an interior one?

(a) Because the folded share of a single lot is highest at 20 °C

(b) Because the purity gate rejects lots refolded above 20 °C

(c) Because the hours a lot occupies the vessel fall faster than its folded share until a third lot fits, after which every further kelvin is yield given away

(d) Because the chiller cannot hold 15 °C once its capacity falls

9.3 Concept check · 1 min

In case 8 the 22 h harvest carries 11 % more precursor than the 16 h harvest and the campaign isolates about half as much insulin. What is the direct cause?

(a) The extra precursor ages in the hold tank, and all of it expires

(b) The larger lots fold worse, because a bigger charge means a higher unfolded concentration

(c) The extra host-cell protein pushes both lots past the 1000 ppm limit

(d) The six extra fermentation hours are six fewer recovery hours, so the train finishes one lot instead of two and the rest is still inside it at the horizon

9.4 Theory and calculation · 6 min

A lot is charged as a bolus to 2.0 g/L at 15 °C and pH 10.6, where $k_{f,15} = 4.61 \times 10^{-5}\,\mathrm{s^{-1}}$ and $k_{a,15} = 2.21 \times 10^{-5}\,\mathrm{m^3/(kg\,s)}$ (Table 9.2).

(a) Compute $\alpha = k_a U_0 / k_f$ for this charge, remembering that 1 g/L is 1 kg/m^3. *Checkpoint:* $\alpha = 0.96$

(b) Compute the folded share the charge would reach if it were left until every unfolded molecule had gone one way or the other. *Checkpoint:* 70.1 %, to within 0.1 point

(c) Case 4 measures 67.7 % folded for exactly this charge. Explain in two sentences why the measured share is below the value of part (b), and say which way the gap moves if the transfer rule waits longer.

9.5 Theory and calculation · 5 min

The temperature dependence of the two constants is Equation (9.5), with $\theta_f = 0.0693\,\mathrm{K^{-1}}$ and $\theta_a = 0.1351\,\mathrm{K^{-1}}$.

(a) Compute k_f and k_a at 25 °C from their values at 15 °C given in the previous exercise. *Checkpoint:* $k_f = 9.22 \times 10^{-5}\,\mathrm{s^{-1}}$, $k_a = 8.54 \times 10^{-5}\,\mathrm{m^3/(kg\,s)}$

(b) Compute α and the folded share of a bolus to 1.5 g/L at that temperature, and compare with the 75.3 % the same charge reaches at 15 °C. *Checkpoint:* 62.7 %, to within 0.2 point

(c) A colleague proposes running every refolding vessel 10 K warmer because the reaction then takes half as long. State in one sentence what that argument leaves out.

9.6 Theory and calculation · 6 min

The scheduler's feedforward is the first line of Equation (9.11), with $C_{sol} = 10\,\mathrm{g/L}$ the solubilized concentration the controller assumes.

(a) Compute $\hat{k}_f$ and $\hat{k}_a$ at 20 °C, where the suggested settings hold the vessel. *Checkpoint:* $0.235\,\mathrm{h^{-1}}$ and $0.157\,\mathrm{L/(g\,h)}$

(b) The vessel holds 500 L part way through a charge and the ceiling is $U_{tgt} = 0.9\,\mathrm{g/L}$. Compute the feedforward charge rate F_{FF} in L/h. *Checkpoint:* 16.9 L/h, to within 0.2 L/h

(c) The PI trim is bounded to $[-F_{FF},\ F_{max} - F_{FF}]$. Name two errors in the feedforward that the trim has to absorb, and say which of them it cannot see.

9.7 App challenge · 8 min

Open scenario **4** · **A bolus charge against a fed one** at the default seed, on the Refolding view, with the speed at 7200× (the whole 72 h campaign takes 36 s).

(a) Run the shipped case, in which *Charge strategy* is Bolus and *Charge concentration* is 2.0 g/L, until the first lot has folded and left the vessel. Read the folded and aggregated shares and the hours in the vessel from the LAST CHARGE line of the refolding vessel panel. *Checkpoint:* 67.7 and 30.0 % in 18.8 h, each to within 0.3

(b) Press Reset, set *Charge strategy* to Scheduler on $\hat{U}$ and the *Unfolded ceiling* $\hat{U}$ to 0.35 g/L, and run the same lot again. *Checkpoint:* 84.7 and 13.1 % in 39.0 h

(c) Let each of the two runs reach the 72 h horizon and read the *Insulin accepted* and the *expired* figures on the scorecard. Say in one sentence which charge the plant should buy. *Checkpoint:* bolus 112.7 g accepted and nothing expired; scheduler nothing accepted and 665 g expired, both to within 1 g

9.8 App challenge · 9 min

Open scenario **7** · **The conversion endpoint behind a delayed assay** at the default seed, with the speed at 7200×. Everything is at its shipped value except the *Stop the conversion* rule on the Supervisor card. Each run needs about half a minute to isolate the first lot.

(a) Run with the shipped rule, After a fixed time (16 h). Read the molar conversion from the scorecard tile *Last lot: folded · converted*, and the insulin and purity of the lot from the LOTS ISOLATED list on the Recovery view. *Checkpoint:* 57.3 % of molecules, 110.3 g, 95.76 %, each to within 0.2

(b) Reset and repeat with When the assay falls, then with At the predicted peak. *Checkpoint:* 71.7 and 72.5 %; 137.9 and 139.4 g

(c) The true insulin peak of this lot is 72.5 % of the molecules at 7.64 h of reaction. The assay rule stops at 9.01 h and the forecast at 7.51 h. Explain in three sentences why the assay rule is late by construction and why being early costs the forecast almost nothing.

9.9 App challenge · 8 min

Open the **Challenge** · **The most insulin a campaign can deliver** at the default seed, with the speed at 7200×, and run each configuration for the full 72 h (36 s of wall clock). Enter every setting before pressing Run.

(a) Run the campaign as shipped and read the *Accepted insulin rate* tile and the *worst lot purity · host-cell protein* tile. Say in one sentence why the rate is what it is. *Checkpoint:* 0.00 g/h; both lots rejected near 1670 ppm

(b) Reset and enter the suggested settings: *Wash intensity* 3, *Charge strategy* Scheduler on $\hat{U}$, *Unfolded ceiling* 0.90 g/L, *Charge concentration* 1.5 g/L, *Refolding temperature* 20 °C, *Conversion temperature* 20 °C, and both supervisory rules on their model-based settings (When the estimate says it is done, At the predicted peak). Read the rate and the accepted mass. *Checkpoint:* 5.56 g/h and 401 g in three lots, to within 0.02 g/h

(c) Change only *Charge strategy* to Bolus, Reset and run again. Read the rate, and state what the difference between (b) and (c) is worth against the 0.3 pH and 1.5 °C tolerances the challenge also imposes. *Checkpoint:* 5.47 g/h

9.10 Application area · 3 min

Section 9.3 assumes the refolding vessel is one well-mixed liquid.

(a) Name the three levers a plant uses to keep the unfolded concentration low where the charge enters, and say which one this model can represent. *Checkpoint:* charge point, agitation, feed rate; only the feed rate

(b) Say whether that assumption flatters the bolus or the fed charge, and why.

9.11 Application area · 4 min

A lot is accepted when its purity is at least 95.0 % and its host-cell protein at most 1000 ppm. The recipe's lots sit at 95.76 %, and they flip to rejected when the carried-over related-substances fraction c_{rel} rises from the declared 0.040 to 0.048.

(a) Say what that margin means for the conclusions drawn from cases 1, 3 and 8, and name the measurement a plant would make before trusting them.

(b) Say where real acceptance criteria come from, and what a simulated pass is not. *Checkpoint:* the product specification, the pharmacopeia and regulatory guidance; not a batch release

9.12 Review the agent · 5 min

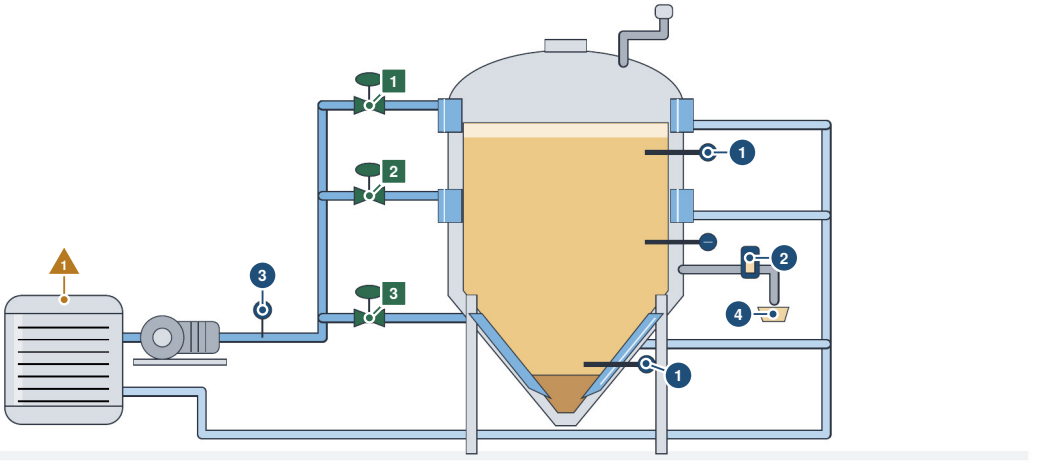
An assistant is asked whether the plant should adopt the scheduler and writes:

> *"In the challenge the scheduler isolates 401 g against the bolus's 394 g, 5.56 g/h against 5.47 g/h, and the advantage survives every perturbation the chapter reports: aggregation 20 % faster or slower, a weaker temperature dependence of aggregation, a 20 % slower enzyme, a heavier over-processing path, an RP-HPLC assay reported 1.5 h late and a conversion assay reported 1 h late. The scheduler is therefore the robust choice, and the plant should run it at the 0.9 g/L ceiling."*

Find the flaw in this recommendation and say how you would test it on the application.

CHAPTER 10

Lager Fermenter Simulator With On/Off Glycol Valves and a Day-Old Diacetyl Result



The controller moves 1 top jacket valve 2 middle jacket valve 3 cone jacket valve

The controller reads 1 zone temperature probes 2 density meter 3 glycol supply temperature 4 daily laboratory sample

What disturbs it 1 other tanks and lost capacity on the shared glycol plant

A tall lager tank is cooled through three jackets by valves that are only open or shut, from a glycol plant it shares with other tanks, and only its temperature is measured in every zone while gravity and diacetyl are sampled at one valve. Illustration, not to scale.

IN PLAIN TERMS Keeping a Lager Tank on Its Recipe

The main idea. Yeast turn sugar into beer and give off heat as they work. A big lager tank is cooled by cold liquid in jackets on its wall, and the valves on the jackets can only be open or shut.

The challenge. The beer is stirred only while the yeast make gas, so one thermometer stops speaking for the whole tank once the gas stops. The test that says the beer is ready comes back a day late, and the cold supply is shared with other tanks.

Why it matters. The batch is finished on time by what starts and ends each step, more than by fine tuning of the loops.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the two history segments (Section 10.2), the two fact segments, both try-it segments (Sections 10.5 and 10.5.2) and the key points
Study	50 min	the whole chapter with its three pause points and the twelve exercises of Section 10.9
Practice	20 min	the first page, the cases and the graded batch (Section 10.5), Technology in Practice (Section 10.6), Limitations (Section 10.7) and the open challenge

The control question A lager tank has only open-or-shut glycol valves, is mixed only while its yeast release gas, ends its diacetyl rest on a result a day old and shares its glycol plant with the cellar. Does the tuning of its temperature loops decide the batch, or do the phase triggers, the endpoint, the probe placement and the shared utility?

Learning objectives

- Explain how a time-proportioned output gives an on/off valve a duty, and why the jacket's heat is a concave function of that duty.
- Compare the limit cycle of hysteresis valves with a time-proportioned PI loop, and relate the swing to the deadband and the probe lag.
- Predict which zone of a quiet tank one probe fails to control during a crash, and in which direction.
- Calculate the correction that a day-old laboratory result gives a model-based estimate, and compare three ways to end a diacetyl rest.
- Choose the valves, probes, triggers and glycol coordination that finish a twelve-day batch on time and in specification.

Prerequisites Sections 1.3 (loop action), 1.4.3 (growth and heat), 1.4.5 (late samples and the bias update) and 1.4.6 (one-way actuators), with Sections 1.6.2 and 1.6.4 on the PID and the scenarios.

The app https://apmonitor.com/apps/brewery-fermentation/

See also Volume 8, Chapter 2 for an exothermic batch under jacket cooling, Volume 8, Chapter 12 for thermostats on a shared refrigeration plant and Volume 8, Chapter 5 for a batch ended on an endpoint. Volume 5, Chapter 10 for a soft sensor corrected by a late laboratory and Volume 5, Chapter 4 for a process steered ahead of its analyzer.

10.1 Introduction

A lager cellar is a set of batch reactors whose heat of reaction arrives on the yeast's schedule rather than the operator's. The tank of this chapter holds 3000 hL of a 12 °P wort that is held at 11 °C through the fermentation, allowed to warm itself to 14 °C for a diacetyl rest, crashed to −1 °C and matured, all inside a twelve-day plan. The temperature program of such a batch has been posed as an optimal control problem with the cooling duty as the single available control, and its solution is bang-bang with a singular arc at the maximum permitted temperature [1]. The rest is often the step the schedule waits for, since lowering the diacetyl of green beer is an expensive and time-consuming part of lager brewing whose rate is usually limited by the spontaneous decarboxylation of a precursor [2].

Four features make the plant hard to control, and none of them appears in a transfer function from valve to thermometer. The heat grows with the yeast and with the temperature, so a tank whose valves stay shut warms only 0.56 °C in its first day and then runs away, to 13.61 °C at 48 h and 21.33 °C on day 5.3. The only actuators are solenoid valves, and a controller that wants a fraction of the cooling has to obtain it in time rather than in position. The tank holds 300 m^3 in three zones, and it is mixed only while the yeast release carbon dioxide (CO_2), with about 330 m^3/h of exchange between zones at the day-two peak against 3 m^3/h of free convection once the gas stops. The measurements that decide each phase are late or partial, an hourly density reading from one sample valve and a diacetyl result a day old, and the glycol comes from a plant that the tank shares with the rest of the cellar.

The misconception this simulator targets is that a batch is kept on its recipe by the tuning of its temperature loops. On the same valves a PI controller with a time-proportioned output holds the primary fermentation between 10.97 and 11.04 °C while a thermostat swings from 10.66 to 11.27 °C, yet the graded batch is decided elsewhere. Shipped with hysteresis valves, one mid-height probe, a free rise timed for day 6, a rest of fixed length and no glycol coordinator, it is crashed on day 12.57 with its unprobed cone 2.73 °C off the recipe. With time-proportioned valves, a probe per zone, a free rise triggered by the density meter, a rest ended on a corrected estimate and the coordinator, it is crashed on day 10.47 with every zone within 0.52 °C. The loop gains are the same in both runs. Hysteresis valves with a probe in every zone hold the zones as closely, but their full openings during

the rest take the shared glycol plant 39.4 kW over its capacity when a neighboring tank starts its crash.

The class of problem is general. It is a batch process with a slow exotherm and a discrete actuator, whose phase changes are decided on measurements that arrive late or see part of the vessel. Jacketed batch reactors with dead time, jacket dynamics and measurable load disturbances share its regulatory part [3], and the exothermic semi-batch reactor of Volume 8, Chapter 2 is one whose heat must likewise be removed when the reaction releases it. A relay in a feedback loop drives a limit cycle whose period and amplitude are set by the process [4], and a phase ended on a result a day old is the problem that soft sensors and inferential control address across the process industries [5], with state estimation already demonstrated on batch beverage fermentation [6]. Wine fermentation shares the failure of a batch that stops before its sugar is gone, described as one of the most challenging problems of wine production and troublesome to diagnose [7], and a refrigeration plant that serves many loads in a brewery is a design and load-management problem of its own [8].

However, process control texts treat on/off control, feedback tuning and inferential measurement as separate topics [9], and the brewing literature treats the temperature profile as an optimization [1] and the kinetics as a model [10]. In neither literature can a thermometer on target, a laboratory a day late and a supply shared with other tanks be set against what the beer actually does. This chapter measures what decides a lager batch, on a browser fermenter that logs every zone's true temperature, gravity, diacetyl and yeast apart from what the probes, the density meter, the laboratory and the estimator report. The kinetics, the time-proportioned output, the bias update and the supervisory rules are established methods, and the chapter works with them through measured comparisons on one plant, among them the concave heat characteristic of the modeled time-proportioned jacket. The eight cases and the graded batch are arranged to separate an actuator command from a heat flow, a probe from a tank, a laboratory from an estimator, and a tank from the plant it shares.

Section 10.2 follows the domain from mechanical refrigeration to kinetic models. Section 10.3 sets out the tank as a 38-state model whose extract, ethanol, CO_2, yeast and energy ledgers close to within 3×10^{-11}, in the detail needed to rebuild it. Section 10.4 describes the zone loops, a time-proportioned output that carries its rounding and minimum-time errors so that the mean opening equals the duty, the glycol coordinator, and a supervisor that declares a stuck fermentation or a late crash from its own measurements. Section 10.5 reports the eight cases and the graded batch, Section 10.6 names the devices and the controller platform of a real cellar, Section 10.7 states the limits of the model, Section 10.8 draws the lessons together, and the exercises of Section 10.9 work through the cases.

10.2 Refrigeration, Fermenting Vessels and Diacetyl

A lager brewery is a refrigeration plant as much as a fermentation plant, and its refrigeration has a record of its own.

> ⧖ HISTORY **Refrigeration Comes to the Brewery (1912 to 1989)**
>
> Mechanical refrigeration as applied to the brewing industry was already the subject of a 39-page paper in the Journal of the Institute of Brewing in 1912 [11]. The design and operation of a brewery's refrigeration plant were later studied by simulating its components against real load variations, and design optimization, a better operating strategy and load management together cut primary fuel use by an ideal 15 % and a practical 10 % [8]. In the tank simulated here the glycol plant carries the whole heat of the batch, and the glycol coordinator is a small instance of that load management.

Practice in the vessel changed as well. A system for fermenting and then maturing lager, the Nathan system, was described to the Institute of Brewing in 1930, when an English brewery was already working on it [12]. The tank modeled here is a closed vessel of the tall cylindroconical kind, with a cone at the bottom where the yeast settles and a jacket zone on each part of its wall. Its interior is not at one temperature. The temperature and velocity fields inside an industrial cylindroconical

fermenting tank of 30 m^3 have been measured with a grid of 43 resistance thermometers and an ultrasound Doppler technique in order to determine the convective flow that carries its heat [13], and a thermowell reports only the liquid around it.

Diacetyl, which the warm rest of the recipe removes, was studied in brewing long before it was modeled.

> ⧖ HISTORY **Diacetyl Becomes a Brewing Problem (1957 to 2013)**
>
> The formation and prevention of diacetyl in beer were under study in 1957 [14]. A review in 1973 found that an excess of diacetyl is a flavor defect in every type of beer, that diacetyl and 2,3-pentanedione form in every brewery fermentation, that much of the trouble comes from the precursor acetolactic acid, which gives rise to diacetyl in the finished beer, and that the yeast strain, the condition of the pitching yeast, the wort and the management of the fermentation all affect what remains [15]. The current account adds that the removal is typically rate-limited by the spontaneous decarboxylation of α-acetolactate to diacetyl, which the yeast then take up with their reductases [2]. A warm rest therefore precedes the crash in the recipe simulated here, and its end is decided on a measurement.

Models followed the practice. Balling's relations among original extract, real and apparent extract and alcohol, on which every gravity reading of this simulator depends, were re-examined against 532 commercial and pilot brews in 2009 [16]. A kinetic model extended with temperature effects turned the temperature program into an optimal control problem in 1988 [1], optimal regulatory control of jacket-cooled fermentation reactors followed in 1991 [3], and in 1996 an extended Kalman filter with recursive parameter identification tracked batch beverage fermentations, simulated and real, well enough despite an imperfect model [6]. The kinetic structure of this simulator follows a model for beer production under industrial operational conditions [10], with constants chosen for teaching. The failure of case 8 has a literature of its own in wine, where the causes of stuck and sluggish fermentations are numerous and difficult to rectify [7]. This simulator gathers the pieces into one tank, with a record that no cellar keeps, the true state of every zone beside what each instrument reported.

10.3 The Tank as Modeled

The simulated tank holds 300 m^3 of wort in a 5.0 m cylinder on a cone, divided into a top and a middle zone of 130 m^3 each and a cone zone of 40 m^3. Each zone has its own jacket, glycol valve and thermowell, the middle zone carries the sample valve of the density meter and the laboratory, and the glycol comes from a plant shared with other tanks, as shown in Figure 10.1.

Table 10.1 classifies the variables by role, as controlled (CV), manipulated (MV) and disturbance (DV) variables, parameters and states. Each zone carries seven states and each jacket one, and with the plant's averaged load, the settled yeast and twelve cumulative ledgers of extract fermented, yeast grown, CO_2, heat released, the heat each jacket removed, the ambient gain, the heat returned to the glycol and each valve's open time, the state vector has 38 entries. Because the ledgers advance with the same integrator as the plant, each balance is checked from the state vector alone. Time is in hours, temperatures in °C, species in kg/m^3, yeast in Mcells/mL (million cells per milliliter), vicinal diketones (VDK) in mg/L, heat in kW and flows in m^3/h and kg/s.

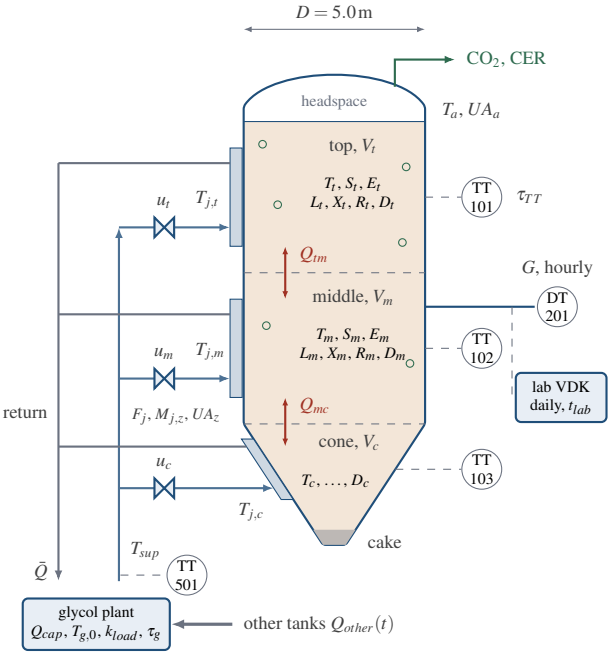


Figure 10.1 Lager fermenter as modeled. The top, middle and cone zones, of volumes V_t, V_m and V_c, each carry a temperature T, fermentable extract S, ethanol E, lag-phase yeast L, active yeast in suspension X, growth reserve R and total vicinal diketones D, and exchange liquid at Q_{tm} and Q_{mc}. Each jacket holds glycol at $T_{j,z}$ behind an on/off valve u_z that passes F_j from a supply at T_{sup} (TT-501), fed by a plant of capacity Q_{cap} whose averaged load $\bar{Q}$ includes the other tanks' $Q_{other}(t)$. Probes TT-101 to TT-103 lag by τ_{TT}, DT-201 reads the gravity G at the middle zone every hour, and the laboratory reports a daily sample from the same valve t_{lab} after it is drawn. Settled yeast collects as the cake and CO_2 leaves at its evolution rate CER. Not to scale.

Table 10.1 Variables of the Tank by Role, With Their Values at Knock-Out

Variable	Role	Knock-out	Range
T_t, T_m, T_c	CVs, TT-101 to TT-103	11.0 °C	−1.5 to 21.3
T_{SP}	setpoint, supervisor	11.0 °C	−1 to 14
G	measured, DT-201	12.0 °P	2.10 to 12.0
D_{lab}, $\hat{D}_c$	laboratory, estimate	0 mg/L	0 to 0.93
u_t, u_m, u_c	MVs, valve duties	0 %	0–100
Q_{other}	DV, metered	300 kW	0–600
ℓ	DV, capacity lost	0 %	0–100
β_G	DV, meter bias, unmeasured	0 °P	±1
T_a	DV, ambient	15 °C	−10 to 35
valve fault	DV, event	none	shut or open
ΔQ_n, τ_n	DV, event, neighbor crash	220 kW, 30 h	—
v	parameter, unmeasured	95 %	5–100
T_z, S_z, E_z	states, per zone	11, 86.8, 0	see text
L_z, X_z, R_z, D_z	states, per zone	14.25, 0, 44, 0	see text
$T_{j,z}$	states, jackets	11.0 °C	−4 to 14
$\bar{Q}$, C_k	states, plant load, cake	300 kW, 0	—
12 ledgers	states	0	—

S_z, E_z in kg/m^3, L_z, X_z, R_z in Mcells/mL, D_z in mg/L.

10.3.1 Kinetics in Each Zone

Lag-phase yeast activates, active yeast grows on a reserve of new growth that the wort's oxygen and the yeast's sterols can support, and active yeast ferments the extract, as shown in Equation (10.1).

$$\begin{aligned} r_a &= k_L\, a(T)\, L, \qquad r_g = \mu_{11}\, a(T)\, X\, \frac{S}{K_x + S}\, \frac{R}{2 + R} \\ r_S &= q_s\, a(T)\, X\, \frac{S}{K_s + S}\, \max\left(0,\, 1 - \frac{E}{E_{max}}\right) \\ r_D &= f_D\, k_D\, r_g - f_R\, k_R\, a_D(T)\, X\, D \end{aligned} \tag{10.1}$$

where r_a and r_g are the activation and growth rates in Mcells/mL/h, r_S is the extract fermented in kg/(m^3 h), r_D is the net VDK rate in mg/(L h), the reserve's half-saturation is fixed at 2 Mcells/mL, every concentration is taken as zero if negative, and f_D and f_R are the strain's factors on the textbook formation and reduction constants. The zone balances are Equation (10.2).

$$\begin{aligned} \frac{\mathrm{d}S_z}{\mathrm{d}t} &= -r_S + \mathscr{M}_z[S], \qquad \frac{\mathrm{d}E_z}{\mathrm{d}t} = Y_E\, r_S + \mathscr{M}_z[E] \\ \frac{\mathrm{d}L_z}{\mathrm{d}t} &= -r_a - k_{s,z} L_z + \frac{V_{z^-}}{V_z} k_{s,z^-} L_{z^-} + \mathscr{M}_z[L] \\ \frac{\mathrm{d}X_z}{\mathrm{d}t} &= r_a + r_g - k_{s,z} X_z + \frac{V_{z^-}}{V_z} k_{s,z^-} X_{z^-} + \mathscr{M}_z[X] \\ \frac{\mathrm{d}R_z}{\mathrm{d}t} &= -r_g + \mathscr{M}_z[R], \qquad \frac{\mathrm{d}D_z}{\mathrm{d}t} = r_D + \mathscr{M}_z[D] \\ \frac{\mathrm{d}C_k}{\mathrm{d}t} &= k_{s,c}\,(X_c + L_c)\,\frac{V_c}{V_T} \end{aligned} \tag{10.2}$$

where the rates are evaluated in zone z, z^- is the zone above it (settling enters the middle zone from the top and the cone from the middle), C_k is the cake expressed over the tank volume $V_T = 300\,\mathrm{m}^3$, and the exchange with the neighbors is

$$\mathscr{M}_z[\psi] = \frac{1}{V_z} \sum_n Q_{zn}\,(\psi_n - \psi_z) \tag{10.3}$$

over the one or two neighbors n of zone z. Ethanol and CO_2 follow the Balling split, in which 2.0665 kg of extract make 1 kg of ethanol, 0.9565 kg of CO_2 and 0.11 kg of yeast and by-products, so $Y_E = 0.484$ and $Y_C = 0.463$, and the tank's CO_2 evolution is $\mathrm{CER} = Y_C \sum_z r_{S,z} V_z$ in kg/h.

10.3.2 Temperature, Settling and Exchange

The yeast's activity is a double Arrhenius law, growth against thermal deactivation, normalized to one at $T_r = 11\,^\circ\mathrm{C}$, and VDK reduction has a single activation energy of its own, as shown in Equation (10.4).

$$\begin{aligned} a(T) &= \max\left(0,\, \frac{\phi_1(T) - \delta\,\phi_2(T)}{1 - \delta}\right), \qquad a_D(T) = \phi_D(T) \\ \phi_i(T) &= \exp\left[-\frac{E_i}{R_g}\left(\frac{1}{T} - \frac{1}{T_r}\right)\right], \quad \delta = \frac{E_1}{E_2}\,\frac{\phi_1(T_{opt})}{\phi_2(T_{opt})} \end{aligned} \tag{10.4}$$

with temperatures in kelvin, R_g the gas constant, $E_1 = 60\,\mathrm{kJ/mol}$, $E_2 = 250\,\mathrm{kJ/mol}$, $E_D = 110\,\mathrm{kJ/mol}$ and $T_{opt} = 30\,^\circ\mathrm{C}$, where δ makes a largest at T_{opt}. The activity is 1.30 at 14 °C, 3.74 at 30 °C and zero at 40 °C, and a_D is 1.63 at the 14 °C rest and 0.128 at −1 °C. Suspended yeast settles faster when cold, when the sugar is gone and when the gas stops stirring, and neighbors exchange liquid at a rate

that rises with the CO_2 and with a density inversion, as shown in Equation (10.5).

$$\begin{aligned} k_s &= \frac{k_{set}\,\mathrm{e}^{-\theta_s(T-11)}\,K_f}{(K_f+S)\,(1+\mathrm{CER}/C_{ref})} \\ Q &= Q_{free}+k_c\,\mathrm{CER}+k_b\left[\rho(T_{up})-\rho(T_{low})\right]^+ \\ \rho(T)-\rho_{max} &= -0.007\,(T-4)^2 \end{aligned} \tag{10.5}$$

where T is in °C, ρ in kg/m^3, $[x]^+ = \max(0,x)$, and T_{up} and T_{low} are the zones above and below a boundary. The density maximum at 4 °C is water's, an illustrative stand-in for beer's, so below 4 °C the colder liquid is the lighter. The settling rate is 0.015 h^{-1} at 11 °C without sugar or gas and 0.30 h^{-1} at −1 °C, and the exchange is 3 m^3/h in a still tank, 333 at 110 kg/h of CO_2 and 150 across a boundary with 1 °C over 0 °C.

10.3.3 Heat, Jackets and the Shared Plant

Each zone gains the heat of fermentation and the ambient gain and loses heat to its jacket, as shown in Equation (10.6).

$$\begin{aligned} \rho c_p V_z \frac{\mathrm{d}T_z}{\mathrm{d}t} = 3600\Big[&q_{f,z}-UA_z\,(T_z-T_{j,z}) \\ &+UA_a\frac{V_z}{V_T}\,(T_a-T_z)\Big]+\rho c_p V_z\,\mathscr{M}_z[T] \end{aligned} \tag{10.6}$$

where $\rho c_p = 4100\,\mathrm{kJ/(m^3\,K)}$, the heat terms are in kW, the factor 3600 converts them to kJ/h, $q_{f,z} = q_f\,r_{S,z}V_z/3600$ is the heat released in the zone, and $q_f = 500\,\mathrm{kJ/kg}$ is the declared heat of fermentation. The tank holds 1.23 GJ/K. Each jacket is a glycol holdup mixed at its outlet temperature, as shown in Equation (10.7).

$$M_{j,z}c_g\frac{\mathrm{d}T_{j,z}}{\mathrm{d}t} = 3600\left[u_z F_j c_g\,(T_{sup}-T_{j,z})+UA_z\,(T_z-T_{j,z})\right] \tag{10.7}$$

where u_z is 0 or 1 for an on/off valve and the position for a modulating one, $c_g = 3.6\,\mathrm{kJ/(kg\,K)}$ for 35 % propylene glycol, $F_j = 2.5\,\mathrm{kg/s}$, and the holdup is $M_j = 600\,\mathrm{kg}$ for each cylinder jacket and $M_j UA_c/UA_t = 369\,\mathrm{kg}$ for the cone. A shut valve keeps cooling until the holdup has warmed to the beer, with a time constant of 5.5 min, and an open one settles in 2.3 min (1.7 min for the cone) at $q = UA_{eff}(T-T_{sup})$, with $UA_{eff} = UA\,F_j c_g/(UA+F_j c_g)$ of 3.77, 3.77 and 2.77 kW/K. With every valve open, beer at 11 °C and glycol at −4 °C, the jackets remove 155 kW. The plant is shared, and its buffer averages the loads it is sent, as shown in Equation (10.8).

$$\begin{aligned} \tau_g\frac{\mathrm{d}\bar{Q}}{\mathrm{d}t} &= Q_{other}(t)+\sum_z u_z F_j c_g\,(T_{j,z}-T_{sup})-\bar{Q} \\ T_{sup} &= T_{g,0}+k_{load}\,\max\left(0,\ \bar{Q}-(1-\ell)Q_{cap}\right) \\ Q_{other}(t) &= Q_{base}+\Delta Q_n\,\mathrm{e}^{-(t-t_n)/\tau_n},\quad t\ge t_n \end{aligned} \tag{10.8}$$

where $\tau_g = 0.5\,\mathrm{h}$, $T_{g,0} = -4\,°\mathrm{C}$, $Q_{cap} = 500\,\mathrm{kW}$, $k_{load} = 0.03\,\mathrm{K/kW}$, ℓ is the share of capacity lost, Q_{base} is the other tanks' base load and a neighbor that starts crashing at t_n adds ΔQ_n decaying over τ_n (220 kW over 30 h for the event in the interface).

10.3.4 Gravity, Alcohol and Freezing

The fermentable extract at knock-out is $S_0 = \varphi\,\varepsilon(\mathrm{OE})$, with $\varepsilon(P) = 10P\,\mathrm{sg}(P)$ in kg/m^3 and $\mathrm{sg}(P) = 1+P/(258.6-227.1P/258.2)$, which for a 12 °P wort and $\varphi = 0.69$ is 86.8 kg/m^3, 26.04 t in the

Table 10.2 Parameters of the Tank. V Marks a Slider of the Show Model Dialog, H a Value Kept From It, and F a Constant of the Code

Symbol	Meaning	Value	Class
OE, φ	original extract, fermentable share	12 °P, 0.69	V
G_f	target apparent extract	2.6 °P	V
R_{O_2}	growth the wort's oxygen supports	44 Mcells/mL	V
X_0, v	pitch rate, viability	15 Mcells/mL, 95 %	V
T_0	pitching temperature	11 °C	V
k_L, μ_{11}	activation and growth at 11 °C	0.05, 0.052 h^{-1}	V
K_x	extract half-saturation of growth	4 kg/m^3	H
n_{gen}	doublings of a viable cell	3	V
q_s	uptake per cell at 11 °C	0.0165[a]	V
K_s	extract half-saturation of uptake	8 kg/m^3	V
E_{max}	ethanol that stops uptake	120 kg/m^3	H
k_D	VDK formed per cell grown	0.021[b]	V
k_R	VDK reduction per cell at 11 °C	0.00026[c]	V
f_D, f_R	strain factors, estimator assumes 1	1.15, 0.8	V
E_1, E_2, E_D	activation energies	60, 250, 110 kJ/mol	F
k_{set}	settling, 11 °C, no sugar or gas	0.015 h^{-1}	V
θ_s	settling increase per kelvin colder	0.25 K^{-1}	V
K_f, C_{ref}	sugar and CO_2 that halve settling	6[d], 15 kg/h	H
Q_{free}	free-convection exchange	3 m^3/h	V
k_c, k_b	exchange per CO_2, per inversion	3, 3000[e]	V
q_f	heat per kg of extract fermented	500 kJ/kg	F
ρc_p, c_g	heat capacities, beer and glycol	4100, 3.6[f]	F
UA_t, UA_m, UA_c	jacket conductances	6.5, 6.5, 4.0 kW/K	V
F_j, M_j	open-valve flow, cylinder holdup	2.5 kg/s, 600 kg	V
$T_{g,0}$, Q_{cap}	supply in capacity, plant capacity	–4 °C, 500 kW	V
k_{load}	supply warming per kW over	0.03 K/kW	V
τ_g, UA_a	plant buffer lag, ambient gain	0.5 h, 0.5 kW/K	H
τ_{TT}, t_{lab}	probe lag, laboratory turnaround	10 min, 24 h	V
t_{plan}	plan from knock-out to finish	12 days	V
V_t, V_m, V_c	zone volumes	130, 130, 40 m^3	F

[a] kg/(m^3 h) per Mcells/mL. [b] mg/L per Mcells/mL. [c] h^{-1} per Mcells/mL. [d] kg/m^3.
[e] m^3/h per kg/h and per kg/m^3. [f] kJ/(m^3 K), kJ/(kg K).

tank. The gauge reads degrees Plato through Balling's relations [16], as shown in Equation (10.9).

$$
\begin{aligned}
f &= \frac{\mathrm{OE}\,(S_0 - \bar{S})}{\varepsilon(\mathrm{OE})}, & w &= 100 - \frac{1.0665\,f}{2.0665} \\
n &= \frac{100\,(\mathrm{OE} - f)}{w}, & A &= \frac{100\,f}{2.0665\,w} \\
G &= \frac{n - 0.1808\,\mathrm{OE}}{0.8192}
\end{aligned}
\tag{10.9}
$$

where $\bar{S}$ is the volume-weighted mean extract, f the extract fermented in °P, n the real extract, A the alcohol by mass and G the apparent extract. Alcohol by volume is $A\,\mathrm{sg}(G)/0.7907$, and the beer's freezing point, used only to flag ice, is an illustrative $-0.42A - 0.04n$ in °C. The target of 2.6 °P is reached when 95.1 % of the fermentable extract is gone, and complete fermentation would leave 2.10 °P with the freezing point at –1.91 °C.

10.3.5 Knock-Out, Mid-Batch Starts and the Integrator

Table 10.2 gives the parameter values. A batch starts at knock-out with every zone and jacket at T_0, $S = S_0$, $E = 0$, the viable pitch $X_0 v/100 = 14.25\,\mathrm{Mcells/mL}$ as lag-phase yeast, $X = 0$, the reserve $R_0 = \min\left(R_{O_2}, X_0 v(2^{n_{gen}} - 1)/100\right) = 44\,\mathrm{Mcells/mL}$, $D = 0$, $\bar{Q}$ equal to the other tanks' load and the ledgers at zero. The wort, the pitch, n_{gen} and t_{lab} are read when a batch starts, so a change takes effect before the run or after a Reset, and the gauge keeps the knock-out OE. Cases 5 to 7 start part-way through a batch. Their history from knock-out is integrated first with every zone held exactly

on the recipe temperature, using a 3 min midpoint step on the kinetics, the free rise at its trigger, the rest ended when the true VDK is below 0.10 mg/L and the true gravity at its target, and a laboratory sample recorded at each whole day. At hand-over each jacket is set to remove its zone's heat of fermentation and ambient gain q_z, with $T_{j,z} = T_{SP} - q_z/UA_z$ and nothing removed during a free rise, and the controller replays the estimator from knock-out on the history's temperatures, passes the results already reported through its bias update and queues those still at the laboratory.

The engine of Section 1.6.1 integrates the plant on a fixed 30 s step, the controllers execute once a minute, the historian records every 5 min, and the default seed is 20261102. The speed selector spans 0.25 to 12 simulated hours per second, one by default, so a twelve-day batch takes 24 s at the top speed, and the trend windows are 12, 48, 168 and 336 h. Halving the step leaves the crash completion of case 3 on the same control step and moves its finishing VDK by 2.1×10^{-5} mg/L and its worst recipe deviation by 0.0011 °C. Over 280 h of each of the nine scenarios the ledgers close to 2.8×10^{-14} relative in case 3 and to 2.9×10^{-11} at worst, and the heat released in case 3, 13.01 GJ from 26.03 t of extract, is the declared 0.5 MJ/kg. The app's own checks require closure below 1×10^{-9}.

10.3.6 Assumptions

The liquid is three well-mixed zones with exchange across two boundaries, while a real tank has continuous temperature and velocity fields set by convection [13]. The zones keep the one fact the lessons need, that the tank is stirred while the yeast are busy and stratifies when they are not, and they fail for the position of a probe within a zone.

The kinetic constants are illustrative, calibrated so that a 12 °P wort held at 11 °C reaches 2.6 °P in about seven days (on day 6.95 in the engine), and are not fitted to a strain. Growth has one oxygen and sterol reserve, and flocculation is an empirical rate. Total VDK is one state formed with growth and reduced by suspended yeast, so the precursor's spontaneous conversion outside the cell, which usually limits the real removal [2], is lumped into one rate with a high activation energy, and the laboratory is assumed to measure that total. The model therefore cannot show diacetyl rising again after a crash from precursor left in the beer.

The heat of fermentation is the declared 0.5 MJ/kg, the heat capacity, density and volume are constant although about 4 % of the mass leaves as CO_2, and the density maximum is water's. Each jacket is one well-mixed holdup behind a valve with no stroke time. The glycol plant is a capacity and a buffer whose supply temperature is an algebraic function of the averaged load above capacity, so the tripped plant of case 4 settles near 6.5 °C instead of warming for as long as the trip lasts, and the other tanks are a load profile that does not respond to the supply temperature. The density meter reads one valve and the laboratory one sample a day, and neither drifts except through the bias disturbance. Ice is flagged below the freezing point and not modeled, and pH, pressure, temperature shock and yeast health beyond viability are outside the model. Every size, capacity and limit was chosen for teaching and is not a design basis.

10.4 Implemented Control System

Three zone loops, TIC-101 to TIC-103, a glycol coordinator, a recipe supervisor and a VDK estimator run every $\Delta t_c = 1$ min, two physics steps, at 1, 2, 3 minutes and so on after a priming execution at $t = 0$, in the order of Figure 1.10. After the integrator advances, the app's own bookkeeping of ice, plant capacity and recipe deviation runs before the outputs and the historian are updated. No part of the controller reads a zone's contents.

10.4.1 Instruments

Each thermowell probe lags its zone's true temperature to first order with $\tau_{TT} = 10$ min, adds first-order colored noise of 0.02 °C standard deviation and 3 min correlation time, and is read to 0.01 °C. With the one-probe option only TT-102 is installed. The density meter DT-201 samples the middle zone every hour through a sample line of 3 min dead time, reports 3 min later with 0.05 °P of noise,

and adds a bubble spike of ±0.6 °P with probability 0.01 per report and any calibration bias. A reading more than 0.4 °P from the last accepted one is rejected unless three in a row fall on the same side, and a reading that repeats exactly or is overdue by three cycles is stale. Over the batch of case 3, to the completion of its crash, the meter reports 250 times, one report carries a spike, and the gate rejects it. The laboratory analyzes a sample from the same valve at every whole day of batch age from day 1 and reports it $t_{lab} = 24\,\text{h}$ later with 0.01 mg/L of noise. The meter and the laboratory are instances of the framework's shared sampled analyzer, and the estimator's correction is the framework's shared bias update, the logarithmic form of Equation (1.8). TT-501 reads the supply through a 3 min lag, and the glycol meters report this tank's return heat through a 10 min filter and the other tanks' through 5 min.

10.4.2 Zone Loops and the Time-Proportioned Output

All three loops run the velocity-form PID law of Section 1.6.2, written for these loops in Equation (10.10), where s_k is the running sum S_k of Equation (1.13) and d_k the limited output u_k.

$$\begin{aligned} e_k &= \gamma(\text{SP}_k - \text{PV}_k), \qquad P_k = K_c\, e_k \\ s_k &= s_{k-1} + (P_k - P_{k-1}), \qquad v_k = s_k + I_{k-1} \\ d_k &= \text{clamp}(v_k,\, 0,\, 100) \\ I_k &= I_{k-1} + \left[\frac{K_c}{\tau_I}\, e_k + \frac{d_k - v_k}{\tau_I}\right]\Delta t_c \end{aligned} \tag{10.10}$$

where $\gamma = -1$ because the valve must open as the zone warms, so that a zone warmer than its setpoint raises its duty d_k in percent. The loops are therefore direct acting in the convention of Section 1.3, since more glycol lowers the temperature. The app's source and its own checks use the same name. The output bias is zero and the loops have no derivative action, so s_k is P_k. The integral is protected by back-calculation with τ_I as the tracking time, the duty has no rate limit, and in MANUAL the integral tracks the operator's duty d_{op}, all as in Section 1.6.2. The shipped tunings are $K_c = 150\,\%$ per °C with $\tau_I = 8\,\text{h}$ for the top and middle loops and $K_c = 60\,\%$ per °C with $\tau_I = 8\,\text{h}$ for the cone, from sliders of 0 to 400 and 0.5 to 48 h.

The duty becomes an on-time. At the start of every cycle of $N = 10$ ticks of 1 min, the latched duty $d \in [0, 1]$ sets the number of open ticks, as shown in Equation (10.11).

$$n_{on} = \text{round}(dN + c), \qquad c \leftarrow dN + c - n_{on} \tag{10.11}$$

where the carry c holds the rounding error and any correction the 2 min minimum on and off times impose. An on-time of one tick becomes zero or two, whichever is nearer to the demand, one of nine ticks becomes eight, and a duty of 0 or 1 shuts or opens the valve for the whole cycle and clears the carry. The three outputs are staggered by 0, 3 and 7 min so that the valves do not open together. Over 200 cycles the mean opening equals duties of 3, 37 and 97 % exactly, where the app's own check asks for 0.002. In the hysteresis mode each valve opens when its probe reads above $\text{SP} + d_b/2$ and shuts below $\text{SP} - d_b/2$, with $d_b = 0.5\,^{\circ}\text{C}$, and the PI tracks the relay. In the modulating mode the valve is positioned at $u = d/100$. With one mid probe, TIC-102 computes the duty for all three valves and the other two loops track it.

10.4.3 Glycol Coordinator

The coordinator acts only from the start of the crash to the end of the batch. It takes as this tank's allotment the plant's capacity less the other tanks' metered load and a reserve, $A = \max\big(0,\, (1 - \ell)Q_{cap} - \hat{Q}_{other} - Q_{res}\big)$ with $Q_{res} = 20\,\text{kW}$, and integrates a scale factor κ on every duty from this tank's metered draw $\hat{Q}_{tank}$, as shown in Equation (10.12).

$$\kappa \leftarrow \text{clamp}\Big(\kappa + \frac{A - \hat{Q}_{tank}}{\max(20, A)}\,\frac{\Delta t_c}{0.25\,\text{h}},\; 0,\; 1\Big) \tag{10.12}$$

Table 10.3 Recipe Supervisor. Every Transition Is Decided on a Measurement or the Batch Clock

Phase	Setpoint	Ends when
primary	$T_{pri} = 11\,^{\circ}\mathrm{C}$	by the chosen trigger, DT-201 reads $G_{rise} = 6.0\,^{\circ}\mathrm{P}$, or day $d_{rise} = 6$ arrives
free rise	$+r_{rise} = 0.07\,^{\circ}\mathrm{C/h}$	the setpoint reaches $T_{rest} = 14\,^{\circ}\mathrm{C}$
rest	T_{rest}	by the chosen rule, $\hat{D}_c < D_{lim} = 0.10\,\mathrm{mg/L}$, or a laboratory result below D_{lim} for a sample drawn in the rest, each with DT-201 at or below G_f, or $t_{rest} = 48\,\mathrm{h}$ have passed
crash	$-r_{crash} = 0.3\,^{\circ}\mathrm{C/h}$ to $T_{crash} = -1\,^{\circ}\mathrm{C}$, waiting while a probe is 0.5 K behind, the coordinator limits, or TT-501 is above $T_{crash} - 1\,\mathrm{K}$	the setpoint is at T_{crash} and every probe at or below $T_{crash} + 0.5\,\mathrm{K}$
maturation	T_{crash}	$t_{mat} = 24\,\mathrm{h}$ have passed
held	$+r_{rise}$ to $\max(T_{pri}, T_{rest})$	the operator decides

It feeds back on the meter rather than on a jacket model, because a time-proportioned jacket removes more heat than its duty suggests (Section 10.5.2). Hysteresis valves cannot be scaled, so the warmest zones are opened first while $\lfloor 3\kappa \rfloor$ remain. The cooling of the fermentation itself is never cut.

10.4.4 Recipe Supervisor

The supervisor publishes the setpoint of all three loops and moves through the phases of Table 10.3. The free rise has no heater behind it, so a zone keeps pace with the rising setpoint only while the fermentation releases enough heat. From day 3, once DT-201 has twelve accepted hourly readings and reads more than 0.1 °P above the target, the supervisor projects the finish from the measured trend, as shown in Equation (10.13).

$$t_{fin} = t + \frac{G_{DT} - G_f}{-\dot{G}\, a(T_{hi})/a(\bar{T})} + \frac{T_{rest} - T_{crash}}{r_{crash}} \tag{10.13}$$

where $\dot{G}$ is the least-squares slope of the last 24 accepted readings, $\bar{T}$ is the mean of the installed probes and $T_{hi} = \max(T_{pri}, T_{rest})$ is the warmest temperature the recipe permits. A projection beyond the plan declares the fermentation stuck and holds the batch. During the crash the supervisor projects the completion by cooling the tank as one lump from $\bar{T}$ to $T_{crash} + 0.5\,\mathrm{K}$ at $q = \min\left(A,\ \sum_z UA_{eff,z}(T - T_{sup})\right)$ in steps of 0.25 h, capped at 1000 h, and announces a late crash when the projection passes the plan, with 3 h of hysteresis. The ranges of the recipe sliders are 8 to 14 °C, 10 to 18 °C and −2 to 4 °C for the three temperatures, 0.02 to 0.3 °C/h for the rise and 0.05 to 1 °C/h for the crash, with 3 to 9 °P for the trigger gravity, day 2 to 10 for the timed rise, 12 to 120 h for the timed rest, 0 to 96 h for the maturation, 0.05 to 0.20 mg/L for the VDK limit, 2 to 30 min for the cycle, 0.1 to 2 °C for the deadband and 0 to 100 kW for the reserve.

10.4.5 VDK Estimator and Its Laboratory Correction

The estimator integrates Equation (10.1) for one well-mixed zone by explicit Euler at Δt_c, on the volume-weighted mean of the installed probes, from the declared pitch with an assumed viability of 95 % and the textbook factors $f_D = f_R = 1$, and records $\ln \hat{D}$ every hour. A laboratory result is compared with what the model said when its sample was drawn, a day before the result arrived, and the correction is an exponentially weighted average of the log ratio, as shown in Equation (10.14).

$$\begin{aligned} b &\leftarrow b + \lambda\left[\ln D_{lab} - \ln \hat{D}(t_s) - b\right], \qquad |b| \le 1.5 \\ \hat{D}_c(t) &= \hat{D}(t)\,\mathrm{e}^{b} \end{aligned} \tag{10.14}$$

where t_s is the sampling time, D_{lab} is floored at 0.005 mg/L, and $\lambda = 0.5$ by default on a slider from 0 to 1, with $\lambda = 0$ ignoring the laboratory. With a constant model error e the bias approaches it as $e[1 - (1 - \lambda)^k]$ after k results, which the app checks to 1×10^{-12}.

Table 10.4 The Eight Cases and the Graded Batch

#	Scenario and event	Teaching point
1	Manual, valves shut, from knock-out	the exotherm arrives
2	Hysteresis valves, $d_b = 0.5\,°C$	a limit cycle and its lags
3	Time-proportioned PI, 10 min cycle	on/off valves, no cycle
4	16 °P wort, plant trips 42 to 54 h	capacity at the peak
5	One mid probe, crash from day 7.95	a quiet tank stratifies
6	Laboratory-ended rest, from day 4.18	a result a day old
7	Crash, others at 478 kW of 500	a shared utility
8	Viability 15 %	honest infeasibility
9	Twelve days, neighbor crash on day 7	graded challenge

> ✦ DID YOU KNOW? **A Flavor Threshold Near the Limit**
>
> A review of beer aroma compounds puts the odor threshold of diacetyl at 0.1 to 0.2 mg/L in lagers and 0.1 to 0.4 mg/L in ales, and above it the beer smells and tastes spoiled [17]. The rest of this chapter's recipe ends below 0.10 mg/L, at the bottom of the lager range.

10.4.6 Plant, Loop Response and Controller Models

Everything in Section 10.3 belongs to the plant, including this strain's $f_D = 1.15$ and $f_R = 0.8$ and whatever viability the scenario sets. No control model is identified by the application, and open-loop steps in MANUAL measure the plant that a loop sees. From the state of case 3 at 36 h, with the duties held where the loops left them, a step of 10 % on the middle valve's duty cools the top, middle and cone zones by 0.142, 0.150 and 0.150 °C after 6 h, because 234 to 302 m^3/h of exchange mixes the tank over those hours. From the start of case 5, with the CO_2 gone and every duty at 50 %, the same step cools them by 0.027, 0.148 and 0.065 °C. The loop sees close to an integrator of 0.025 °C/h per 10 % of one jacket's duty while the tank is stirred, and a zone of its own once it is not. The controller's internal models are the estimator's kinetics with textbook constants and an assumed viability, the activity ratio and the lumped cooling curve in the supervisor's two projections, and the capacity and reserve in the coordinator's allotment. None of them contains the strain's own diacetyl constants or the mixing.

10.5 Cases and What They Measure

The eight cases and the graded batch appear in Table 10.4. Every case begins from the same reference settings, three probes on time-proportioned valves, the free rise at 6.0 °P on the density meter, the rest ended on the corrected estimate and the coordinator on, and each alters only what its row states. The engine supplied every number below, at the default seed and without the interface, stepped exactly as the page steps it, with each change made after the case is loaded and before it runs (Section 1.6.4), which gives the same run as the change carried through a Reset, and days count batch age from knock-out. The scorecard of each case run to the completion of its crash is collected in Table 10.5. Its worst deviation of 0.51 to 0.53 °C in most rows is set by the crash rules, which let a probed zone trail the ramp by 0.5 K before the ramp waits and complete the crash with every probe within 0.5 K of −1 °C.

> ✦ DID YOU KNOW? **Many More Strokes, the Same Batch**
>
> From the simulator: run to the completion of the crash, hysteresis valves in case 2 and the time-proportioned PI loops in case 3 finish on days 10.28 and 10.39, both at 2.10 °P, with 0.090 and 0.089 mg/L of VDK and the same worst deviation of 0.52 °C, while the valves switch 81 times against 2783 (Table 10.5).

Table 10.5 Scorecard of Each Case Run to the Completion of Its Crash. *G* and *D* Are the True Tank Values Then, and the Worst Deviation Is Any Zone From the Published Setpoint

Case and configuration	Day	*G* (°P)	*D* (mg/L)	Dev. (°C)	Over (kW, h)	Switches
2 hysteresis	10.28	2.10	0.090	0.52	0	81
3 time-proportioned	10.39	2.10	0.089	0.52	0	2783
3 modulating	10.39	2.10	0.089	0.51	0	—
3 one probe	10.21	2.10	0.090	1.38	0	2485
4 trip[a]	9.63	2.90	0.073	0.52	348.5, 12.0	3379
5 one probe	10.73	2.10	0.071	1.95	0	873
5 three probes	10.93	2.10	0.070	0.52	0	1199
6 laboratory	11.99	2.10	0.044	0.52	0	1417
6 estimate	10.37	2.10	0.089	0.53	0	1399
6 estimate, $\lambda = 0$	9.64	2.12	0.123	0.52	0	1331
6 rest of 48 h	10.95	2.10	0.069	0.52	0	1413
7 coordinator on	12.71	2.10	0.040	0.52	0	2129
7 coordinator off	11.11	2.10	0.067	0.51	70.9, 41.5	1225
8 viability 15 %[b]	held	4.72	—	1.41	0	—
3 at viability 45 %[c]	11.36	2.11	0.108	0.52	0	2789
3 at viability 75 %[c]	10.66	2.11	0.099	0.51	0	2787
9 as shipped	12.57	2.10	0.058	2.73	0	99
9 suggested	10.47	2.10	0.087	0.52	0	2825

[a]Target 3.5 °P, and the plant has no capacity during the trip. [b]*G* on day 12, deviation to day 12.5. [c]Viability set before the run. Cases 5 to 7 count switches from their start.

10.5.1 The Exotherm With the Valves Shut

Case 1 pitches 15 Mcells/mL of 95 % viable yeast into the 12 °P wort at 11 °C with all three valves shut, in MANUAL. The fermentation releases 5.7 kW at 12 h and 13.7 kW at 24 h, which against 1.23 GJ/K warms the tank only to 11.17 °C and 11.56 °C, and the first degree takes until 31.9 h. Then the heat compounds, because $a(T)$ is 1.30 times larger at 14 °C than at 11 °C. The beer is at 13.61 °C at 48 h and releasing 40.6 kW, against the 32.8 kW peak at 47 h of the same batch held at 11 °C, it is at 16.61 °C at 72 h with 44.8 kW, and it peaks at 21.33 °C at 127.2 h. At 72 h the hot tank is at 7.04 °P and the controlled one at 8.13 °P, and the supervisor, still publishing its recipe, starts the free rise on day 3.38 without being obeyed. Opening all three valves at 24 h cools the tank from 11.56 to 9.29 °C in 6 h and to 7.39 °C in 12 h, and fixed duties of 20, 20 and 10 % from knock-out overcool it to 8.35 °C at 24 h and 6.13 °C at 72 h. The jackets can remove far more than the yeast make, and the task is to open them for the right share of the time.

> **TRY IT IN THE APP** **1 · Manual Valves on Day One**
>
> Load the case and press run. All three valves stay shut. Watch the beer temperature on the Tank view: it barely moves for a day, warming only 0.56 °C, and then runs away, to 13.61 °C at 48 h, because warmer yeast work faster and release more heat.

10.5.2 A Thermostat, a Pulse Train and a Positioned Valve

Case 2 opens each valve above 11.25 °C and shuts it below 10.75 °C. Between days 1 and 3 the zones swing from 10.66 to 11.27 °C and the valves switch 29 times in 48 h, with about 8 h (7.4 to 9.4 h) between successive openings of a valve. The bottom of the swing lies 0.09 °C below the lower switch point, mainly because the probe reports the cooling 10 min late. The glycol left in the jacket after the valve shuts adds little, as the holdup runs below show. The cycle is the relay limit cycle, whose amplitude and period the plant sets [4]. A 0.2 °C deadband gives 10.80 to 11.12 °C with 62 switches and cycles of 3.3 to 4.7 h, and a 1.0 °C deadband gives 10.43 to 11.52 °C with 17 switches and cycles of 13.8 to 14.7 h. The holdup barely matters in a tank this large, since 150 or 1500 kg per jacket puts the bottom of the swing at 10.66 or 10.65 °C, while the probe lag does matter, with 10.70 to 11.25 °C

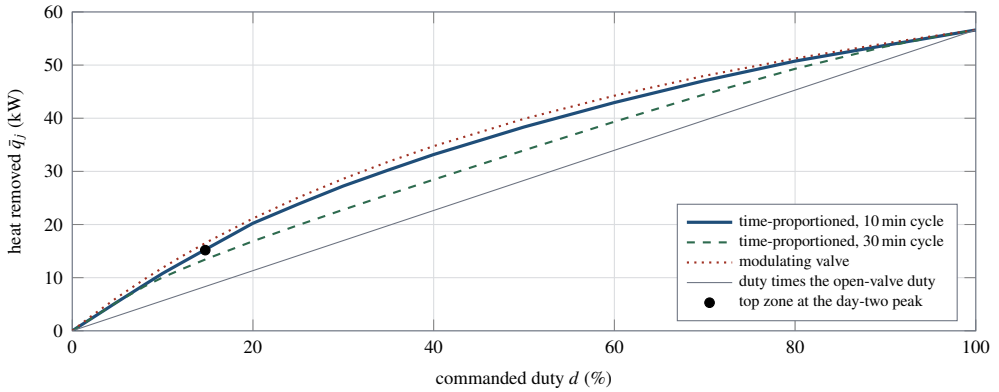


Figure 10.2 Heat removed by one cylinder jacket against its commanded duty, with the beer held at 11 °C and the glycol at −4 °C: the time-proportioned output on a 10 min cycle (thick solid line) and a 30 min cycle (dashed), the modulating valve (dotted) and the duty times the open-valve heat (thin solid). The curves come from the jacket equation and the application's own time-proportioned output averaged over 180 h, and the round marker is the engine's top zone at the day-two peak of case 3.

at 2 min and 10.55 to 11.29 °C at 30 min. These swings are read at the 30 s integration step, and the 5 min log agrees with them within 0.003 °C.

> **TRY IT IN THE APP** **2 · On/Off With Hysteresis**
>
> Run the case to day 3 and watch the zone temperatures on the trend. Each valve opens above 11.25 °C and shuts below 10.75 °C, and the zones swing from 10.66 to 11.27 °C with about 8 h between openings of a valve. Then load case 3, whose time-proportioned PI loops hold the same valves between 10.97 and 11.04 °C.

Case 3 drives the same valves from the PI loops through the ten-minute cycle. Between days 1 and 3 the zones stay between 10.97 and 11.04 °C, nearly nine times tighter, at 802 switches in 48 h against 29. A 30 min cycle gives 10.96 to 11.05 °C with 492 switches, a 5 min cycle 10.97 to 11.04 °C with 810, and a modulating valve 10.99 to 11.03 °C. The pulses are smoothed twice before they reach the beer, by the glycol held in the jacket and by the 1.23 GJ/K of the tank, so the cycle length barely matters to the temperature.

The duty is not a heat meter. At the day-two peak, between 44 and 50 h, the top valve is open 14.7 % of the time and its jacket removes 15.2 kW, 1.8 times the open share times the 56.7 kW that an open valve removes at steady state, because each pulse fills the jacket with glycol at the supply temperature that goes on cooling after the valve shuts. Figure 10.2 gives the whole characteristic for one cylinder jacket. It is concave, like a quick-opening valve, with a slope of about 1.1 kW per percent of duty near zero and 0.3 kW near full opening, and on the ten-minute cycle it follows the modulating valve within 1.6 kW. The loops therefore see an actuator gain that falls by more than three times between the primary fermentation, where they ask for 12 to 15 % at the peak, and the crash, where the top and middle valves stay fully open.

Pause. Try Exercise 10.1 before reading on.

10.5.3 Sizing for the Peak or for the Crash

Case 4 ferments a 16 °P wort pitched at 24 Mcells/mL with the oxygen for 65 Mcells/mL of growth and trips the glycol plant from 42 to 54 h. Without the trip its heat peaks at 49.2 kW at 43.9 h, against 32.8 kW for the 12 °P batch, while every open valve could remove 155 kW, and the loops never ask

for more than 32 % or let a zone past 11.05 °C. With the trip the supply warms to 5.8 °C within half an hour and to 6.46 °C before the restart, the capacity of the open jackets falls to 51 to 56 kW while the fermentation releases 50.0 kW at 44.9 h, the top and middle valves reach 100 % by 48.5 h, loops are saturated for 6.45 h in all, and the beer reaches 11.42 °C at the restart. The margin and the thermal mass carry the batch through twelve hours without a plant.

The jackets are sized for the crash. Cooling 300 m^3 from 14 to −1 °C removes 18.45 GJ, and over the batch of case 3 the jackets remove 31.03 GJ against the 13.01 GJ the fermentation releases. The crash ramp of 0.3 °C/h asks for 102.5 kW, and the open jackets can deliver that only while the beer is above about 5.9 °C. The ramp runs unhindered until day 8.63, then waits 21.3 h in all for the slowest zone, the top and middle valves stay at full duty for about 50 h, the two-hour average of the jackets' heat peaks at 111.7 kW on day 7.77, and the crash that started on day 7.42 is complete on day 10.39.

10.5.4 One Probe in a Quiet Tank

Case 5 starts on day 7.95 as the crash begins, with TT-102 the only probe and TIC-102 driving all three valves. With the CO_2 gone the zones exchange 7.4 m^3/h after 6 h and 3.8 after 24 h. The cone has 0.069 kW/K of open jacket per cubic meter against 0.029 for the cylinder zones, so the duty that suits the middle overcools it. It is 1.95 °C below the recipe on day 8.98, and the zones spread by 2.69 °C on day 9.12. Once the cone is colder than 4 °C and further from 4 °C than the middle, it is the lighter, and the exchange across its boundary rises to 20.3 m^3/h at 36 h and 28.4 at 48 h. The cold cone drops its yeast early, 18.1 Mcells/mL of cake over the tank volume 36 h into the crash against 14.7 Mcells/mL with three probes. The crash is declared complete on day 10.73, when TT-102 reads −0.51 °C, with the zones at −0.39, −0.49 and −0.63 °C. With a probe and a loop per zone the spread stays within 0.47 °C, every zone within 0.52 °C of the recipe, and the crash completes on day 10.93. The same single probe is harmless while the yeast are busy. Over the whole batch of case 3 with one probe the spread stays below 0.10 °C through the primary fermentation and the rest, first passes 0.5 °C on day 7.53, 2.7 h into the crash, when the CO_2 has fallen to 7.2 kg/h, and the cone runs up to 1.38 °C off the recipe.

🕮 **Pause.** Try Exercise 10.2 before reading on.

10.5.5 The Diacetyl Rest and a Result a Day Old

Case 6 starts on day 4.18 at the free-rise trigger, with the rest ended on the laboratory. The beer reaches 14 °C on day 5.97. The daily samples read 0.242 mg/L on day 1, 0.923 on day 2, then 0.681, 0.505, 0.377, 0.253, 0.179 and 0.083 on day 8, each reported a day after it was drawn. The true VDK falls below 0.10 mg/L on day 7.96, the result for day 8 arrives on day 9.00 and the crash starts then, with the true VDK at 0.062 mg/L, and the beer is crashed on day 11.99 at 0.044 mg/L, just inside the plan. The run 120 h after its start, 4.4 h into that crash, is shown in Figure 10.3.

Ending the rest on the corrected estimate starts the crash on day 7.40, with the estimate at the limit, its uncorrected model at 0.066 mg/L and the truth at 0.130 mg/L. By then the truth is 1.96 times the model, because this strain forms 15 % more VDK and reduces it 20 % more slowly than the estimator's textbook yeast, and the correction, built from results a day old, has reached a factor of 1.52. The yeast keep working into the crash, and the beer is crashed on day 10.37 at 0.089 mg/L. Without the correction ($\lambda = 0$) the crash starts on day 6.70 at 0.180 mg/L and the beer finishes at 0.123 mg/L, over the limit. A gain of 0.25 starts the crash on day 7.19 and finishes at 0.098 mg/L, and a gain of 1 on day 7.57 at 0.083. A limit of 0.08 mg/L on the estimate starts it on day 7.78 at 0.109 and finishes at 0.075, and a timed rest of 48 h starts it on day 7.97 and finishes at 0.069. The laboratory's turnaround is read when the batch starts, so it is set before the run. With a 12 h turnaround the day-8 result reports 0.101 mg/L, just above the limit through the analysis noise, and the crash waits for the day-9 result on day 9.50. The laboratory is exact but a day old and the estimator is current but only as good as its constants, so the useful arrangement gives the model the timing and the laboratory the level, as inferential schemes do [5, 6]. The spray dryer of Volume 5, Chapter 10 corrects a soft

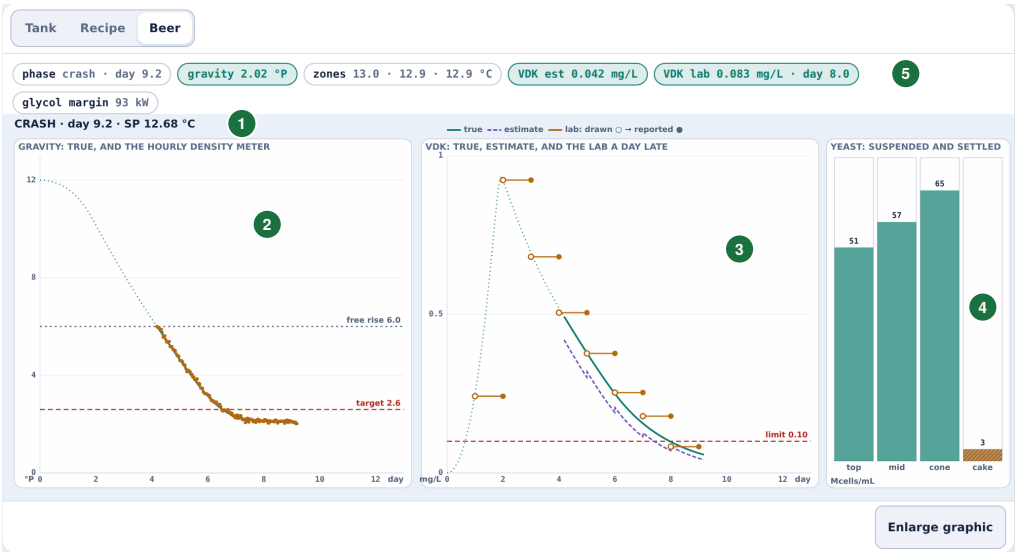


Figure 10.3 The diacetyl rest of case 6 at $t = 120$ h (5d 00:00), on batch day 9.2, in the Beer view with no control changed, cropped to the plant view. Callout 1 is the status line: the supervisor has started the crash on the laboratory's result for the day-8 sample, 0.083 mg/L, and the setpoint has ramped to 12.68 °C, with the probes at 13.0, 12.9 and 12.9 °C in the badge row above. Callout 2 is the gravity panel, with the history as a dotted line and the true gravity and the hourly density readings falling from the 6.0 °P free-rise line to 2.02 °P, below the 2.6 °P target. Callout 3 is the VDK panel, with the true value, the estimate and eight laboratory results, each drawn from its sample day to its report day, from 0.242 mg/L for day 1 and the peak of 0.923 for day 2 down to 0.083, against the 0.10 mg/L limit. Callout 4 is the yeast bars, 51, 57 and 65 Mcells/mL in suspension in the top, middle and cone and 3 in the cake, and callout 5 the badges of the estimate, 0.042 mg/L, and the latest laboratory result.

sensor's bias from a delayed laboratory in the same way, and Chapter 8 corrects its culture model from a daily count.

🕮 **Pause.** Try Exercise 10.3 before reading on.

10.5.6 A Crash on a Short Glycol Plant

Case 7 starts its crash on day 7.95 while the other tanks draw 478 kW of the plant's 500, easing over the next days. With the coordinator on, this tank may draw between 1.8 and 85.3 kW during its crash, the ramp waits, and the supervisor marks the crash late at its first control step. The lumped projection is pessimistic while the neighbors' load is easing, beyond day 19 at the start, day 14.5 after one day and day 12.4 near day 11, and the beer is crashed on day 12.71 with the plant never over capacity and the supply at −4 °C. With the coordinator off the valves take what they pass, the plant runs up to 70.9 kW over capacity for 41.5 h, the supply warms to −1.87 °C for every tank on it, and on day 8.06 the crash itself is held, because glycol less than a degree colder than −1 °C cannot reach it. That beer is crashed on day 11.11. A shared utility turns the tank's schedule into a plant-wide question, and the coordinated answer is a late batch announced from the first hour rather than a warmer supply for every tank. The refrigeration pack of Volume 8, Chapter 12 shows the same coupling through a shared suction pressure, where thermostats that exchange no signal still act on one another. Load management of this kind was one of the three levers by which a brewery's refrigeration plant was found to save fuel [8].

Table 10.6 The Twelve-Day Batch, Shipped, Suggested and With One Lever Returned to Its Shipped Setting

Configuration	Rise (day)	Crash (day)	Done (day)	D (mg/L)	Dev. (°C)	Over (kW)
as shipped	6.00	9.79	12.57	0.058	2.73	0
suggested	4.21	7.42	10.47	0.087	0.52	0
rest of 48 h	4.21	8.00	10.98	0.069	0.52	0
hysteresis valves	4.25	7.41	10.53	0.083	0.52	39.4
one mid probe	4.21	7.42	10.28	0.088	1.45	0
free rise on day 6	6.00	8.51	11.49	0.088	1.59	0
rest on the laboratory	4.21	9.00	11.98	0.044	0.52	0
coordinator off	4.21	7.42	10.39	0.089	0.52	10.2

10.5.7 A Stuck Fermentation

Case 8 pitches yeast of 15 % viability. On day 3.00 the density meter reads 11.16 °P and falls 0.51 °P a day, where the healthy batch of case 3 is at 8.13 °P, and projected at 14 °C the target is out of reach until day 17.9. The supervisor declares the fermentation stuck, warms the beer to 14 °C and holds it there for a re-pitch decision, and the true gravity is still 4.72 °P on day 12, so a crash on schedule would have shipped unfinished beer. The rule knows nothing of the viability. The pitch is read at knock-out, so a viability is set on case 3 before the run. A 30 % viability is declared stuck on day 7.98, on a projected finish of day 12.0 at the edge of the plan, while at 45 % the batch is crashed on day 11.36 at 2.11 °P but at 0.108 mg/L, because the estimator still assumes a healthy pitch and ends the rest too soon, and at 75 % on day 10.66 at 0.099 mg/L. Honest infeasibility is a control function here. In wine, where stuck and sluggish fermentations have been reviewed at length, their causes are many and hard to diagnose [7], and a rule on the measured trend flags the symptom without having to know the cause.

10.5.8 The Twelve-Day Challenge

The challenge runs a whole batch from knock-out, with a neighbor tank starting its crash on day 7 and drawing 150 kW that eases over about a day. The target is a crash to −1 °C completed by day 12 at no more than 2.6 °P and below 0.10 mg/L of true VDK, with every zone within 1 °C of the recipe throughout and the plant never over capacity. As shipped, with one mid-height probe, hysteresis valves, a free rise on day 6, a 48 h rest and no coordinator, the free rise comes too late for the fermentation's own heat, and the beer reaches only 12.77 °C while the recipe asks for 14. The rest starts on day 7.79, the crash on day 9.79, and the unprobed cone runs 2.73 °C off the recipe on day 10.55. The beer is crashed on day 12.57 at 2.10 °P and 0.058 mg/L, too late and outside the band. The suggested configuration of three probes on time-proportioned valves, the free rise at 6.0 °P, the rest ended on the corrected estimate and the coordinator on completes the crash on day 10.47 at 2.10 °P and 0.087 mg/L, every zone within 0.52 °C and the plant never over capacity, which leaves margins of 1.53 days, 0.50 °P, 0.013 mg/L and 0.48 °C on the target. The challenge is scored on the tank as given, so a change to a model parameter or a disturbance voids the attempt, and a change to a valve mode, a probe arrangement, a trigger or the coordinator does not. Table 10.6 returns one lever at a time. A timed rest and a rest ended on the laboratory both pass, the latter on day 11.98 with half a day to spare. Hysteresis valves with three probes hold every zone within 0.52 °C on 103 valve switches against 2825, but during the rest their full openings meet the neighbor's crash and hold the plant 39.4 kW over capacity for 1.2 h. One probe breaks the band at 1.45 °C, a free rise on day 6 at 1.59 °C with the beer reaching only 12.57 °C, and without the coordinator the neighbor's crash and this one hold the plant 10.2 kW over capacity for 3.1 h. Every failing lever is a trigger, a measurement, the valve law or the shared utility, and none is a gain. Figure 10.4 draws the table as bars.

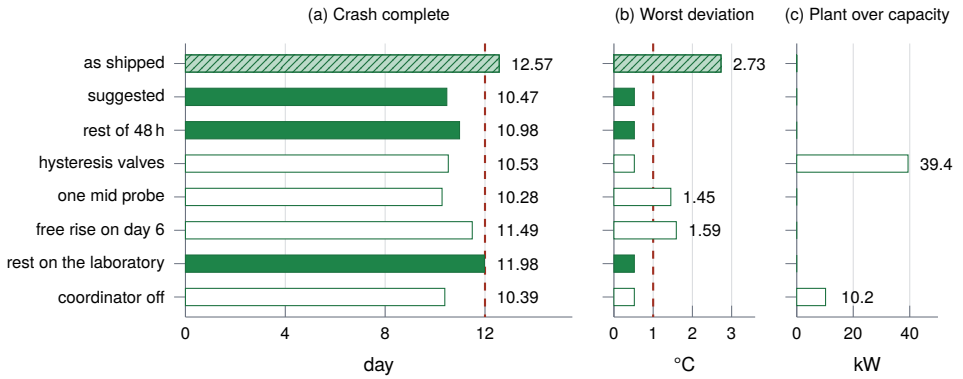


Figure 10.4 The twelve-day batch of Table 10.6 drawn as bars, with the challenge's targets dashed: the crash complete by day 12 and every zone within 1 °C of the recipe. The hatched bar is the batch as shipped, solid bars pass and white bars fail. Notice that each white bar fails on one panel only, on the deviation or on the plant, and that the passing ones differ only in when the rest ends.

10.6 Technology in Practice

The simulated tank carries three temperature loops, a density meter, a daily laboratory sample and a shared glycol supply. Figure 10.5 draws them with the devices a cellar would install.

What Measures It

Each zone has a thermowell with a resistance thermometer, TT-101 to TT-103, and the supply carries a fourth, TT-501. The chapter models each probe as a first-order lag of 10 min with a little colored noise, which is what a well and its element add to the true beer temperature. The gravity comes from DT-201 on the sample line of the middle zone, read once an hour, and the diacetyl only from the laboratory, a day after the sample was drawn. No device in the model measures the cone's gravity or any zone's diacetyl.

> **TECHNOLOGY Sensor: Resistance Thermometer in a Thermowell**
>
> A zone temperature is read by a resistance temperature detector, a fine wire of pure metal, usually platinum, whose resistance rises with temperature. It is less sensitive than a thermistor but far more linear, which is why platinum is the common industrial choice [18]. The element sits in a thermowell, a pressure-tight sheath that lets it be replaced without opening the vessel. The well's mass adds a first-order lag, and an element not seated against the bottom of the well adds a second lag through the air gap, which can push a feedback loop toward oscillation [18]. On a trend that fault shows as a probe that trails its zone and a limit cycle that grows. The chapter's 10 min first-order lag stands for the well.

What Moves It

Three on/off valves admit glycol to the three jackets. A solenoid valve has two positions and a resting position set by its spring, so the controller's only freedom is when to switch it. The time-proportioned output of Section 10.4 turns a duty into an on-time, and its minimum on and off times of 2 min keep it from asking for shorter pulses. The chapter's valves open and shut instantly, which no real valve does.

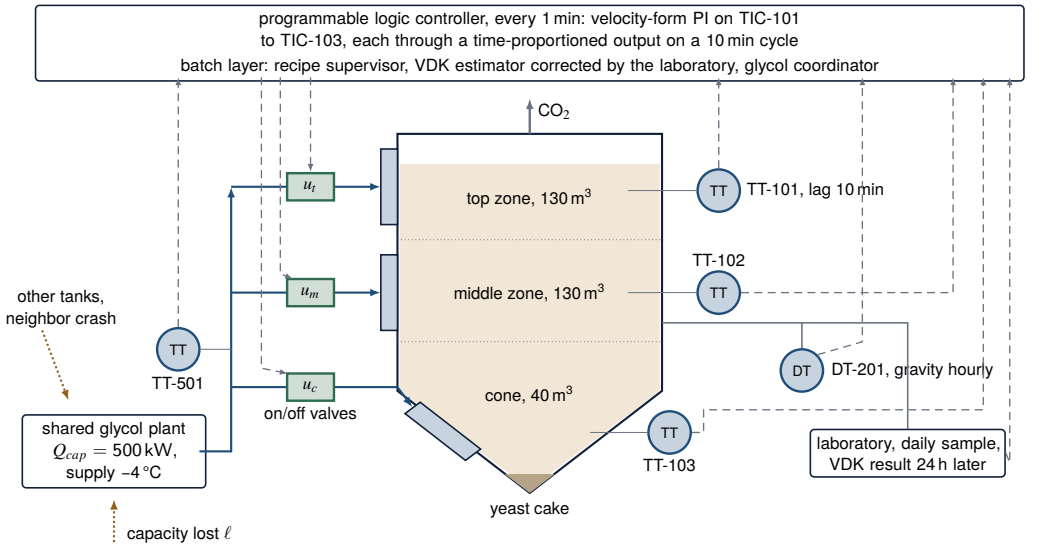


Figure 10.5 The chapter's loops with the devices a cellar would install. Thick blue lines carry glycol, dashed gray lines the signals and dotted amber arrows the disturbances. Notice that the three valves are on/off solenoid valves fed from one shared plant, and that only temperature is measured in every zone: the gravity is read at the middle zone once an hour and the diacetyl only by the laboratory, a day late.

TECHNOLOGY Final Element: On/Off Solenoid Valve

A jacket valve that is only open or shut is commonly a two-way solenoid valve. An energized coil pulls a ferrous armature that lifts a small plug off its seat, and a return spring sets the resting position, normally closed or normally open, which is where the valve goes when the coil is de-energized [18]. Process pressure should act with the spring, or it may keep the valve from returning to rest, and fast action needs a solenoid force well above the spring force [18]. A valve that fails to shut shows on a trend as a zone that keeps cooling below its setpoint while its duty sits at zero, the stuck-valve fault of the chapter's variables table. The simulated valves have no stroke time.

Where the Controller Runs

The zone loops would run on a programmable logic controller or in a distributed control system. A programmable logic controller holds instructions for input and output, logic, timing, counting and three-mode PID control, and in a distributed control system it serves as a local controller within a supervisory scheme [19]. Supervisory and lower loops run continuously with cycle times from milliseconds to minutes [19]. The chapter's loops execute once a minute on a ten-minute valve cycle, slow for a controller and fast for a tank of 1.23 GJ/K. In a function block program the PID block feeds a time-proportioning block that switches a discrete output, the arrangement used where a final element can only be fully on or fully off [18]. The recipe supervisor, the estimator and the glycol coordinator belong one layer above the loops, in the batch logic that moves the tank from phase to phase.

What Is Modernizing It

Two lines of work would move the measurement boundary this chapter draws. The first is wider monitoring. Automated monitoring of fermentation has long meant little more than temperature, and electronic noses, electronic tongues and other sensors are being studied for the variables that decide a batch, so far mostly in laboratories [20]. The second is connected, low-cost instrumentation.

Large breweries adopted sensors and digitization before smaller ones, and connected devices are now spreading to mid-sized and craft breweries, with real-time monitoring feeding data-driven decision support, while data ownership and standardization remain open [21]. Both would feed model-based state estimation, which has already tracked batch beverage fermentations with an imperfect model [6]. Each shortens the day that the simulated estimator must bridge between a sample and its result.

> **IN PRACTICE** **A Differential Gap, and the Wear of Every Stroke**
>
> On/off control with a single setpoint is possible but usually not practical, because the final element then cycles too often. A working on/off loop has two setpoints with a gap between them, a differential gap, and the process variable cycles between the upper and lower setpoints without ever settling at the setpoint [18]. Each cycle is mechanical wear, and frequent cycling of the final element may contribute to its premature failure [18]. The gap therefore trades the swing of the temperature against the number of strokes, which the chapter's scorecard counts as valve switches.

10.7 Limitations

The kinetic constants are illustrative and not fitted to a strain, the VDK is one lumped state that cannot reappear from its precursor after the crash, and the heat of fermentation is a declared constant. The tank is three well-mixed zones, so it shows stratification between zones but not a gradient within one, and the density maximum is water's.

The glycol plant is a capacity with an algebraic supply temperature, the other tanks are a load profile that does not respond to it, and the valves have no stroke time. The density meter fails only through its bias disturbance, noise and bubble spikes, and the laboratory only through its noise. The estimator shares the plant's structure and differs only in its constants and the viability, where a real estimator would also differ in structure. The supervisor's crash projection cools the tank as one lump at the present allotment, so it is pessimistic while the other tanks' load is easing, beyond day 19 at the start of case 7 for a crash completed on day 12.71, and optimistic on an uncontended plant, where it first projects day 9.7 for a crash completed on day 10.39 because it ignores the ramp and the slowest zone. Mid-batch starts replay a history held exactly on the recipe, and every size, capacity and limit is a teaching value.

> **OPEN CHALLENGE** **Seeing More of a Fermentation Than Its Temperature**
>
> Automated monitoring of alcoholic fermentation is often limited to a few parameters, mainly temperature. A review of the sensors proposed for beer, wine and spirits found that the solutions are mostly isolated laboratory studies yet to be tested in industrial plants, that a public database of recorded sensor data is needed but missing, and that the machine learning in these studies generally falls short of the established standards of data science [20]. Progress would look like an in-line measurement of gravity or diacetyl proven in a production cellar against the laboratory, with its data open to others.

10.8 Conclusion

This chapter studied a lager fermenter, a batch reactor whose heat arrives on the yeast's schedule, whose valves are open or shut, whose mixing comes and goes with the CO_2 and whose endpoint is a day-old laboratory result, on a simulator that logs every zone's truth beside every instrument. With the valves shut the beer runs away to 21.33 °C, and on the same valves a thermostat holds 10.66 to 11.27 °C where a time-proportioned PI holds 10.97 to 11.04 °C, although a 14.7 % open share removes 1.8 times its linear share of heat. One mid-height probe lets the quiet crash spread the zones by 2.69 °C. A rest ended on the laboratory starts the crash 1.04 days after the beer was ready, the corrected estimator 1.60 days earlier, and without the correction the beer ships at 0.123 mg/L. On the short plant the coordinated crash is announced late and completes on day 12.71 without the 70.9 kW overload of the uncoordinated crash, and a fermentation at 15 % viability is declared stuck on day 3.00 rather than crashed off target. The graded batch improves from day 12.57, its cone 2.73 °C off

the recipe, to day 10.47 within 0.52 °C on unchanged gains. The model is lumped in its zones and illustrative in its constants. A fuller model would add a diacetyl precursor, a glycol inventory that warms in an outage, responsive neighboring tanks and valve stroke times. The tank makes visible that a lager batch is decided by what triggers each phase, what ends it and what its measurements can see.

KEY POINTS

- A lager batch is decided by what triggers each phase, what ends it and what its measurements can see, not by the tuning of its temperature loops.
- On the same valves a thermostat holds 10.66 to 11.27 °C where a time-proportioned PI holds 10.97 to 11.04 °C.
- One mid-height probe lets the quiet crash spread the zones by 2.69 °C.
- A rest ended on the laboratory starts the crash 1.04 days after the beer was ready, and without the correction the estimator ships the beer at 0.123 mg/L.
- The graded batch improves from day 12.57 to day 10.47 on unchanged gains.

Where else this applies. Volume 8, Chapter 2 removes the heat of an exothermic batch through a jacket, Volume 8, Chapter 12 runs thermostats on a shared refrigeration plant, and Volume 5, Chapter 10 corrects a soft sensor from a late laboratory.

References

[1] D. A. Gee and W. F. Ramirez, "Optimal temperature control for batch beer fermentation," *Biotechnology and Bioengineering*, vol. 31, no. 3, pp. 224–234, 1988.

[2] K. Krogerus and B. R. Gibson, "125th anniversary review: Diacetyl and its control during brewery fermentation," *Journal of the Institute of Brewing*, 2013.

[3] D. A. Gee and W. F. Ramirez, "Optimal temperature control of jacket-cooled fermentation reactors," *Optimal Control Applications and Methods*, vol. 12, no. 1, pp. 49–62, 1991.

[4] K. J. Åström and T. Hägglund, "Automatic tuning of simple regulators with specifications on phase and amplitude margins," *Automatica*, vol. 20, no. 5, pp. 645–651, 1984.

[5] M. T. Tham, G. A. Montague, A. J. Morris, and P. A. Lant, "Soft-sensors for process estimation and inferential control," *Journal of Process Control*, vol. 1, no. 1, pp. 3–14, 1991.

[6] D. A. Gee and W. F. Ramirez, "On-line state estimation and parameter identification for batch fermentation," *Biotechnology Progress*, vol. 12, no. 1, pp. 132–140, 1996.

[7] L. F. Bisson, "Stuck and sluggish fermentations," *American Journal of Enology and Viticulture*, vol. 50, no. 1, pp. 107–119, 1999.

[8] P. M. Petersen and B. Qvale, "Control and optimization of the refrigeration plant in a brewery," in *Proceedings of the 24th Intersociety Energy Conversion Engineering Conference*, 1989, pp. 1807–1812.

[9] D. E. Seborg, T. F. Edgar, D. A. Mellichamp, and F. J. Doyle, III, *Process Dynamics and Control*, 4th ed. Hoboken, NJ: John Wiley & Sons, 2016.

[10] B. de Andrés-Toro, J. M. Girón-Sierra, J. A. López-Orozco, C. Fernández-Conde, J. M. Peinado, and F. García-Ochoa, "A kinetic model for beer production under industrial operational conditions," *Mathematics and Computers in Simulation*, vol. 48, no. 1, pp. 65–74, 1998.

[11] L. Chew, "Mechanical refrigeration as applied to the brewing industry," *Journal of the Institute of Brewing*, vol. 18, no. 4, pp. 274–312, 1912.

[12] L. Nathan, "Improvements in the fermentation and maturation of beers," *Journal of the Institute of Brewing*, vol. 36, no. 6, pp. 538–544, 1930.

[13] H. Meironke, J. A. Szymczyk, and A. Leder, "Thermofluiddynamics inside industrial cylindroconical fermenting tank," in *36th AIAA Fluid Dynamics Conference and Exhibit*, 2006.

[14] M. Burger, P. R. Glenister, and K. Becker, "Diacetyl studies. II. formation and prevention of diacetyl in beer," *Proceedings. Annual Meeting - American Society of Brewing Chemists*, vol. 15, no. 1, pp. 110–115, 1957.

[15] T. Wainwright, "Diacetyl-a review: Part I-analytical and biochemical considerations: Part II-brewing experience," *Journal of the Institute of Brewing*, vol. 79, no. 6, pp. 451–470, 1973.

[16] A. J. Cutaia, A.-J. Reid, and R. A. Speers, "Examination of the relationships between original, real and apparent extracts, and alcohol in pilot plant and commercially produced beers," *Journal of the Institute of Brewing*, vol. 115, no. 4, pp. 318–327, 2009.

[17] C. M. Şutea, E. Mudura, C. R. Pop, L. C. Salanţă, A. C. Fărcaş, P. C. Balaş, E. Gal, E.-I. Geană, H. Zhao, and T. E. Coldea, "Beer aroma compounds: Key odorants, off-flavour compounds and improvement proposals," *Foods*, vol. 14, no. 24, p. 4287, 2025.

[18] T. R. Kuphaldt, *Lessons in Industrial Instrumentation*, version 2.33 ed. Open textbook, Creative Commons Attribution 4.0 license, ibiblio.org/kuphaldt/socratic/sinst, 2024.

[19] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule, and M. Thompson, "Guide to operational technology (OT) security," National Institute of Standards and Technology, Tech. Rep. NIST Special Publication 800-82r3, 2023.

[20] T. Horváth, S. Kun, L. Sipos, and D. T. Pham, "Automated monitoring of alcoholic fermentation: Trends and challenges," *Journal of Food Science and Technology*, vol. 63, no. 2, pp. 224–237, 2026.

[21] S. Violino, S. Figorilli, C. Costa, and F. Pallottino, "Internet of beer: A review on smart technologies from mash to pint," *Foods*, vol. 9, no. 7, p. 950, 2020.

10.9 Exercises

10.1 Concept check · 2 min

At the day-two peak of case 3 the top valve is open 14.7 % of the time, and its jacket removes 15.2 kW. That is 1.8 times the open share times the 56.7 kW that a valve held open removes at steady state. Which explanation is right?

(a) Each pulse fills the jacket with supply glycol that goes on cooling after the valve shuts

(b) The probe lag makes the loop ask for a larger duty than the zone needs at the peak

(c) The carry of the time-proportioned output keeps the valve open longer than its duty

(d) The glycol supply runs colder at the peak than in the steady test of the open valve

10.2 Concept check · 1 min

Case 5 crashes a quiet tank, with the CO_2 gone, on one mid-height probe, and TIC-102 sends the same duty to all three valves. Which zone ends furthest from the recipe, and in which direction?

(a) The top zone, too warm, because warm beer rises once the gas stops

(b) The cone, too cold, because it has the most open jacket per cubic meter

(c) The middle zone, too warm, because its probe lags the beer by ten minutes

(d) The cone, too warm, because the settled yeast insulate it from its jacket

10.3 Concept check · 2 min

In case 6 the rest is ended on the estimate twice, once with the laboratory bias update gain λ at 0.5 and once at 0. What does the run at $\lambda = 0$ show, and why?

(a) The crash starts later, because the uncorrected model forms more diacetyl than the strain

(b) The crash starts on the same day, because the estimate takes its timing from the model alone

(c) The crash starts earlier and the beer finishes over the limit, because the model's yeast clear diacetyl faster

(d) The crash starts earlier but the beer finishes inside the limit, because the yeast keep working in the crash

10.4 Theory and calculation · 6 min

A cylinder jacket has $UA = 6.5\,\mathrm{kW/K}$ and a glycol holdup $M_j = 600\,\mathrm{kg}$ with $c_g = 3.6\,\mathrm{kJ/(kg\,K)}$, and an open valve passes $F_j = 2.5\,\mathrm{kg/s}$. The cone jacket has $UA = 4.0\,\mathrm{kW/K}$.

(a) From Equation (10.7) at steady state with the valve open, find the effective conductance UA_{eff} between the beer and the supply for a cylinder jacket and for the cone. *Checkpoint:* 3.77 kW/K and 2.77 kW/K.

(b) Find the time constant of the cylinder holdup with the valve open and with it shut. *Checkpoint:* 2.3 min open and 5.5 min shut.

(c) Find the heat the three open jackets remove with the beer at 11 °C and the supply at −4 °C, and compare it with the 32.8 kW peak of the 12 °P batch held at 11 °C. *Checkpoint:* 155 kW, 4.7 times the peak.

10.5 Theory and calculation · 5 min

The crash cools 300 m^3 of beer with $\rho c_p = 4100\,\mathrm{kJ/(m^3\,K)}$ from the rest at 14 °C to −1 °C on a ramp of 0.3 °C/h.

(a) Find the heat capacity of the tank and the heat the crash removes. *Checkpoint:* 1.23 GJ/K and 18.45 GJ.

(b) Find the heat flow that the ramp asks for. *Checkpoint:* 102.5 kW.

(c) The three open jackets give $\sum UA_{eff} = 10.32\,\mathrm{kW/K}$ against a supply at −4 °C. Find the beer temperature below which they cannot keep up with the ramp. *Checkpoint:* About 5.9 °C.

10.6 Theory and calculation · 7 min

The estimator's VDK reduction uses $a_D(T)$ of Equation (10.4) with $E_D = 110\,\mathrm{kJ/mol}$ and $T_r = 11\,^\circ\mathrm{C}$, and its level is corrected by Equation (10.14).

(a) Find a_D at the 14 °C rest and at the −1 °C crash temperature. *Checkpoint:* 1.63 and 0.128.

(b) Suppose the truth stays at a constant 1.96 times the model, so the log error is $e = \ln 1.96$. With $\lambda = 0.5$ and $b = 0$ at the start, find the correction factor e^b after one, two and three results, and the number of results after which it is within 10 % of 1.96. *Checkpoint:* 1.40, 1.66 and 1.80. Three results.

(c) Repeat the factor after three results for $\lambda = 0.25$, and say what a smaller gain buys. *Checkpoint:* 1.48.

10.7 App challenge · 8 min

Load scenario 2, On/off with hysteresis, and change no setting. Run to 72 h at the top speed with the default seed and export the CSV file. Then do the same with scenario 3, Time-proportioned PI.

(a) For each run find the lowest and the highest zone temperature in the columns `T_t_C`, `T_m_C` and `T_c_C` over the rows from 24 h to 72 h, within 0.01 °C. *Checkpoint:* 10.66 to 11.27 °C, and 10.97 to 11.04 °C.

(b) Read `metric_switches` in the rows at 24 and 72 h and find the valve switches between them in each run. *Checkpoint:* 29 and 802.

(c) Find how many times narrower the PI swing is, and say why the hysteresis swing reaches below its lower switch point of 10.75 °C. *Checkpoint:* 8.7 times.

10.8 App challenge · 9 min

Load scenario 6, The diacetyl rest. Use the default seed, run at the top speed, and let each run go on until it stops itself at the end of the batch. For each run read the Crashed (finished) and VDK tiles of the scorecard, and in the exported CSV file find the first row whose `phase` is `crash`, where `batch_age_h` divided by 24 gives the day the crash starts and `metric_vdk_at_crash_mgL` the true VDK then.

(a) Run the case as shipped, with the rest ended on the laboratory. *Checkpoint:* Crash on day 9.00 at 0.062 mg/L. Tiles: day 12.0, 0.044 mg/L.

(b) Press Reset, set Rest ends on to Estimate, and run again. *Checkpoint:* Crash on day 7.40 at 0.130 mg/L. Tiles: day 10.4, 0.089 mg/L.

(c) Press Reset, keep the Estimate, set the Lab bias update gain to 0, and run again. *Checkpoint:* Crash on day 6.70 at 0.180 mg/L. Tiles: day 9.6, 0.123 mg/L.

Accept a day within 0.01 and a VDK within 0.002 mg/L. The true VDK falls below 0.10 mg/L on day 7.96. Say what each ending costs.

10.9 App challenge · 8 min

Load the challenge, A twelve-day lager, at the default seed and run at the top speed. Before Run set Jacket valves to Time-proportioned, Temperature probes to Three zones, Free rise starts to At a gravity, Rest ends on to Estimate and the Glycol coordinator On. Run until the batch stops itself.

(a) Read the Crashed (finished), VDK, Worst recipe deviation and Glycol over capacity tiles and the Challenge target tile. *Checkpoint:* Day 10.5, 0.087 mg/L, 0.52 °C (top), 0.0 kW for 0.0 h, met.

(b) Press Reset, set Jacket valves to Hysteresis and change nothing else. Run again and read the same tiles. *Checkpoint:* Day 10.5, 0.083 mg/L, 0.52 °C (top), 39.4 kW for 1.2 h, not met.

Explain why the hysteresis valves fail although they hold every zone as closely as the time-proportioned valves.

10.10 Application area · 4 min

In case 7 the other tanks draw 478 kW of the glycol plant's 500 kW at the start of this tank's crash, and the coordinator keeps a reserve of 20 kW.

(a) Find this tank's allotment at the start from the coordinator's rule. *Checkpoint:* About 2 kW.

(b) The allotment rises to at most 85.3 kW during the crash. Find the shortest time in which that draw could remove the 18.45 GJ of a crash from 14 to −1 °C. *Checkpoint:* About 60 h.

(c) Give the engineering case for announcing a late crash rather than taking the glycol.

10.11 Application area · 3 min

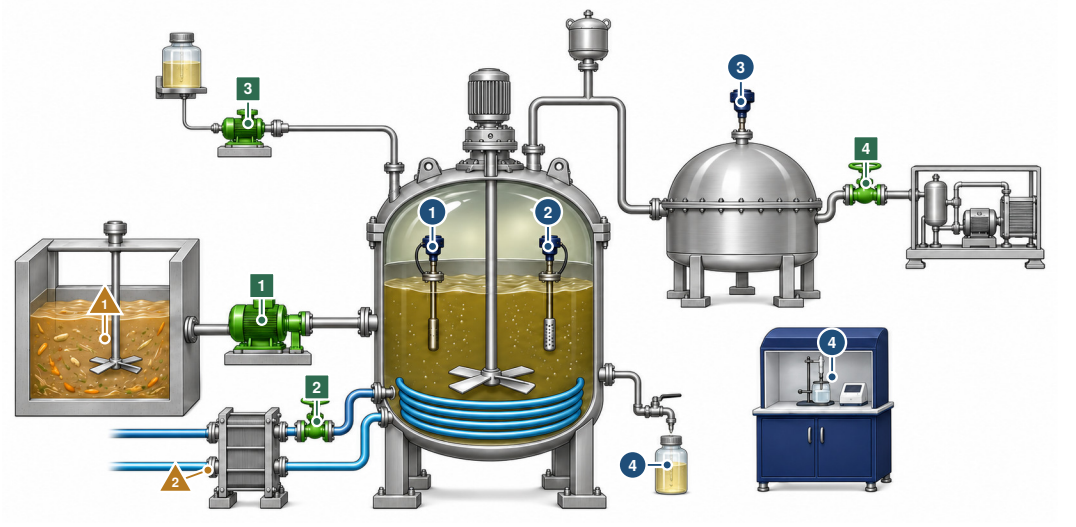
A brewer reads that ending the rest on the laboratory cost 1.04 days of tank time, and proposes to end the rests of a real cellar on a model-based estimate corrected by the laboratory, as case 6 does. Name two limitations stated in the chapter that bear on this proposal, and say what each could do in a real tank.

10.12 Review the agent · 5 min

An AI assistant is asked why the challenge fails as shipped and replies as follows. "The cone runs 2.73 °C off the recipe, and the cone loop has the weakest tuning, $K_c = 60\,\%$ per °C against 150 for the other two. Raise the cone gain to 150 and shorten all three integral times from 8 to 2 h, and the cone will stay inside the 1 °C band." Find the flaw in the proposal and say how to check it.

CHAPTER 11

Feeding an Anaerobic Digester Toward Its Fold on a Daily Titration



The controller moves 1 feed pump 2 heating valve 3 bicarbonate dosing 4 gas draw to the upgrader

The controller reads 1 tank temperature 2 pH probe 3 gas holder level 4 daily acids test of a sample

What disturbs it 1 change in the feed blend 2 heater fault

Notice that the controller reads the tank's temperature and pH at once, but its acids only through a sample tested once a day, hours after it was drawn. Illustration, not to scale.

IN PLAIN TERMS **Feeding a Digester Without Souring It**

The main idea. A digester is a warm, closed tank where microbes turn manure and food waste into a gas that burns. Fast microbes make acids within hours, and slow ones eat the acids over days to give off the gas.

The challenge. More feed makes more gas for a day or two, then the acids pile up until the slow microbes stop and the tank turns sour. The test that shows the acids comes from a laboratory once a day, hours after the sample was taken.

Why it matters. A sour tank takes weeks to bring back, so the feed must be held back before the test warns.

THREE WAYS TO READ THIS CHAPTER

Explore	11 min	the first page, the fact on farm digesters (Section 11.1), the two history segments (Section 11.2), the try-it segments and the fact of Section 11.6, and the key points
Study	45 min	the whole chapter with its three pause points and the twelve exercises of Section 11.11
Practice	15 min	the first page, the measured cases (Section 11.6), Technology in Practice (Section 11.8), Limitations (Section 11.9) and the open challenge

The control question More feed raises a digester's biogas for a day before the acids consume its buffer and the tank sours for weeks. How close to that fold can a feed loop run when the only timely warning of the acids is a titration drawn once a day and reported six hours later?

Learning objectives

- Locate the fold of the Haldane growth law and explain why the retention, not the feed's strength, sets it.
- Explain why the biogas rises past the fold before the titration warns, and why bicarbonate buys time without removing acid.
- Predict how far a heater fault moves the fold from the heat balance and the temperature factors.
- Relate the gain of a loading loop to the age of the titration it acts on, and recognize a loop that rings on a sampled measurement.
- Weigh an economic supervisor's revenue against the margin it spends near the fold.

Prerequisites Sections 1.4.3 (Haldane growth and the fold), 1.4.5 (sparse and late samples), 1.4.6 (one-sided actuators and stores), 1.6.2 (velocity form) and 1.6.4 (scenarios). Section 1.6.3 is needed only for the predictive feed controller.

The app https://apmonitor.com/apps/anaerobic-digester/

See also Volume 7, Chapter 2 for the ignition and extinction folds of an exothermic reactor and Volume 7, Chapter 4 for a loop held at the peak of its plant characteristic. Volume 5, Chapter 3 for a tuning limited by a slow analyzer and Volume 5, Chapter 4 for a cook graded by late analyzer and laboratory results.

11.1 Introduction

A 4000 m^3 digester fed 160 m^3/d of dairy manure with 12 % food waste makes 4674 Nm^3/d of biogas at 60.0 % methane in the simulator of this chapter. On the second day of its overload case the feed is raised to 400 m^3/d. The biogas climbs to a peak of 9504 Nm^3/d on day 2.63, twice where it started, and an operator paid by the cubic meter would read that as success. The titration that shows the acids past the limit is drawn at 08:00 on day 3 and reported on day 3.58, almost a day after the gas has turned down, and the digester sours on day 4.51. Bringing a soured tank back takes 45.3 days at half feed with the pH held by bicarbonate, and without the bicarbonate it does not come back at all.

The behavior has a simple cause that a block diagram hides. Acidogens turn the feed into volatile fatty acids within hours, and methanogens turn the acids into methane over days while being slowed by those same acids. Above a critical feed no methanogen population can grow as fast as the liquid carries it out, and the steady states on either side of that point form a fold. The misconception the simulator corrects is that more gas means a healthy digester. Near the fold more gas is the methanogens burning a surplus they cannot keep up with, plus the carbon dioxide that the accumulating acids drive out of the bicarbonate.

The class of problem is wider than biogas. Growth on an inhibitory substrate can make a continuous culture unstable, a result published in 1968 with wastes containing phenols, ammonia and volatile acids named as the cases where it matters [1]. The controlled variable comes from a laboratory on a slow cycle, as in any plant run on a delayed assay, such as the kraft digester of Volume 5, Chapter 4, and the economics push the operator toward the edge, the setting of extremum seeking and best-operating-point studies for digesters [2, 3]. A reader who has run a nitrifying basin behind organic removal (Volume 9, Chapter 10), a fermentation with an inhibitory product or a reactor near a washout limit has met this shape.

DID YOU KNOW? Four Hundred Farm Digesters

As of June 2024 the United States had 400 anaerobic digestion systems on livestock manure, 343 of them taking dairy manure, and 105 of them complete-mix tanks like the one simulated here. Vehicle fuel was by then the most common use of their biogas [4].

However, the pieces sit in different literatures. The operating diagram of the two-step model is mathematical analysis [5, 6], the titration ratio is operating guidance [7], feed control is reviewed for energy production [8], and none of them lets a student watch a gas peak, a day-late titration and a fold move with temperature on one tank. This chapter works with one digester simulator in which the fold, the daily titration, bicarbonate dosing, a heater fault and the revenue from renewable natural gas act on one plant, and it measures each one's effect on that plant. The model is published, its methanogen constants are declared teaching values, and feed control for maximum biogas is established. The simulator has a 22-state model whose chemical oxygen demand and gas ledgers close to round-off, a souring map computed from the same growth law the engine integrates, and a strict split between what the controllers may read and the populations and true acids they may not. Its application grades ten cases, from a feed raised by hand to a sixty-day season.

Section 11.2 follows digester control from failing digesters to the titration loop. Section 11.3 gives the model: the two populations, the pH and the gas, the heat balance, the events and the measurements. Section 11.4 locates the fold, shows that the retention sets it and that temperature moves it. Section 11.5 describes the three control layers and the predictive feed controller. Section 11.6 measures the ten cases, the season and the predictive controller against the sampled loop. Section 11.7 sorts the evidence for the model into code verification, agreement with published results and the missing educational evaluation, Section 11.8 names the devices and the controller platform of a real plant, and Section 11.9 states the limitations. The exercises in Section 11.11 use the cases of Table 11.3.

11.2 Failing Digesters, Inhibition and the Titration

Digesters earned a reputation for instability long before they earned revenue, and the first dynamic models were built to explain it.

HISTORY A Model That Could Fail (1969 to 1971)

Andrews opened his 1969 model of the process by naming its major problem: a poor record of stability, as the reports of failing digesters showed. A model that predicts failure, he argued, needs an inhibition function that relates the volatile acids to the growth rate of the methane bacteria. He took the un-ionized acid as both the growth-limiting substrate and the inhibitor, which settled the conflict between inhibition by the acids and inhibition by pH, since the un-ionized share depends on both. His simulations reproduced what operators saw in the field at start-up, at failure and in recovery [9]. Andrews and Graef then added the interactions within and between the liquid, the gas and the biological phases of the digester [10].

Operators read their own indicators from a laboratory titration.

HISTORY The Operator's Ratio (1976 to 1987)

A federal operations manual for anaerobic sludge digestion, published in 1976, ranked the indicators of an upset in order of importance: the ratio of volatile acids to alkalinity first, then the production rates of methane and carbon dioxide, then pH. It found that the ratio can vary from less than 0.1 to about 0.35 without significant changes in digestion and called a rising ratio the first warning, while the pH, held up by the alkalinity, may not move until the digester is completely upset [11]. A survey of seventy observations in 1987 related failure to acetic acid above 800 mg/L or a propionic-to-acetic ratio above 1.4 [12], rules for a person reading a laboratory result.

Table 11.1 Variables by Role at the Design Point (160 m^3/d, 35 °C)

Variable	Role	Design	Range
r_{lab}	CV, daily titration	0.095	limit 0.4
T	CV, thermometer	35.0 °C	setpoint 30–40
pH	CV, probe (Auto in cases 3, 5)	7.39	setpoint 6.0–7.8
L	CV, holder level	50 %	setpoint 20–90
Q	MV, feed	160 m^3/d	0–400
Q_h	MV, heater duty	215 kW	0–600
u_{alk}	MV, bicarbonate	0	0–80 kmol/d
F_{up}	MV, upgrader feed	—	0–8640 Nm^3/d
f_{fw}	DV, food-waste share	12 %	slug 50, change 25
Q_{max}	DV, heater fault	600 kW	0 for 6 d
I	DV, inhibitor shock	0	0.35, decays
X_1, X_2	states, populations	0.889, 2.120	kg VSS/m^3, truth
S_1	state, substrate	0.374	kg COD/m^3, truth
S_2	state, acids	733 mg/L	12.20 mmol/L, truth
Z, C	states, alkalinity, carbon	154.4, 153.9	mmol/L
T	state, temperature	35.0 °C	—
H_m, H_c	states, holder gas	1000 Nm^3 total	methane, CO_2
q_{gas}	output, biogas	4674 Nm^3/d	60.0 % CH_4

Model-based control needed a model small enough to identify. Bernard and coauthors reduced the process to acidogenesis and methanogenesis with an alkalinity balance, identified it step by step from designed dynamical experiments, validated it on 70 days of a 1 m^3 fermenter, and made the alkalinity central to monitoring and control [13]. The International Water Association's task group published its structured model, ADM1, a year later, with disintegration, hydrolysis, acidogenesis, acetogenesis and two methanogenic steps, ion equilibria and gas transfer in 26 dynamic states [14]. Fifteen years of instrumentation, control and automation on high-rate digesters were then reviewed with the sensors, the model complexity and the strategies each one needs [15]. The two-step model's bifurcations were analyzed [5] and its operating diagram, the map of feed and dilution on which washout and bistability occur, was drawn in full [6]. Its constants identified on a full-scale 6000 m^3 reactor at a water resource recovery facility differ widely from the original values [16].

Feed control followed the models. Volatile fatty acids and total alkalinity have been regulated together [7], robust control has been applied to improve digester stability [17], biogas production has been intensified by extremum seeking with a variable gain [2], the operating conditions that maximize biogas in simple models have been characterized [3], and feed control for renewable energy has been reviewed [8]. A recent model predictive controller moves the feed against temperature disturbances without controlling the temperature [18]. Biogas upgrading and utilization have their own review literature [19], and in this simulator the declared fuel credits on the upgraded gas are what make the last cubic meter of feed valuable.

11.3 The Digester as Modeled

The simulated plant is one generic mesophilic digester with synthetic teaching values, drawn in Figure 11.1 in the arrangement of the application's Plant view. A pit blends dairy manure with food waste, and a pump feeds the blend continuously to a 4000 m^3 stirred tank held at 35 °C by a heater that burns the tank's own gas. The biogas collects in a 2000 Nm^3 membrane holder. A membrane upgrader draws from the holder and sells the methane as renewable natural gas (RNG), and the holder relieves to a flare when it is nearly full. A dosing pump can add sodium bicarbonate. Table 11.1 classifies the variables at the design point, a feed of 160 m^3/d for a hydraulic retention of 25 days and an organic loading rate (OLR) of 2.18 kg COD/m^3/d.

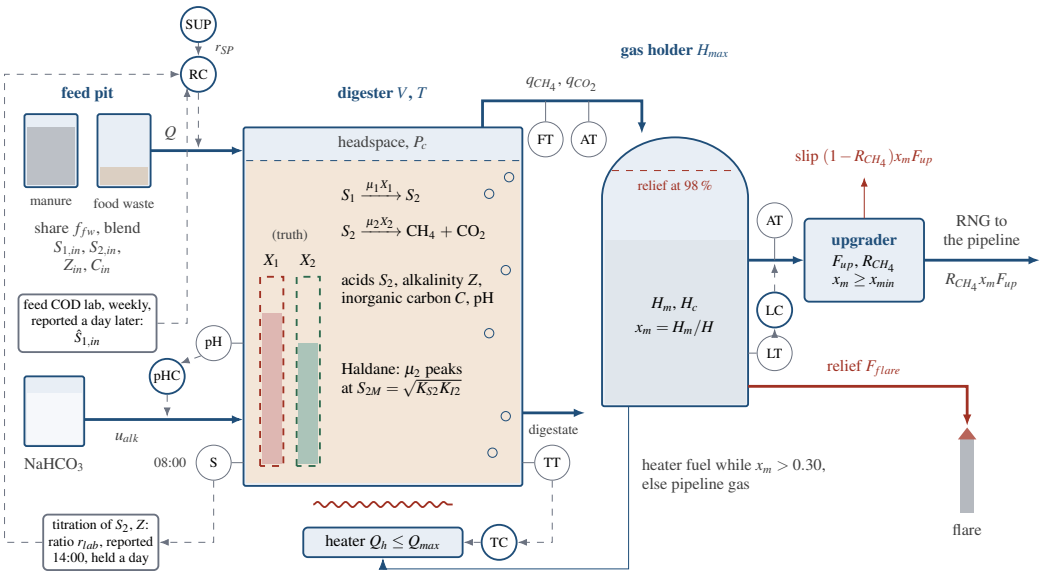


Figure 11.1 The digester as modeled. The feed pit blends manure and food waste at share f_{fw} and pumps Q into the tank of volume V, where acidogens X_1 and methanogens X_2 (teaching truth, dashed) turn the substrate S_1 into acids S_2 and the acids into methane and carbon dioxide, with alkalinity Z, inorganic carbon C and the headspace pressure P_c setting the pH. Bicarbonate enters at u_{alk}, the heater supplies Q_h, and the gas fills the holder (H_m, H_c) that feeds the upgrader at F_{up} and relieves to the flare above 98 % of H_{max}. A sample drawn at 08:00 is titrated and reported at 14:00, and the ratio controller (RC) acts on it under the supervisor (SUP). The temperature (TC), pH (pHC) and holder level (LC) loops act hourly. Not to scale.

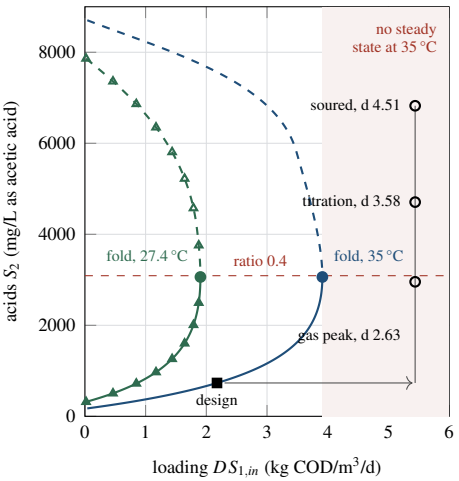


Figure 11.2 The souring map. Steady acids against loading from the application's map functions at the design alkalinity and dissolved CO_2, at 35 °C (blue, no markers) and at the heater fault's 27.4 °C (green, triangle markers). Solid branches are stable and dashed ones bound the basin the tank sours into. The thin dashed horizontal line is the ratio limit of 0.4 at the design alkalinity. The circles are case 2's true state at the gas peak, the first titration over 0.4 and the souring.

11.3.1 Two Populations

The liquid is the two-step model of Bernard and coauthors [13] written in units in which the chemical oxygen demand (COD) closes, with ADM1's first-order decay and pH inhibition [14] and a temperature factor per population. The populations grow on their substrates, decay, and leave with the liquid, as shown in Equation (11.1).

$$\begin{aligned}
\frac{\mathrm{d}X_1}{\mathrm{d}t} &= (\mu_1 - \alpha D - k_d)X_1 \\
\frac{\mathrm{d}X_2}{\mathrm{d}t} &= (\mu_2 - \alpha D - k_d)X_2 \\
\mu_1 &= \mu_{1,max}\,\theta_1^{T-35}\,I_{pH,1}\,\frac{S_1}{K_{S1}+S_1} \\
\mu_2 &= \mu_{2,max}\,\theta_2^{T-35}(1-I)\,I_{pH,2} \\
&\quad \times \frac{S_2}{K_{S2}+S_2+S_2^2/K_{I2}}
\end{aligned} \tag{11.1}$$

where $D = Q/V$ is the dilution rate in d^{-1}, $\alpha = 1$ says that the stirred tank keeps no biomass back (the fixed bed of the original work had 0.5), k_d is the decay rate, θ_1 and θ_2 are temperature factors per degree from 35 °C, and I is the activity an inhibitor shock removes. The acids S_2 are counted as acetic-acid equivalents in mmol/L. The Haldane term makes μ_2 peak at $S_{2M} = \sqrt{K_{S2}K_{I2}}$ and fall beyond it, the form of Equation (1.5). The pH factor is ADM1's lower inhibition, as shown in Equation (11.2).

$$I_{pH} = \begin{cases} 1 & \mathrm{pH} \geq \mathrm{pH}_{UL} \\ \exp\Big[-3\Big(\dfrac{\mathrm{pH}-\mathrm{pH}_{UL}}{\mathrm{pH}_{UL}-\mathrm{pH}_{LL}}\Big)^2\Big] & \mathrm{pH} < \mathrm{pH}_{UL} \end{cases} \tag{11.2}$$

with limits 7 and 6 for the methanogens (ADM1's acetate degraders) and 5.5 and 4 for the acidogens.

The substrate, the acids, the alkalinity and the inorganic carbon follow Equation (11.3).

$$\begin{aligned}
\frac{\mathrm{d}S_1}{\mathrm{d}t} &= D(S_{1,in} - S_1) - k_1\mu_1 X_1 + 1.42\,k_d(X_1 + X_2) \\
\frac{\mathrm{d}S_2}{\mathrm{d}t} &= D(S_{2,in} - S_2) + k_2\mu_1 X_1 - k_3\mu_2 X_2 \\
\frac{\mathrm{d}Z}{\mathrm{d}t} &= D(Z_{in} - Z) + \frac{u_{alk}}{V} \\
\frac{\mathrm{d}C}{\mathrm{d}t} &= D(C_{in} - C) - q_C + k_4\mu_1 X_1 + k_5\mu_2 X_2 + \frac{u_{alk}}{V}
\end{aligned} \tag{11.3}$$

where S_1 is the biodegradable COD in $\mathrm{kg/m^3}$, Z and C are in mmol/L, the dead biomass returns to the substrate at 1.42 kg COD per kg of volatile suspended solids, and bicarbonate adds one equivalent of alkalinity and one mole of carbon per mole, so it raises the pH and removes no acid. The methane flow is $q_M = k_6\mu_2 X_2$. Bernard and coauthors' identified yields close the COD of both steps with no single acid, since their $k_2 = 116.5$ would need 350 g COD per mole of acid while acetic acid carries 64. The application keeps their $k_1 = 42.14$ and $k_3 = 268$ and derives the other two so that COD closes exactly, as shown in Equation (11.4), which gives $k_2 = 636.25$ and $k_6 = 245.81$ mol/kg.

$$k_2 = \frac{k_1 - 1.42}{0.064}, \qquad k_6 = \frac{0.064\,k_3 - 1.42}{0.064} \tag{11.4}$$

11.3.2 pH, Headspace and Gas

The pH is the root of the charge balance on the liquid, as shown in Equation (11.5).

$$\begin{aligned} Z = C\,\frac{K_1 h + 2K_1K_2}{h^2 + K_1 h + K_1K_2} \\ + S_2\,\frac{K_a}{K_a + h} + \frac{K_w}{h} - h \end{aligned} \tag{11.5}$$

where $h = 10^{-\mathrm{pH}}$ in the same units, $\mathrm{p}K_1 = 6.31$ is ADM1's 6.35 at 25 °C moved to 35 °C, $\mathrm{p}K_a = 4.76$, $K_2 = 5 \times 10^{-11}$ and $K_w = 2.1 \times 10^{-14}$. The root is found by 48 bisection steps between pH 2 and 13, a fixed count so that the derivative stays a pure function of the state. Between pH 5.5 and 8 the balance reduces to the two-step model's closed form, and at the design point the two agree within 0.0063 (7.3884 against 7.3947). The headspace is the two-step model's quasi-steady one, with the carbon dioxide transfer and its partial pressure given by Equation (11.6).

$$\begin{aligned} q_C &= k_L a\,(\mathrm{CO_2} - K_H P_c) \\ P_c &= \frac{\phi - \sqrt{\phi^2 - 4K_H P_T\,\mathrm{CO_2}}}{2K_H} \\ \phi &= \mathrm{CO_2} + K_H P_T + \frac{q_M}{k_L a} \end{aligned} \tag{11.6}$$

where $\mathrm{CO_2}$ is the dissolved carbon dioxide from the speciation of C, $k_L a = 20\,\mathrm{d}^{-1}$, $K_H = 27\,\mathrm{mmol/(L\,atm)}$ and $P_T = 1\,\mathrm{atm}$. The gas flows are $q_{CH_4} = q_M V v_m$ and $q_{CO_2} = q_C V v_m$ with $v_m = 0.022\,414\,\mathrm{Nm^3/mol}$.

The holder keeps methane and carbon dioxide apart, so that its methane share $x_m = H_m/(H_m + H_c)$ is a state, and every draw takes gas at that share, as shown in Equation (11.7).

$$\begin{aligned} \frac{\mathrm{d}H_m}{\mathrm{d}t} &= q_{CH_4} - x_m(aF_{up} + F_{flare}) - F_{fuel} \\ \frac{\mathrm{d}H_c}{\mathrm{d}t} &= q_{CO_2} - (1 - x_m)(aF_{up} + F_{flare}) \\ &\quad - F_{fuel}\,\frac{1 - x_m}{x_m} \\ F_{flare} &= 200\,\max(0,\ H - 0.98H_{max}) \\ a &= \min\left(1, \frac{H}{0.02H_{max}}\right) \end{aligned} \tag{11.7}$$

where $H = H_m + H_c$, the relief gain is $200\,\mathrm{d}^{-1}$, and a stops an empty holder from being drawn. The heater needs $Q_h \cdot 86.4/(\eta_h L_{CH_4})$ $\mathrm{Nm^3}$ of methane per day at an efficiency $\eta_h = 0.85$ and a lower heating value of $35.88\,\mathrm{MJ/Nm^3}$, and the holder supplies the fraction F_{fuel} of it scaled by a and by a burner factor that falls from 1 at 40 % methane to 0 at 30 %, the rest being bought as pipeline gas. The upgrader sells $R_{CH_4} x_m a F_{up}$ with $R_{CH_4} = 0.97$, and the other 3 % of its methane is the declared slip. Revenue accrues at 0.03774 MMBtu per $\mathrm{Nm^3}$ of RNG (a higher heating value of $39.82\,\mathrm{MJ/Nm^3}$) times \$48 per MMBtu, the sum of a \$3 gas price, a \$25 federal fuel-standard credit and a \$20 state low-carbon fuel credit, all declared teaching prices.

11.3.3 Heat, Feed and Events

The tank's temperature follows Equation (11.8).

$$\rho c_p V\,\frac{\mathrm{d}T}{\mathrm{d}t} = 86.4\,Q_h - \rho c_p Q(T - T_{feed}) - 86.4\,UA(T - T_{amb}) \tag{11.8}$$

with $\rho c_p = 4.18\,\mathrm{MJ/(m^3\,K)}$, Q_h and UA in kW and kW/K, a 15 °C feed and 5 °C outdoors. At $250\,\mathrm{m^3/d}$ with the heater tripped the tank cools with a time constant of $\rho c_p V/(86.4\,G) = 13.73\,\mathrm{d}$, where $G = 14.09\,\mathrm{kW/K}$ is the feed's and the shell's conductance, toward 13.58 °C. The feed is a blend of manure ($40\,\mathrm{kg\,COD/m^3}$, 8 mmol/L acids, 170 meq/L alkalinity, 75 mmol/L carbon) and food waste (160, 80, 40 and 10 in the same units) at the food-waste share f_{fw}, which gives $S_{1,in} = 54.4\,\mathrm{kg\,COD/m^3}$ at the design share of 12 %. Five events change the plant between steps, each leaving a mark on the trends: a slug that makes half of one day's feed food waste (1.84 times the COD, 32 % less alkalinity), a feedstock change to 25 % food waste, a heater fault that sets Q_{max} to zero for six days, an inhibitor shock that removes 35 % of the methanogens' activity and decays as $\exp(-t/25\,\mathrm{d})$, and a drift of the pH probe at $0.06\,\mathrm{d^{-1}}$.

11.3.4 Measurements

The controllers read only declared measurements. Online, the biogas meter (2 %), the methane analyzer (0.3 points), the pH probe (0.02), the thermometer (0.05 °C), the holder level (0.5 points) and the upgrader's inlet methane analyzer (0.3 points) carry first-order filtered Gaussian noise with a correlation time of 0.05 d (0.02 d for the level), seeded so that a case replays exactly, and the heater reports a trip contact. The titration samples the true acids and alkalinity at 08:00 every $T_{lab} = 1\,\mathrm{d}$ and reports them $\tau_{lab} = 6\,\mathrm{h}$ later with 3 % and 1 % relative noise, as the ratio

$$r_{lab} = \frac{60.05\,S_2}{50.04\,Z}, \tag{11.9}$$

of acids in mg/L as acetic acid to alkalinity in mg/L as calcium carbonate, held until the next report. A weekly laboratory reports the feed COD a day after its Monday sample with 4 % noise. The populations, the true acids and the true pH are logged under a `truth_` prefix, and one of the application's tests searches the controller source to prove that it never reads them.

11.3.5 Parameters, Initialization and Numerics

Table 11.2 lists the parameters. The acidogen kinetics and k_1, k_3 and k_4 are the values identified by Bernard and coauthors, read through the table that Attar and Haugen reproduce [16], because the original was not available to read. The methanogen constants are declared for teaching and place the fold at 1.8 times the design feed. With the original methanogens ($\mu_{2,max} = 0.74\,\mathrm{d^{-1}}$, $K_{S2} = 9.28$, $K_{I2} = 256\,\mathrm{mmol/L}$) in a stirred tank the fold would sit at a retention of about 1.9 days, and the full-scale fit to a municipal reactor has $K_{I2} = 2396\,\mathrm{mmol/L}$ [16], so no single published set fits a farm digester on manure.

Every case starts lined out. The liquid begins at the analytic steady state of Equation (11.1) and (11.3) at the case's feed and 35 °C, the two Monod and lower Haldane roots giving S_1 and S_2, a linear solve giving the populations and 80 bisection steps on the carbon balance giving C. The holder starts at its level setpoint with the gas's own composition. Case 3 starts instead from the design state after eight days at twice the design feed with no dosing. The 22 states (seven for the liquid and the temperature, two for the holder and thirteen ledgers) advance by fourth-order Runge-Kutta steps of $\Delta t = 600\,\mathrm{s}$ in an engine whose time base is the day, the controllers execute every hour, and a row is logged every hour (the engine of Section 1.6.1). Every event and every laboratory sample and report happens once, after a step.

11.3.6 Assumptions

The two-step model lumps hydrolysis, acidogenesis and acetogenesis into one Monod step and every acid into acetic-acid equivalents, so a fibrous manure is digested faster here than in a real tank, where hydrolysis often limits. The tank is perfectly mixed and keeps no solids back. The methanogens are inhibited only by the total acids and by pH, so free ammonia, sulfide and the sodium that bicarbonate adds are absent, and the inhibitor shock is a declared factor rather than a speciation. The headspace is quasi-steady and the holder's pressure is not modeled. The carbon dioxide yield k_5 is a teaching

Table 11.2 Model Parameters. A: Adjustable in the Parameters Panel, H: Hidden, D: Derived. Units of Concentration in mmol/L Unless Given

Symbol	Meaning	Value	Unit	
$\mu_{1,max}$, K_{S1}	acidogen growth, half-saturation[a]	1.2, 7.1	d^{-1}, kg/m^3	A
$\mu_{2,max}$	methanogen growth[t]	0.164	d^{-1}	A
K_{S2}, K_{I2}	Haldane constants[t]	20, 130	—	A
k_d	decay, both populations[b]	0.02	d^{-1}	A
α	biomass leaving with liquid	1	—	H
k_1, k_3, k_4	yields[a]	42.14, 268, 50.6	per kg VSS	H
k_2, k_6	COD-consistent yields	636.25, 245.81	mol/kg VSS	D
k_5	CO_2 per methanogen[t]	170	mol/kg VSS	H
θ_1, θ_2	temperature factors[t]	1.04, 1.07	per °C	A
pH_{UL}, pH_{LL}	methanogen pH limits[b]	7.0, 6.0	—	A
$pH_{UL,1}$, $pH_{LL,1}$	acidogen pH limits[b]	5.5, 4.0	—	H
pK_1, pK_a	carbonic, acetic acid	6.31, 4.76	—	A, H
K_H, k_La	CO_2 solubility, transfer	27, 20	–, d^{-1}	A
P_T	headspace pressure	1	atm	H
V, UA	volume, shell loss	4000, 2.0	m^3, kW/K	A
T_{amb}, T_{feed}	outdoor, feed temperature	5, 15	°C	A
Q_{max}, η_h	heater capacity, efficiency	600, 0.85	kW, —	A, H
$S_{1,man}$, Z_{man}	manure COD, alkalinity	40, 170	kg/m^3, meq/L	A
$S_{2,man}$, C_{man}	manure acids, carbon	8, 75	—	H
$S_{1,fw}$, Z_{fw}	food-waste COD, alkalinity	160, 40	kg/m^3, meq/L	A
$S_{2,fw}$, C_{fw}	food-waste acids, carbon	80, 10	—	H
f_{fw}, f_{new}, f_{slug}	food-waste shares	12, 25, 50	%	A
I_{NH_3}, t_{fault}	inhibitor shock, fault length	35, 6	%, d	A
H_{max}	holder volume	2000	Nm^3	A
$F_{up,max}$, $F_{up,min}$	upgrader capacity, turndown	360, 30	Nm^3/h, %	A
R_{CH_4}, x_{min}	methane recovery, inlet floor	97, 45	%	A
prices	gas, federal, state credits	3, 25, 20	$/MMBtu	A
c_{alk}	bicarbonate price	0.40	$/kg	A
T_{lab}, τ_{lab}	titration period, delay	1, 0.25	d	A
T_{COD}	feed COD period	7	d	A

[a]Bernard et al. through Attar and Haugen. [b]ADM1. [t]Teaching value.

value that puts the design biogas at 60.0 % methane, and the inorganic carbon is not balanced against the COD. The ratio limit of 0.4, the supervisor's back-off at 0.30, the upgrader's 45 % methane floor, its 97 % recovery and the 98 % relief are teaching values. A plant takes its ratio limit from its own titration history (a healthy range of 0.1 to 0.35 is the published guidance [7]), and its upgrader limits from the membrane maker's data and the pipeline's gas-quality tariff.

11.4 The Fold

11.4.1 Where the Methanogens Lose

A steady state with methanogens present needs $\mu_2(S_2) = \alpha D + k_d$, the condition Section 1.4.3 sets out for Equation (1.5). Below the peak of the Haldane curve there are two roots, a lower one that is stable and an upper one that bounds the basin of the sour state, and above the peak there are none. With the teaching constants $S_{2M} = \sqrt{20 \cdot 130} = 51.0\,\text{mmol/L}$ (3062 mg/L) and $\mu_2(S_{2M}) = 0.0919\,d^{-1}$, so the critical dilution is $D_c = 0.0719\,d^{-1}$, a feed of 287.6 m^3/d, a retention of 13.9 days and 1.80 times the design. At S_{2M} with the design alkalinity the pH is 7.25, above ADM1's limit of 7, so the pH term does not move the fold at the design point. It sharpens the collapse once the acids have consumed the bicarbonate. The application's map function computes the fold and both branches from the same growth law the engine integrates, at the present temperature, inhibitor, alkalinity and dissolved carbon dioxide, as shown in Figure 11.2. An exothermic reactor has steady states of the same shape, with an ignition and an extinction fold (Volume 7, Chapter 2).

11.4.2 A Fold Set by the Retention

At a steady state the acids depend on D alone, because $\mu_2(S_2) = \alpha D + k_d$ contains no feed concentration. At 250 m^3/d the steady acids are 25.184 mmol/L whether the feed is 12, 25 or 50 % food waste. A richer feed needs more methanogens, which grow into it at the same retention, not less volume. The steady ratio still moves a little with the feed's alkalinity, from 0.196 to 0.220 at 250 m^3/d after the change to 25 % food waste. The consequence for control is that a feedforward dividing the loading by a measured feed COD answers the wrong model. At a ratio of 0.25 the steady feed falls only 3.3 %, from 271.0 to 261.9 m^3/d, after the feedstock change, while the feedforward cuts it 22.3 %. This mechanism is behind the revenue the feedforward costs in case 6 and in the challenge.

🕮 **Pause.** Try Exercise 11.2 before reading on.

11.4.3 A Ratio That Stays Quiet

The steady ratio along the stable branch is 0.084, 0.104, 0.163, 0.215, 0.259, 0.303, 0.328 and 0.396 at 50, 60, 80, 90, 95, 98, 99 and 100 % of the critical feed. The ratio barely moves until the last few percent, so in this model 0.3 already marks the edge and 0.4 is the fold itself. The 0.4 limit is therefore a teaching value placed at this model's fold, between the published healthy range of 0.1 to 0.35 [7] and the value near 0.5 that operator training material reads as a sour digester.

11.4.4 Temperature Moves the Fold

The methanogens' factor $\theta_2 = 1.07$ against $\theta_1 = 1.04$ makes a cooling tank produce acids almost as fast while consuming them more slowly. A fall of 7 °C costs the methanogens 38 % of their rate. In case 7 the tank reaches 27.41 °C on day 8.42, and the fold at that temperature and the state's alkalinity and dissolved carbon dioxide is 139.9 m^3/d, about half its warm value, while S_{2M} is unchanged. The same feed is safe warm and beyond the fold cold, as the green branch with triangle markers in Figure 11.2 shows.

11.5 The Control Layers as Implemented

Three layers act on the measurements of Section 11.3. Every loop uses the velocity-form PID routine of Section 1.6.2 with derivative action off, so its clamp and its back-calculation anti-windup, whose tracking constant is τ_I, are those of Equations (1.14) and (1.15). In the convention of Section 1.3, the temperature, pH and loading loops are reverse acting, because more heat, more bicarbonate and a higher loading raise the temperature, the pH and the ratio. The level trim of the upgrader is direct acting, because a holder above its setpoint calls for a larger draw. Each layer-one loop can be switched between Manual and Auto on its own, and the loading loop follows the header's switch.

11.5.1 Temperature, pH and the Holder

Every hour the temperature loop moves the heater duty with $K_c = 300\,\mathrm{kW/°C}$ and $\tau_I = 1\,\mathrm{d}$ between 0 and the capacity the trip contact says is available, so the integral does not wind up during a fault. The pH loop doses bicarbonate between 0 and 80 kmol/d with $K_c = 100\,\mathrm{kmol/d}$ per pH unit and $\tau_I = 0.5\,\mathrm{d}$, and it ships in Manual at zero dose except in cases 3 and 5. The upgrader's feed follows the metered biogas less the heater's fuel, trimmed on the holder level, as shown in Equation (11.10).

$$F_{up} = q_{gas,m} - \frac{86.4\,Q_h}{\eta_h L_{CH_4} \max(0.3,\, x_{m,m})} + K_L (L_m - L_{SP}) H_{max} \tag{11.10}$$

where the subscript m marks a measurement, $K_L = 2\,\mathrm{d}^{-1}$ and $L_{SP} = 50\,\%$. The feed is held between its turndown of 30 % and its capacity of 8640 Nm^3/d, runs at the turndown only while the holder is above its setpoint, and stops below a level of 10 % (restarting above 40) or below the 45 % methane floor (restarting five points above it).

11.5.2 The Loading Loop on the Titration

The loading loop acts once per report, on a measurement of the kind Section 1.4.5 describes. At the k-th report it runs the PI with a step equal to the time since the previous report and holds the requested loading until the next one, as shown in Equation (11.11).

$$\begin{aligned} e_k &= r_{SP,k} - r_{lab,k}, \qquad \Delta t_k = t_k - t_{k-1} \\ OLR_k &= \mathrm{PI}(e_k;\ K_c,\ \tau_I,\ \Delta t_k), \qquad 0.2 \le OLR_k \le 6 \\ Q &= \min\left(400,\ \frac{OLR_k V}{\hat{S}_{1,in}}\right) \end{aligned} \tag{11.11}$$

where OLR is in kg COD/m^3/d, $\hat{S}_{1,in}$ is the nameplate 54.4 kg COD/m^3 of the design blend or, with the feedforward on, the weekly laboratory's last report, and a ratio below its setpoint raises the loading. The defaults are $K_c = 8$ kg COD/m^3/d per unit of ratio, $\tau_I = 8$ d and $r_{SP} = 0.20$, about twice the design ratio. When the pump limit cuts the requested loading, the integral is moved by the shortfall times $\Delta t_k / \tau_I$, a second back-calculation against what the pump delivered. In Manual the loop tracks the operator's feed.

11.5.3 The Economic Supervisor

With the loading loop in Auto, the supervisor acts on the same report before the PI does. If the report exceeds the back-off ratio $r_{off} = 0.30$ it drops the setpoint to its floor of 0.20, so the PI's proportional action cuts the loading on that report. Otherwise, if the report lies within 0.03 of the setpoint, it raises the setpoint by 0.01 up to a ceiling of 0.28. A variant that also cut the loading by a fifth at once was measured when the application was built, and it lost revenue while breaking the limit, and it was not kept.

11.5.4 Plant, Map and Controller Models

The plant model of Section 11.3 is integrated. The map of Figure 11.2 is a second model, the steady branches at the present slow quantities, used only for display. The sampled controllers carry no internal biological state. The predictive feed controller (MPC) instead integrates an independent six-state nominal AM2 observer. It reads measured temperature and methane flow, reported acid and alkalinity, the selected feed COD and applied feed and dose. Feed COD is the declared design value unless COD feedforward is On, when the available laboratory report is used. Neither actual populations, hidden inhibition, current kinetic parameters nor future events are inputs.

Every six hours the observer balance is centrally linearized and discretized with its affine term. The framework's quadratic program of Section 1.6.3 predicts sixty steps, fifteen days, in twelve move blocks. Ratio weight 100, feed-move weight 0.001 and unit methane reward compete. Feed bounds 0–400 m^3/d and planned block moves of 10 m^3/d are hard. The smaller of 0.38 and the supervisor back-off ratio is a soft ceiling, with quadratic and linear prices 10^5 and 10^3. The methane reward omits holder and upgrader constraints. This follows the predictive optimization framework reviewed in [20]. Bounded output-bias correction follows the offset-correction principle of [21]: ratio reports are compared with predictions at their sample times, with gain 0.4 and bound 0.25, while hourly methane correction uses gain 0.05 and bound 5 10^3 Nm^3/d. It does not identify biomass or inhibitor kinetics. The other loops and supervisor remain active. The sampled PI tracks the applied feed. A rejected plan holds feed on the handover scan, marks the historian and returns to PI until a valid plan resumes. The first plan is computed while the engine primes.

A hand dilution Jacobian agrees within 1.4×10^{-13} and the ratio Jacobian within 6.6×10^{-14}. Over one six-hour interval, feed perturbations up to ± 10 m^3/d give ratio prediction errors below 6.3×10^{-6} against the nonlinear nominal observer. This is a local check, not a bound on a fifteen-day forecast.

Table 11.3 The Cases as Shipped

#	Setting and scripted events	Teaching point
1	Manual, 160 m^3/d, learner raises to 200 at day 1	fast gas, slow acids
2	Manual, feed to 400 m^3/d at day 1.35	gas peak before the warning
3	Soured start, 80 m^3/d, pH held at 7.0	recovery takes weeks
4	Auto, r_{SP} 0.20, K_c 8, τ_I 8 d	gain a daily sample allows
5	Feed to 300 m^3/d at day 1.35, pH held at 7.2	a buffer, not a cure
6	Auto at 0.22, slug day 3.35, change day 13.9	co-digestion, feedforward
7	Manual 250 m^3/d, heater fault day 2.4	the fold moves with T
8	Auto with supervisor, inhibitor day 20.4	revenue against margin
9	Auto MPC, target 0.25, inhibitor day 20.4	nominal forecast versus PI
10	Challenge: 250 m^3/d held (gain 0), change day 15.5, fault day 38.5	the whole season

Table 11.4 Measured Results at the Default Seed (Days From the Start)

#	Setting	Quantity	Value
1	200 at day 1	biogas day 0, 2, 15	4674, 5667, 5779 Nm^3/d
		acids start, peak (day 4.5)	732, 1106 mg/L
2	400 at day 1.35	upgrader trip, gas peak	day 2.01, 2.63
		titration over 0.4, soured	day 3.58, 4.51
3	pH held at 7	recovered, bicarbonate	day 45.3, 1227 kmol
	seeded before Run	recovered, bicarbonate	day 25.4, 866 kmol
	no bicarbonate	pH at day 30	4.42, still sour
4	$K_c = 8$	feed day 20, 40; worst	233, 248; 0.207
	$K_c = 30$	days at 400, under 100	7, 6 of 40
5	pH held at 7.2	soured, bicarbonate	day 18.85, 1002 kmol
	no bicarbonate	soured	day 10.22
6	feedforward off, on	titration peak after change	0.289, 0.275
		revenue to day 25	$159,425, $153,734
7	held at 250	minimum T, soured	27.4 °C, day 10.60
	cut to 160 at day 7.58	reports over 0.4	4, not soured
8	supervisor on, off	revenue to day 40	$190,704, $187,176
		worst titration	0.326, 0.305

11.6 Nine Cases and a Season, Measured

The cases appear in Table 11.3. Every value below was measured without a display on the engine the browser runs, at the default seed 20261026, and any change named in the text is applied at its stated day after Run. Section 1.6.4 describes how a case, its completion rule and its scorecard are set up. Table 11.4 gathers the headline values.

> ✦ **DID YOU KNOW? Knowing the Feed Lost Money**
>
> From the simulator: in case 6 a feedforward on the weekly laboratory's feed COD lowers the titration peak after the feedstock change only from 0.289 to 0.275, and the revenue to day 25 from $159,425 to $153,734 (Table 11.4). The steady acids do not depend on the feed's strength.

11.6.1 A Long Forecast With Unmeasured Biology

Case 9 compares the same 0.25 target and instruments for forty days (Figure 11.3). MPC ratio IAE is 3.130 versus 1.403 ratio-days for PI, feed travel is 640 versus 462 m^3/d, and six reports exceed 0.4 versus one. Revenue is $206,058 versus $213,896, with 4.04% flared versus zero. The hidden inhibitor at day 20.4 makes the nominal biological forecast optimistic. Scalar output correction cannot reconstruct lost methanogen activity, and methane production reward does not imply sold gas revenue. Halving the integration step retains both revenue and IAE orderings (MPC 3.086 versus PI 1.403 ratio-days). A hindsight PI grid of targets 0.20, 0.24 and 0.26 with gains 4 and 8 includes two settings

Table 11.5 Matched Feed Control Comparisons, PI / MPC

Case	Ratio IAE	High reports	Revenue ($1000)	Plans / µs
New, 40 d	1.403 / 3.130	1 / 6	213.90 / 206.06	160 / 539
Ratio, 40 d	1.188 / 1.145	0 / 0	215.05 / 233.63	160 / 390
Co-feed, 40 d	1.157 / 2.690	1 / 5	274.27 / 279.36	160 / 353
Supervisor, 40 d	1.593 / 2.963	0 / 5	190.70 / 204.33	160 / 395
Season, 60 d*	2.000 / 2.412	0 / 3	428.20 / 435.39	240 / 359

*Original suggested settings. Plans and mean Node solve microseconds are for MPC, with zero rejected plans in these runs.

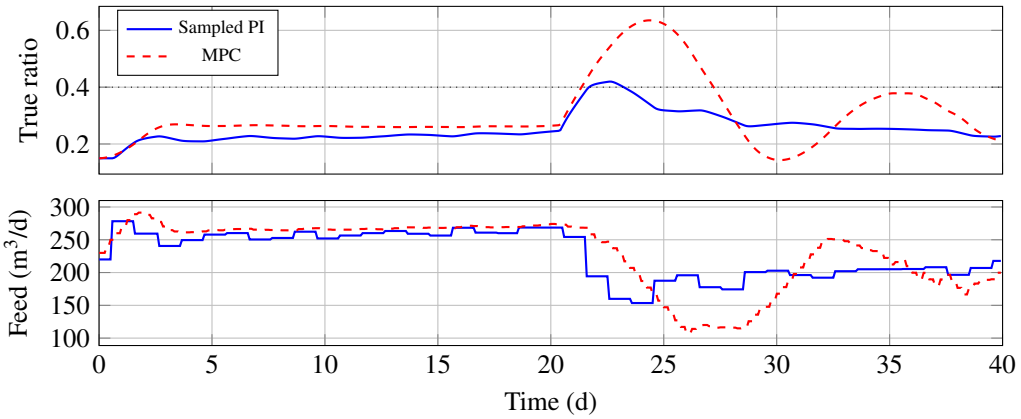


Figure 11.3 Case 9 with matched target and measurements, sampled PI as the solid line and MPC as the dashed line. Hidden inhibition begins on day 20.4. The nominal MPC accumulates more acids and flares more gas than sampled PI, despite its longer forecast. The dotted line is the 0.4 ratio indicator, not a hard state constraint.

with no report over 0.4. Fixed feeds 185, 220 and 250 m^3/d earn less than either matched controller, and the latter two sour the tank.

The ordinary ratio-loop case instead earns $233,627 with MPC versus $215,050 with PI, and ratio IAE is 1.145 versus 1.188. Co-digestion earns $279,363 versus $274,268, but has five high titrations versus one. In the suggested sixty-day challenge MPC earns $435,389 versus $428,198, but three high reports make it fail while PI passes. Thus the MPC does not replace the suggested settings of the challenge. The manual heater-fault case gives identical outputs when only the architecture option is changed. Active Auto operation is checked separately on all ten cases.

Forty browser plans per viewport average 1.13–2.46 ms, excluding observer propagation. At 0.05 d/s playback this costs 0.023–0.049% of one core, and 0.45–0.98% at the top 1 d/s. The six-hour decision is held between plans. Table 11.5 uses identical seeded instruments and no preview.

11.6.2 The Gas Answers in a Day, the Acids in a Week

Case 1 raises the feed from 160 to 200 m^3/d on day 1. The biogas goes from 4674 to 5667 Nm^3/d by day 2, and by day 7 it is within 0.11 % of its day-15 value of 5779 Nm^3/d, 24 % above the start. The acids rise from 732 to a peak of 1106 mg/L on day 4.5 and still read 1044 mg/L on day 15. The methanogens already present burn more acid at once because their rate rises with the acids, but only a larger population brings the acids back down, and at a 20-day retention it needs more acids to grow as fast as it is washed out.

11.6.3 The Overload

Case 2 raises the feed to 400 m^3/d at day 1.35, past the fold, as shown in Figure 11.4. The methane rises 41 % to its peak on day 2.61, while the biogas doubles to 9504 Nm^3/d on day 2.63, because the

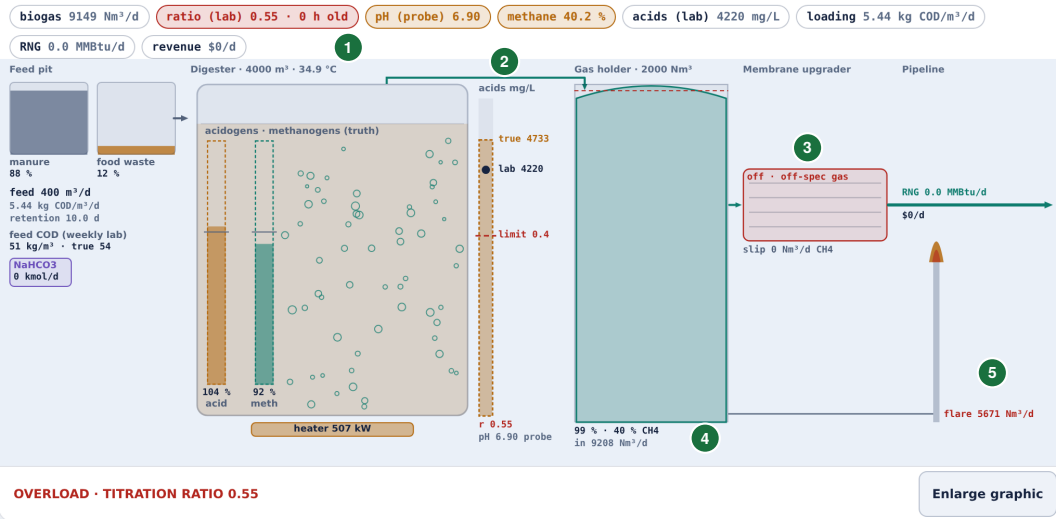


Figure 11.4 The overload of case 2 on the Plant view at $t = 3.60$ d, just after the first titration over the limit, cropped to the plant view. The feed has run at 400 m^3/d since day 1.35. Callout 1 is the report of the 08:00 sample, a ratio of 0.55, in the badge row, where the meter reads 9149 Nm^3/d of biogas at 40.2 % methane and the probe still reads pH 6.90. Callout 2 is the acids column, with the laboratory's 4220 mg/L against the true 4733 mg/L (a true ratio of 0.61). Callout 3 is the upgrader, off on off-spec gas since day 2.01. Callout 4 is the holder, 99.4 % full at 40.5 % methane, and callout 5 the flare, which takes 5671 Nm^3/d of relief.

acids displace bicarbonate and the carbon dioxide they release leaves with the gas, which reaches 2.97 times its design flow on day 3.08. That carbon dioxide gives the first warning of all. On day 2.01 the holder's gas falls below the upgrader's 45 % methane floor, the upgrader trips and the RNG sales stop, after the revenue rate had peaked 43 % above design on day 1.88. The true ratio passes 0.4 on day 2.71. The titration that shows it is drawn at 08:00 on day 3 and reported on day 3.58, 0.95 days after the gas peak, and the digester sours on day 4.51 (methane below half its three-day high with the true ratio above 0.4), when 42.5 % of the biogas made so far has been flared and the run has earned $8,268. On the map the case's state sits at a loading of 5.44 kg COD/m^3/d, beyond the fold where no steady state exists, and its acids climb from 2956 mg/L at the gas peak to 6824 mg/L at the souring.

Pause. Try Exercise 11.1 before reading on.

> **TRY IT IN THE APP** **2 · The Overload: Feast, Then the Crash**
>
> Load the case and press run. Watch the biogas trend after the feed jumps to 400 m^3/d on day 1.35: the gas doubles to a peak on day 2.63, and the titration that shows the acids over the limit is reported only on day 3.58. The tank sours on day 4.51.

11.6.4 Recovery

Case 3 starts from the overload continued for eight days. The pH is 5.33, the ratio 1.50, and the methanogens are still 79 % of their design population but make almost no methane at that pH. At half feed with the pH loop holding 7.0, the methanogens fall to 52 % of design on day 12.6 before regrowing, and the digester is recovered (true pH above 7, ratio below 0.3, methane above 45 % of design) on day 45.3 after 1227 kmol of bicarbonate, $41,218 at the teaching price. Without bicarbonate the pH stands at 4.42 on day 30 and the tank stays in its sour steady state. Replacing

a tenth of the contents with healthy digestate before Run shortens the recovery to 25.4 days and 866 kmol. Dilution alone at half feed needs 50 days per e-fold, which is why the acids must be consumed rather than washed out.

11.6.5 The Gain a Daily Sample Allows

Case 4 closes the loading loop at $r_{SP} = 0.20$. The feed rises from 160 to 233 m^3/d by day 20 and 248 m^3/d by day 40, with no titration above 0.207 and every report within 0.03 of the setpoint after day 19.6. At a gain of 30 the first report sends the feed to 390 m^3/d and the next to 69 m^3/d, and at the day boundaries over 40 days it reads the pump's 400 m^3/d seven times and less than 100 m^3/d six times. Each step is decided on acids that have not yet answered the step before. A column tuned against a slow analyzer meets the same limit (Volume 5, Chapter 3). At the shipped gain a report 0.05 above the setpoint moves the loading by 0.4 kg COD/m^3/d, about 29 m^3/d of the design blend.

11.6.6 Bicarbonate, a Buffer and Not a Cure

Case 5 raises the feed to 300 m^3/d, a 13.3-day retention, just past the fold, with the pH loop holding 7.2. The titration passes 0.4 on day 5.58 and the digester sours on day 18.85 after 1002 kmol (84 t, $33,666) of bicarbonate, against day 10.22 without it. Holding the pH bought 8.6 days and removed no acid. Past the fold the Haldane term inhibits the methanogens whatever the pH, and the time bought is time to cut the feed.

Pause. Try Exercise 11.3 before reading on.

> **TRY IT IN THE APP** **5 · Bicarbonate on pH: A Buffer, Not a Cure**
>
> Run the case as shipped, with the pH loop holding 7.2: the tank sours on day 18.85 after 1002 kmol of bicarbonate. Reload it, set the pH loop to Manual at zero dose and run again: the tank sours on day 10.22. The time the bicarbonate buys is time to cut the feed.

11.6.7 Co-Digestion and the Feedforward

Case 6 runs the loop at 0.22 from 240 m^3/d. The food-waste slug on day 3.35 raises the biogas from 7133 to 11 844 Nm^3/d, more than the upgrader's 8640 Nm^3/d, so the holder relieves to the flare from day 3.97 to 4.83 and 3.9 % of the biogas to day 25 is flared, while one titration reads 0.454 and the loop rides it out. After the feedstock change on day 13.9 the weekly laboratory reports a COD of 73.7 on day 15.33 against a true 70.0. With the feedforward on, the titrations after the change peak at 0.275 instead of 0.289 and the revenue to day 25 falls from $159,425 to $153,734. The feedforward cuts the volume of a feed whose steady acids it does not change, which is the mechanism of Section 11.4.

11.6.8 The Heater Fault and an Infeasible Variant

Case 7 feeds 250 m^3/d warm and trips the heater on day 2.4 for six days. The tank falls from 35.0 to 27.4 °C by day 8.4, the fold moves to 140 m^3/d, the titration crosses 0.4 on day 7.58 and the digester sours on day 10.60 if nothing is done. Cut to 160 m^3/d at that report, it survives with four titrations over the limit, the last reported on day 10.58. Held at 250 m^3/d, it sours at every bicarbonate dose from 0 to 80 kmol/d in steps of 10, on days 10.60, 12.78, 18.30, 22.22, 25.26, 27.95, 30.49, 33.00 and 35.56. Only less feed answers a fold that has moved below the feed, and the application says so. A fault that cooled the tank by 8 °C in two days would need a mean loss of 774 kW, 2.6 times the 302 kW this tank loses at 35 °C, so the fault cools it by 7.6 °C over six days instead.

11.6.9 The Economic Supervisor

Case 8 adds the supervisor to case 4's loop and removes 35 % of the methanogens' activity on day 20.4. The feed reaches 240 m^3/d before the shock (233 with the setpoint fixed at 0.20), the shock cuts it to 143 (152), and the setpoint reaches 0.25 on day 31.6. The revenue to day 40 is $190,704 against

Table 11.6 The Sixty-Day Challenge at the Default Seed

Policy	over 0.4	worst ratio	flared (%)	revenue ($)	verdict
shipped, gain 0 at 250	41	sour	9.11	166,751	fails
constant 185	0	0.368	0.00	341,870	fails
constant 190	1	0.411	0.00	350,866	fails
loop, fixed 0.20	0	0.303	0.00	406,506	fails
loop, fixed 0.24	0	0.360	0.00	426,881	passes
loop, supervisor (suggested)	0	0.359	0.00	428,198	passes
suggested + feedforward	0	0.331	0.00	416,550	passes
loop, fixed 0.26	0	0.389	0.00	434,400	passes
loop, fixed 0.28	1	0.417	0.02	440,312	fails

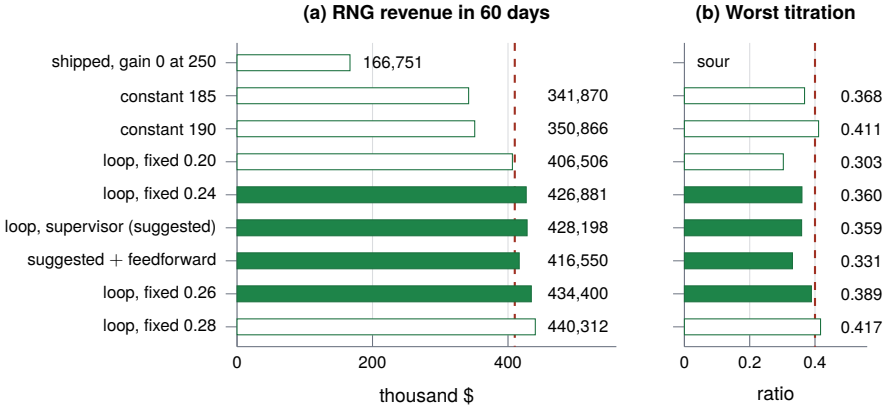


Figure 11.5 The nine policies of Table 11.6 drawn as bars: the RNG revenue of the sixty-day challenge against its $410,000 floor, and the worst titration against the 0.4 limit (dashed lines). Solid bars pass and white bars fail. Notice that revenue rises with the setpoint while the worst report climbs toward the limit, so the policy that earns most, the fixed 0.28, fails on one report, and that the constant feeds fail on revenue even where they keep the ratio low.

$187,176, 1.9 % more, with a worst titration of 0.326 against 0.305. Near a fold the last few percent of gas cost most of the margin.

11.6.10 The Season

The challenge runs 60 days at 250 m^3/d, about 87 % of the warm fold, with the feedstock change on day 15.5 and the heater fault on day 38.5. It passes with no titration report above 0.4, at most 2 % of the biogas flared and at least $410,000 of RNG revenue, on the case's own parameters. Table 11.6 lists nine policies, each with the same measurements, and Figure 11.5 draws their revenue and worst titration against the two limits. As shipped, the loop in Auto with no gain holds the feed, the methanogens cannot grow into the richer feed fast enough (their spare growth at 87 % of the fold is 0.0094 d^{-1}), the titration passes 0.4 on day 19.58 and the tank sours on day 26.74. The suggested loop with the supervisor passes. A fixed setpoint of 0.26 earns more than the supervisor with a worst report of 0.389, and the supervisor's share is finding the edge from 0.20 without hindsight. The feedforward, added to the suggested settings, costs $11,648. The first report of the richer feed arrives on day 22.33, when the loop has already cut the feed from 267 to 221 m^3/d on the titrations, and the feedforward cuts it again to 173 m^3/d. With a noise-free COD it costs $18,164, so the laboratory's noise is not the mechanism. Across seeds 1 to 5 the shipped run sours on days 26.65 to 26.78 and earns $166,101 to $167,274, and the suggested run passes every time with $426,332 to $429,879 and worst reports of 0.351 to 0.369.

11.6.11 Margin, Revenue and the Age of the Titration

Two further measurements bear on the supervisory question. The first is the distance to the fold that each policy keeps. Computed every hour from day 2 on from the true state with the map’s fold at the present temperature, inhibitor, alkalinity and dissolved carbon dioxide, the mean margin over the season falls from 54.7 to 41.8, 40.0, 36.8 and 32.7 m^3/d for the fixed setpoints 0.20 and 0.24, the supervisor, and 0.26 and 0.28, while the revenue rises from $406,506 to $440,312. Every one of these policies runs past the instantaneous fold during the heater fault, by 9.3 to 34.9 m^3/d at its worst hour, and none of them sours. A short excursion past the fold is not a souring, because the acids need days to climb through the basin, which is why the titration’s reaction time matters more than its threshold.

The second is the age of the titration. With the case’s parameters changed (so the attempt is no longer graded), sampling every half day with a 3 h delay moves case 2’s first warning from day 3.58 to 2.96, and every quarter day with 1.5 h to 2.90, while a daily sample reported at once warns on day 3.33. On the challenge the suggested settings then earn $432,438 and $435,857 instead of $428,198, with worst reports of 0.356 and 0.337. The literature treats the same trade with extremum seeking for biogas [2], best operating conditions for the two-step model [3] and predictive feed control against temperature disturbances [18]. The two measurements here are properties of this teaching plant.

11.7 Three Kinds of Evidence

The first kind is code verification, which asks whether the engine solves the equations written above. The COD balance closes at the design steady state with a relative residual of 4.2×10^{-15} and over the 60 days of the suggested challenge, through both events, to 3.9×10^{-14} of the feed. The holder balance closes to 2.1×10^{-15} of the production over the same run and to 4.0×10^{-15} over case 6 to day 25, where no flaring step occurs below 98 % full, and the slip is 3.000 000 % of the upgrader’s methane. The pH root agrees with the closed form within 0.0063 at the design point and within 0.008 at three checked points. Halving the step to 300 s moves case 2’s souring from day 4.5069 to 4.5035, and the suggested challenge’s revenue from $428,198 to $428,254 with every verdict and the ranking unchanged. Sixty days of the challenge take 170 ms headless and 60.1 s of browser time at the fastest setting, 1 d/s. The application runs 28 checks of its own and 18 generic ones, and every case passes the same run loaded first, from the menu and after Reset.

The second kind, agreement with published results, is weaker. With the parameter set of Sari and Benyahia’s published case (from Benyahia and coauthors [5]), the engine’s equations reproduce their critical dilution of 0.818557467 d^{-1} and their $S_{2M} = 36.332$ mmol/L to round-off [6], and the integrated steady state at $D = 0.7$ d^{-1} equals their closed form. That is code verification against a published steady state, not validation. The acidogen constants are identified values from a fixed bed on another substrate, the methanogen constants are declared, and no digester’s data were fitted, so the model has not been validated against any operating digester. The third kind, educational evaluation, is absent. No study of learning has been run, so no gain in learning is claimed.

11.8 Technology in Practice

The simulated digester carries four hourly loops, a daily titration, a weekly feed analysis and an economic supervisor. Figure 11.6 draws them with the devices a farm plant would install.

What Measures It

The loading loop’s measurement is not an instrument on the tank but a laboratory result: the ratio of volatile acids to alkalinity from a sample drawn at 08:00. The chapter models it as the true acids and alkalinity of the sample hour, with 3 % and 1 % relative noise, reported 6 h later and held for a day. The online instruments serve the faster loops: the biogas meter, the two methane analyzers, the pH probe, the thermometer and the holder level, each with filtered noise.

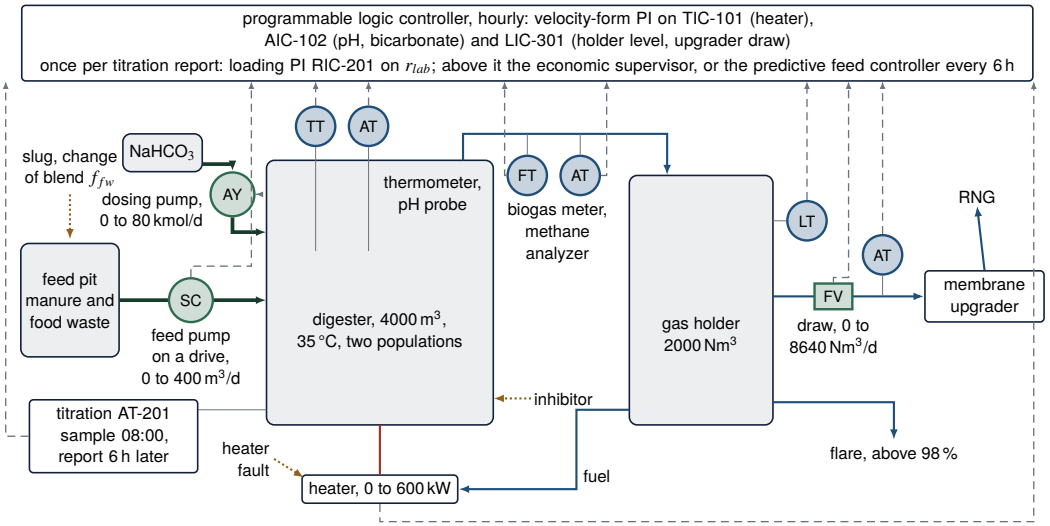


Figure 11.6 The chapter's loops with the devices a farm plant would install. Thick green lines carry the feed, blue lines the gas, dashed gray lines the signals and dotted amber arrows the disturbances. Notice that every instrument on the tank and on the gas is continuous except the one the loading loop acts on: the titration reports once a day, six hours after its sample.

> **TECHNOLOGY** **Sensor: Volatile Acids and Alkalinity by Titration**
>
> The ratio comes from two titrations of one sample. In the 1976 method a settled sample is titrated with sulfuric acid to pH 4.0, which gives the total alkalinity, then boiled lightly to drive off carbon dioxide and titrated with sodium hydroxide from pH 4.0 to 7.0, which gives the volatile acid alkalinity. The method was judged acceptable for digester control but not accurate enough for research [11]. Titration assays remain common for their low cost, while offline sampling introduces delays and limits the time resolution [22]. A sample taken just after feeding, or from a line not run clear, is not representative [11], and on a trend it shows as one report out of step with its neighbors. The chapter's 6 h report delay and its relative noise stand for the laboratory.

What Moves It

The feed pump sets the loading, a dosing pump adds bicarbonate, the heater duty holds the temperature and the upgrader's draw holds the holder level. In the model the feed is a flow between 0 and 400 m^3/d that arrives at once, so the delay of the loading loop lies wholly in the titration and in the biology.

> **TECHNOLOGY** **Final Element: Feed Pump on a Variable-Speed Drive**
>
> A positive-displacement pump discharges a fixed volume of fluid with every revolution of its shaft, so at a constant motor speed it delivers a fixed flow. A variable frequency drive sets the speed of an alternating-current motor by varying the frequency of its supply, and such drives are important final control devices in industry [23]. A blocked discharge stalls a positive-displacement pump and overloads its motor unless a relief path exists [23]. The 1976 manual found that the hydraulic overload of a digester is caused most often by leaving the wrong valve open or a pump on by accident [11]. On a trend that fault shows as a feed that does not match the request. The simulated pump delivers the requested feed at once, between 0 and 400 m^3/d.

Where the Controller Runs

On a farm plant these loops would run on a programmable logic controller or in a small distributed control system. A programmable logic controller holds instructions for input and output, logic, timing,

counting and three-mode PID control, and in a distributed control system it serves as a local controller within a supervisory scheme [24]. Supervisory and lower loops run continuously with cycle times from milliseconds to minutes [24]. The chapter's temperature, pH and level loops execute once an hour, and the loading loop once per titration report, so the platform's scan is never the limit. In a function block program the loading loop is a PI block whose input is a laboratory value entered once a day, and the supervisor that moves its setpoint, like the predictive feed controller that replans every six hours, belongs one layer above it.

What Is Modernizing It

Three lines of work aim at the day the titration costs. The first is online acid sensing. Microbial electrochemical sensors have measured volatile fatty acids with an accuracy comparable to gas chromatography, but electrochemical sensors in digestate are still limited by their longevity, reproducibility and biofouling, and need cleaning and recalibration [22]. The second is the soft sensor. Machine learning models have predicted the ratio of volatile acids to alkalinity in real time from the operating data of a pilot bio-electrochemical digester, the best of them a neural network with an explained variance of 0.77 [25]. The third is model-based feed control. A predictive controller has moved the feed rate to hold production against weather-driven temperature changes without controlling the temperature itself [18], and a supervisory policy has been learned offline on a calibrated digital twin of a full-scale digester [26]. Each would shorten or bridge the delay that sets the gain of the chapter's loading loop in case 4.

> ⚠ **IN PRACTICE** **Read the Ratio as a Trend**
>
> Operator training for wastewater plants reads the ratio in bands: 0.05 to 0.1 optimum, 0.3 to 0.4 stress, 0.5 to 0.7 deep trouble, 0.8 and above failure. It calls the acid-alkalinity relationship the first warning and the pH one of the last, and it notes that gas with more than 45 % carbon dioxide will not burn. A sour digester is cured by adding alkalinity, most often lime, or sodium bicarbonate at a higher cost [27]. The 1976 manual cautions the operator against an absolute number: the rate of change is much more significant, and trends of the indicators best predict an upset [11].

11.9 Limitations

All values describe a consistent generic digester, not a measured one. The methanogen kinetics are chosen to place the fold at 1.8 times the design, and the published constants span an order of magnitude. Hydrolysis is fast, the acids are one pool, and ammonia, sulfide and sodium are absent, so a real tank on manure has loading limits that this model does not have and a bicarbonate dose that this model treats as harmless. The steady state is set by the retention alone, which is what makes the feedforward lose, and a model with hydrolysis limitation or an ammonia balance could reward it. The ratio limit, the upgrader's limits and the prices are teaching values. The titration is the only acids measurement, so the online analyzers and direct online acid measurements that shorten its delay in practice are absent. The nominal MPC observer is an unvalidated forecast, and its soft ratio ceiling does not ensure stability. A simulated pass is not a guarantee of a real digester's stability.

OPEN CHALLENGE **Feeding Closer to the Edge on a Real Plant**

Full-scale digesters are often operated conservatively because their process feedback is delayed, their stability constraints are nonlinear and their state is hard to observe. A supervisory policy learned offline on a calibrated digital twin of one plant showed a potential gain of 2.7 % in biogas on a year of its operating data, with a probability of 0.524 that the gain is real [26]. A review of digester sensing asks for more reliable sensors, open datasets and interpretable models [22]. Progress would look like a feed supervisor proven on a live plant, with a gain that holds outside its own model.

11.10 Conclusion

An anaerobic digester is a plant in which more feed first looks like success, and this simulator shows why with the published two-step model, a carbonate charge balance, ADM1's pH inhibition, a heated tank, a holder, an upgrader and a flare. The fold sits at 287.6 m^3/d, 1.8 times the design, and is set by the retention, so a richer feed needs more methanogens rather than less volume, and a feedforward on the weekly COD costs $11,648 over the season. At 400 m^3/d the gas peaks on day 2.63, the titration warns on day 3.58 and the tank sours on day 4.51, after the upgrader has already tripped on off-spec gas on day 2.01, and recovery takes 45.3 days with bicarbonate. On a daily titration a loading gain of 8 settles in about three weeks while 30 swings the feed between its limits, and a supervisor that raises the setpoint toward 0.25 earns $428,198 in sixty days with no report over 0.4, against $166,751 for a constant feed that sours. The model is lumped, its methanogens are declared, and its only acids measurement is the titration. A fuller model would add hydrolysis, ammonia and sodium inhibition, an online acids analyzer and biological state identification for the nominal predictive feed controller. A digester is fed on its slowest measurement, and the simulator shows what that costs. The fifteen-day MPC earns more in ordinary ratio control but loses under hidden biological inhibition, so forecast horizon alone does not replace state information.

KEY POINTS

- More feed makes more gas for a day or two, and then the acids the slow methanogens cannot keep up with sour the tank for weeks.
- The fold sits at 287.6 m^3/d, 1.8 times the design, and is set by the retention, so a feedforward on the weekly COD costs $11,648 over the season.
- At 400 m^3/d the gas peaks on day 2.63, the titration warns on day 3.58 and the tank sours on day 4.51.
- On a daily titration a loading gain of 8 settles in about three weeks while 30 swings the feed between its limits.

Where else this applies. Volume 7, Chapter 2 has ignition and extinction folds of the same shape, Volume 5, Chapter 3 tunes a column against a slow analyzer, and Volume 5, Chapter 4 grades a cook on late laboratory results.

References

[1] J. F. Andrews, "A mathematical model for the continuous culture of microorganisms utilizing inhibitory substrates," *Biotechnology and Bioengineering*, vol. 10, no. 6, pp. 707–723, 1968.

[2] J. Liu, G. Olsson, and B. Mattiasson, "Extremum-seeking with variable gain control for intensifying biogas production in anaerobic fermentation," *Water Science and Technology*, vol. 53, no. 4–5, pp. 35–44, 2006.

[3] T. Sari, "Best operating conditions for biogas production in some simple anaerobic digestion models," *Processes*, vol. 10, no. 2, p. 258, 2022.

[4] United States Environmental Protection Agency, "AgSTAR data and trends," AgSTAR program web page, last updated March 10, 2026, 2026. [Online]. Available: https://www.epa.gov/agstar/agstar-data-and-trends

[5] B. Benyahia, T. Sari, B. Cherki, and J. Harmand, "Bifurcation and stability analysis of a two step model for monitoring anaerobic digestion processes," *Journal of Process Control*, vol. 22, no. 6, pp. 1008–1019, 2012.

[6] T. Sari and B. Benyahia, "The operating diagram for a two-step anaerobic digestion model," *Nonlinear Dynamics*, vol. 105, no. 3, pp. 2711–2737, 2021.

[7] B. Palacios-Ruiz, H. O. Méndez-Acosta, V. Alcaraz-González, V. González-Álvarez, and C. Pelayo-Ortiz, "Regulation of volatile fatty acids and total alkalinity in anaerobic digesters," *IFAC Proceedings Volumes*, vol. 41, no. 2, pp. 13 611–13 616, 2008.

[8] D. Gaida, C. Wolf, and M. Bongards, "Feed control of anaerobic digestion processes for renewable energy production: A review," *Renewable and Sustainable Energy Reviews*, vol. 68, pp. 869–875, 2017.

[9] J. F. Andrews, "Dynamic model of the anaerobic digestion process," *Journal of the Sanitary Engineering Division*, vol. 95, no. 1, pp. 95–116, 1969.

[10] J. F. Andrews and S. P. Graef, "Dynamic modeling and simulation of the anaerobic digestion process," in *Anaerobic Biological Treatment Processes*, ser. Advances in Chemistry. American Chemical Society, 1971, pp. 126–162.

[11] C. Zickefoose and R. B. J. Hayes, "Operations manual: Anaerobic sludge digestion," United States Environmental Protection Agency, Office of Water Program Operations, Washington, DC, Tech. Rep. EPA 430/9-76-001, 1976. [Online]. Available: https://permanent.fdlp.gov/gpo81586/00000IKT.pdf

[12] D. T. Hill, S. A. Cobb, and J. P. Bolte, "Using volatile fatty acid relationships to predict anaerobic digester failure," *Transactions of the ASAE*, vol. 30, no. 2, pp. 496–501, 1987.

[13] O. Bernard, Z. Hadj-Sadok, D. Dochain, A. Genovesi, and J.-P. Steyer, "Dynamical model development and parameter identification for an anaerobic wastewater treatment process," *Biotechnology and Bioengineering*, vol. 75, no. 4, pp. 424–438, 2001.

[14] D. J. Batstone, J. Keller, I. Angelidaki, S. V. Kalyuzhnyi, S. G. Pavlostathis, A. Rozzi, W. T. M. Sanders, H. Siegrist, and V. A. Vavilin, "The IWA Anaerobic Digestion Model No 1 (ADM1)," *Water Science and Technology*, vol. 45, no. 10, pp. 65–73, 2002.

[15] J. P. Steyer, O. Bernard, D. J. Batstone, and I. Angelidaki, "Lessons learnt from 15 years of ICA in anaerobic digesters," *Water Science and Technology*, vol. 53, no. 4–5, pp. 25–33, 2006.

[16] S. Attar and F. Haugen, "Dynamic model adaptation to an anaerobic digestion reactor of a water resource recovery facility," *Modeling, Identification and Control*, vol. 40, no. 3, pp. 143–160, 2019.

[17] H. O. Méndez-Acosta, B. Palacios-Ruiz, V. Alcaraz-González, V. González-Álvarez, and J. P. García-Sandoval, "A robust control scheme to improve the stability of anaerobic digestion processes," *Journal of Process Control*, vol. 20, no. 4, pp. 375–383, 2010.

[18] A. Moradvandi, S. Heegstra, P. Ceron-Chafla, B. De Schutter, E. Abraham, and R. E. F. Lindeboom, "Model predictive control of feed rate for stabilizing and enhancing biogas production in anaerobic digestion under meteorological fluctuations," *Journal of Process Control*, vol. 147, p. 103375, 2025.

[19] I. Angelidaki, L. Treu, P. Tsapekos, G. Luo, S. Campanaro, H. Wenzel, and P. G. Kougias, "Biogas upgrading and utilization: Current status and perspectives," *Biotechnology Advances*, vol. 36, no. 2, pp. 452–466, 2018.

[20] S. J. Qin and T. A. Badgwell, "A survey of industrial model predictive control technology," *Control Engineering Practice*, vol. 11, no. 7, pp. 733–764, 2003.

[21] G. Pannocchia and J. B. Rawlings, "Disturbance models for offset-free model-predictive control," *AIChE Journal*, vol. 49, no. 2, pp. 426–437, 2003.

[22] M. Marycz, I. Turowska, S. Glazik, and P. Jasiński, "Artificial intelligence in anaerobic digestion: A review of sensors, modeling approaches, and optimization strategies," *Sensors*, vol. 25, no. 22, p. 6961, 2025.

[23] T. R. Kuphaldt, *Lessons in Industrial Instrumentation*, version 2.33 ed. Open textbook, Creative Commons Attribution 4.0 license, ibiblio.org/kuphaldt/socratic/sinst, 2024.

[24] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule, and M. Thompson, "Guide to operational technology (OT) security," National Institute of Standards and Technology, Tech. Rep. NIST Special Publication 800-82r3, 2023.

[25] H. Rutland, J. You, H. Liu, and K. Bowman, "Application of machine learning for FOS/TAC soft sensing in bio-electrochemical anaerobic digestion," *Molecules*, vol. 30, no. 5, p. 1092, 2025.

[26] A. I. Yunus, S. Jalla, J. F. Bozeman, and Y. Chen, "Reinforcement-learning-enabled supervisory control of a full-scale anaerobic digestion facility using a calibrated digital twin," *Environmental Science & Technology*, vol. 60, no. 39, pp. 28 064–28 078, 2026.

[27] Pennsylvania Department of Environmental Protection, "Wastewater treatment plant operator certification training, module 6: Solids handling and disposal," Pennsylvania Department of Environmental Protection, Bureau of Water Supply and Wastewater Management, Training module, 2009. [Online]. Available: https://files.dep.state.pa.us/water/BSDW/OperatorCertification/TrainingModules/ww06_solids_handling_wb.pdf

11.11 Exercises

11.1 Concept check · 2 min

In case 2 the feed rises from 160 to 400 m^3/d, beyond the fold. The biogas doubles to 9504 Nm^3/d on day 2.63 while the methane rises by only 41 %. What makes up the rest of the rise?

(a) The acidogens grow within hours and turn the extra feed straight into methane

(b) The heater burns less of the holder's gas, so more of it passes the biogas meter

(c) The higher feed strips dissolved methane out of the liquid into the headspace

(d) The acids drive carbon dioxide out of the bicarbonate, and it leaves with the gas

11.2 Concept check · 1 min

A digester at 250 m^3/d is moved slowly from 12 to 25 % food waste, slowly enough that its methanogens grow into the richer feed. Where do its steady acids settle?

(a) At the same 25.184 mmol/L, because the steady acids depend on the dilution rate alone

(b) Higher in proportion to the feed COD, because a richer feed brings more acids each day

(c) Lower, because the richer feed grows more methanogens than this dilution rate needs

(d) Nowhere, because the loading now exceeds the fold loading of the design blend

11.3 Concept check · 2 min

In case 5 the feed rises to 300 m^3/d, just past the fold, and the pH loop holds 7.2 with bicarbonate. The tank sours on day 18.85 instead of day 10.22. Why does it sour at all?

(a) Bicarbonate removes the acids only slowly, at one mole of acid per mole dosed

(b) Past the fold the Haldane term inhibits the methanogens whatever the pH

(c) The sodium the bicarbonate adds builds up and inhibits the methanogens

(d) The pH probe drifts high, so the true pH sits below the loop's setpoint

11.4 Theory and calculation · 6 min

The methanogens of the chapter have $\mu_{2,max} = 0.164\,\mathrm{d}^{-1}$, $K_{S2} = 20\,\mathrm{mmol/L}$, $K_{I2} = 130\,\mathrm{mmol/L}$ and $k_d = 0.02\,\mathrm{d}^{-1}$. The tank holds 4000 m^3, keeps no biomass back ($\alpha = 1$) and is designed for 160 m^3/d. Take $T = 35\,^{\circ}\mathrm{C}$ and the pH factor as 1.

(a) Find the acids S_{2M} at the peak of the Haldane curve and the peak growth rate. *Checkpoint:* 51.0 mmol/L and 0.0919 d^{-1}.

(b) Find the critical dilution rate, the feed and retention it corresponds to, and its ratio to the design feed. *Checkpoint:* 0.0719 d^{-1}, 287.6 m^3/d, 13.9 days and 1.80.

(c) Find both steady acids at the design feed and say which one the tank sits at. *Checkpoint:* 12.2 and 213 mmol/L; the lower one.

11.5 Theory and calculation · 8 min

The tank of case 7 is fed 250 m^3/d at 15 °C and loses heat through a shell of $UA = 2.0\,\mathrm{kW/K}$ to 5 °C outdoors, with $\rho c_p = 4.18\,\mathrm{MJ/(m^3\,K)}$ and $V = 4000\,\mathrm{m^3}$. The heater trips at 35.0 °C and stays off for six days. Use Equation (11.8).

(a) Find the heater duty that holds the tank at 35 °C at this feed. *Checkpoint:* 302 kW.

(b) With the heater off, find the conductance G, the time constant, the temperature the tank cools toward and its temperature after six days. *Checkpoint:* 14.09 kW/K, 13.73 d, 13.58 °C and 27.4 °C.

(c) With $\theta_2 = 1.07$ and the methanogen constants of the previous exercise, estimate the fold at that temperature, taking the pH factor as 1. *Checkpoint:* About 140 m^3/d.

11.6 Theory and calculation · 6 min

The titration reports acids in mg/L as acetic acid over alkalinity in mg/L as calcium carbonate, Equation (11.9). The design state has $S_2 = 12.20\,\mathrm{mmol/L}$ and $Z = 154.4\,\mathrm{mmol/L}$.

(a) Find the design ratio. *Checkpoint:* 0.095.

(b) At the design alkalinity, find the acids at which the ratio reaches the limit of 0.4, in mg/L and in mmol/L, and compare them with S_{2M}. *Checkpoint:* 3090 mg/L, 51.5 mmol/L, against 51.0 mmol/L.

(c) The loading loop runs with $K_c = 8$ per unit of ratio on the nameplate feed COD of $54.4\,\mathrm{kg\,COD/m^3}$. Find the proportional part of its step, in loading and in feed, when a report comes in 0.05 above the setpoint, and the same at $K_c = 30$. *Checkpoint:* $0.4\,\mathrm{kg\,COD/m^3/d}$ or $29\,\mathrm{m^3/d}$; $1.5\,\mathrm{kg\,COD/m^3/d}$ or $110\,\mathrm{m^3/d}$.

11.7 App challenge · 9 min

Load case 2, The overload, and change no setting. Run at the default speed of 0.25 d/s with the default seed until the run pauses at the souring, about 18 s, and export the CSV file.

(a) Find the largest true biogas, column `truth_biogas_Nm3d`, and the day of that row. *Checkpoint:* $9504\,\mathrm{Nm^3/d}$ on day 2.63, within $20\,\mathrm{Nm^3/d}$.

(b) Find the first titration report above 0.4 in column `ratio_lab`: its day, its value, and how long after the gas peak it arrives. *Checkpoint:* Day 3.58, 0.55, about 0.95 days later.

(c) Find the first row in which `upgrader_on` reads 0, and the day the souring tile names. Explain why the upgrader stopped first. *Checkpoint:* Day 2.04 (the trip at day 2.01); soured day 4.5.

11.8 App challenge · 8 min

Load case 4, The ratio loop on the daily titration, and run 40 days at the default seed at 1 d/s (40 s). Export the CSV file.

(a) Read the feed, column `feed_m3d`, at days 20 and 40, and the largest report in column `ratio_lab`. *Checkpoint:* 233 and $248\,\mathrm{m^3/d}$, within 2; 0.207, within 0.005.

(b) Set Loading gain to 30, press Reset, run 40 days again and export. The first two reports come on days 0.58 and 1.58. Read the feed in the row after each, at days 0.625 and 1.625, since the row at a report still holds the old feed. *Checkpoint:* 390 and $69\,\mathrm{m^3/d}$, within 2.

(c) Explain the swing with the step size of the previous exercise and the timing of the reports. *Checkpoint:* Each step is decided on acids that have not yet answered the one before.

11.9 App challenge · 8 min

Load case 7, The heater fault, with the default seed.

(a) Run without acting until the run pauses at the souring. Read the lowest true temperature, column `truth_T_C`, and its day, the day of the first report above 0.4, and the souring day on the tile. *Checkpoint:* 27.4 °C on day 8.42, report on day 7.58, soured day 10.6.

(b) Press Reset. Run at 1 d/s to day 7, change the speed to 0.05 d/s, and at the report of day 7.58 cut Feed rate to $160\,\mathrm{m^3/d}$. Return to 1 d/s and run to day 25. Read the titrations tile. *Checkpoint:* Not soured, four reports over 0.4 if the cut is made before day 7.8.

(c) The previous exercise put the cold fold near $140\,\mathrm{m^3/d}$. Explain why a feed of $160\,\mathrm{m^3/d}$ saves the tank. *Checkpoint:* The excursion past the fold is short.

11.10 Application area · 3 min

Published guidance keeps a healthy digester between 0.1 and 0.35 on the acids-to-alkalinity ratio. The chapter's steady ratio along the stable branch is 0.259, 0.303, 0.328 and 0.396 at 95, 98, 99 and 100 % of the critical feed.

(a) What share of the critical feed does the supervisor's back-off ratio of 0.30 correspond to in this model? *Checkpoint:* About 98 %.

(b) Why should a real plant take its ratio limit from its own titration history rather than from this model? *Checkpoint:* The mapping from ratio to margin depends on the kinetics and the alkalinity.

11.11 Application area · 3 min

The upgraded gas is sold at \$3 per MMBtu with a \$25 federal and a \$20 state credit, at 0.037 74 MMBtu per Nm^3. Sodium bicarbonate costs \$0.40 per kg and weighs 84.0 kg/kmol.

(a) Find the revenue per Nm^3 of renewable natural gas and the share of it that the credits pay. *Checkpoint:* \$1.81 per Nm^3; 94 %.

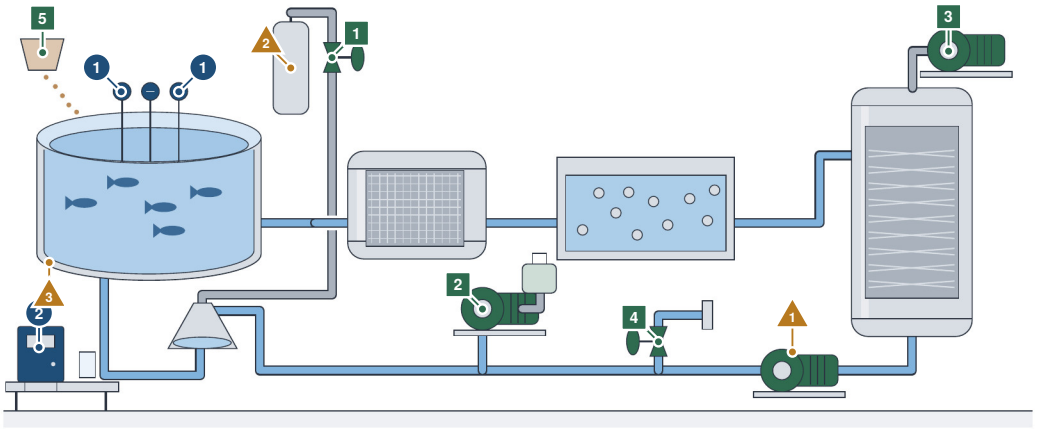
(b) The recovery of case 3 takes 1227 kmol of bicarbonate. Find its mass and cost, and compare the cost with the \$3,528 the supervisor of case 8 earns over 40 days. *Checkpoint:* 103 t, about \$41,200, about 12 times as much.

11.12 Review the agent · 5 min

An AI assistant is asked how to raise the digester's revenue and replies as follows. "The titration is a day late but the biogas meter answers within hours. Close the feed loop on the gas instead: raise the feed by 5 % each day the metered biogas has risen since the day before, and cut it by 5 % each day it has fallen. The climb stops by itself at the feed that makes the most gas." Find the flaw in the proposal and say how it could be checked.

CHAPTER 12

Recirculating Fish Farm Simulator Where Nitrite Answers the Feed Days Later



The controller moves 1 oxygen gas valve 2 bicarbonate dosing 3 degasser blower 4 makeup water valve 5 feeder, the daily ration

The controller reads 1 tank probes: oxygen, pH, carbon dioxide 2 sampled nitrogen analyzer

What disturbs it 1 power failure at the pump 2 oxygen supply sag 3 warmer surroundings

Notice that the probes read the tank's oxygen at once, while the nitrogen the biofilter must remove reaches the controller only through a sample tested every few hours. Illustration, not to scale.

IN PLAIN TERMS **Feeding Fish Faster Than Their Filter Can Grow**

The main idea. A fish farm that reuses its water needs a filter full of living bacteria to clean out the fish waste. Two kinds of bacteria do the job in turn, and each kind grows only on the food it gets.

The challenge. When the farm feeds more, the first kind keeps up within a day. The second kind needs days to grow, so a poison builds up in between, and the water is tested only every few hours.

Why it matters. Feeding faster than the bacteria can grow kills fish instead of growing them.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the fact on world fish farming (Section 12.1), the two history segments (Section 12.2), the two try-it segments (Section 12.4), the fact of Section 12.6, and the key points
Study	45 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	15 min	the first page, the measured cases (Section 12.6), Technology in Practice (Section 12.7), Limitations (Section 12.8) and the open challenge

The control question A recirculating farm removes its fish's ammonia with a biofilter of two nitrifier populations that grow only on the load they carry, so a larger ration returns as nitrite days after the feed changes. How can a farm that measures its water every four hours and weighs its fish every two weeks feed for growth without poisoning the fish?

Learning objectives

- Explain why a feed step or a new biofilter shows an ammonia peak first and a nitrite peak days later.
- Compute the un-ionized share of the total ammonia from pH and temperature, and the total ammonia that reaches its limit.
- Estimate the minutes of oxygen reserve after a power failure from the tank volume and the fish's uptake.
- Describe what a feed governor on a sampled analyzer can and cannot prevent, and how a soft sensor infers biomass from the oxygen balance.
- Compare rations by feed conversion and margin, and recognize a stocking plan that no ration makes feasible.

Prerequisites Sections 1.4.3 (populations that grow on their load), 1.4.5 (sparse and late samples, bias update), 1.4.6 (overrides on limits) and 1.6.2 (velocity form). Chapter 11 for a culture that is the actuator.

The app https://apmonitor.com/apps/aquaculture-ras/

See also Volume 3, Chapter 2 and Volume 3, Chapter 3 for cascade structure and the speed an oxygen inner loop needs. Volume 5, Chapter 3 for control through a slow analyzer and Volume 5, Chapter 10 for a soft sensor whose bias a delayed laboratory corrects.

12.1 Introduction

Ten thousand rainbow trout of 150 g eat four meals a day in a 50 m^3 tank, and the oxygen they draw from a fixed gas flow of 0.50 kg/h falls after each meal to a lower low than the one before, from 9.99 mg/L at 11:00 to 7.63 mg/L at 21:40, two hours and ten minutes after the last meal ends. That is the fast half of the farm. The slow half is the biofilter. When the same farm triples a ration that its biofilter had grown to carry, the total ammonia peaks at 4.34 mg/L that midnight, but the nitrite peaks at 3.07 mg N/L 1.35 days after the step and stays above its 0.5 mg N/L teaching limit for 57.0 h. The nitrite comes late because the bacteria that remove it must first grow on it.

Recirculating systems dissolve pure oxygen, strip carbon dioxide and convert ammonia in a biological filter so that the same water can carry fish many times a day, and control of the dissolved gases is an essential part of such systems [1]. The water quality a reuse system must hold has been set out for each constituent [2], the nitrifying filters that remove the ammonia have been reviewed for their design and operation [3], and flows and filters are sized by mass balances on the feed [4]. Dynamic models that couple the fish to the water treatment have been written for recirculating systems [5], and precision fish farming proposes to apply control-engineering principles to fish production [6]. The broader class is any plant whose actuator is a culture with its own growth rate, so that the capacity to act is itself a slow state. The wastewater treatment that integrated recirculating models couple to the fish [5] and anaerobic digestion have that shape. Volume 9, Chapter 10 treats the nitrifiers of a wastewater works as the plant, and Chapter 11 treats a digester's methanogens as a culture that can be lost.

> **DID YOU KNOW? Farmed Fish Overtook the Catch**
>
> In 2022, for the first time, aquaculture produced more aquatic animals than capture fisheries: 94.4 million tonnes, 51 % of the world total. Inland farms, the setting of this chapter's freshwater tank, gave 62.6 % of the farmed animals [7].

A common belief, and the one this plant is meant to unsettle, holds that a biofilter is a treatment unit with a capacity fixed at design, so that the farm can feed what the fish will eat and leave the water to the equipment. In this model a mature film carries the load it was grown on and little more, a

new film takes weeks to grow from a trace, and the nitrite that a ration beyond the film's capacity makes acts on the fish. However, the sources a student is likely to read describe each piece on its own. Toxicity is reported as lethal concentrations from separate tests [8, 9], filter sizing is a steady balance [4], and growth is a curve against degree days [10]. This chapter works with a recirculating farm, simulated in a browser, in which the meal, the biofilter's growth, the fish's response to nitrite and the farm's sampled measurements are integrated together, and it measures what a cascade, a governor, a soft sensor and a daily economic ration achieve on the same plant. The growth law, the nitrification kinetics, the carbonate chemistry, the proportional integral loops and the golden-section search are established methods. The simulator built from them is a 41-state teaching model whose nitrogen and oxygen ledgers close to round-off, with an observer that follows each nitrifier population from the laboratory's conversion and a strict separation in which no controller or estimator reads a simulated truth. The app carries ten graded cases running from a day of meals to an infeasible stocking plan and a ninety-day challenge.

Section 12.2 follows the domain, from the first reuse farms to the toxicity data behind the limits. Section 12.3 sets out the farm as modeled: the cohort and its ration, the water's action on the fish, the balances in two volumes, the two nitrifier populations, the cone, the degasser, the chemistry and the instruments. Section 12.4 shows the farm's two clocks, the meal in hours and the biofilter in weeks, and the chemistry that sets how much of the ammonia is toxic. Section 12.5 describes the oxygen cascade, the pH, carbon dioxide and nitrate loops, the feed governor, the economic ration, the two estimators and the startup sequence. Section 12.6 measures the cases and the challenge and verifies the code, Section 12.7 sets the loops beside the devices a real farm installs, Section 12.8 states the limitations and Section 12.9 concludes.

12.2 Reused Water and the Farms Built Around It

Intensive fish culture that reconditions its water instead of discharging it was described in 1974 as a way to combine production with pollution abatement [11]. The engineering that followed treated each constituent the fish add to the water as a separate unit operation. Oxygen is dissolved from pure gas in a cone or a column, carbon dioxide is stripped in a packed degasser for which design procedures have been published [1], solids leave in a drum or a settler, and ammonia is oxidized in a biofilter. By 2000 the flow and filter sizes of such a farm were routine mass balances on the daily feed, written in spreadsheets [4], and the same volume of the journal that published those balances reviewed process control and artificial intelligence software for aquaculture [12]. Design reviews of nitrifying filters for recirculating systems [3] and of the water quality reuse systems must hold [2] appeared in 2006, and new European systems were reviewed from the perspective of environmental sustainability in 2010 [13].

> **HISTORY** **Oxygen Pushed Into Water by Depth (1969)**
>
> In 1969 a federal fish hatchery had run a U-tube oxygenator continuously for over a year. The water is directed down a chamber about forty feet deep, under a baffle and back up, and air bubbles injected just below the surface are dragged down with it. The hydrostatic head pressurizes the bubbles, so one pass can saturate or supersaturate the water. A free fall of three feet entrained enough air to saturate 300 gallons per minute with no external power [14]. Farms today reach the pressure with a pump instead: an oxygen cone mixes water and pure oxygen at typically around 1.4 bar and supersaturates part of the circulating flow [15].

The toxicity data behind the limits came from fisheries research. Ninety-six-hour tests on rainbow trout in soft water gave median lethal concentrations of nitrite from 0.19 to 0.39 mg N/L [8]. A later review found that each milligram per liter of chloride raises the 96-hour LC50 by 0.29 to 2.0 mg N/L, that chronic effects are difficult to detect below one fifth of the 96-hour value, and that the ratio of the 24-hour to the 96-hour value has a median of 2.0 [9]. Nitrite accumulates in fish over 24 to 48 hours, and a tenth of the 96-hour LC50 caused neither growth suppression nor tissue damage

in long exposures [16]. Ammonia is toxic mainly as its un-ionized form, whose share follows the ammonium dissociation constant as a function of temperature [17, 18], and the national freshwater criteria tabulate rainbow trout among the acute data [18]. Growth on the farm was reduced to a single coefficient when the cube root of weight was found to rise linearly with degree days in hatcheries [10].

> ⧖ HISTORY **Salt Against Nitrite (1977 to 1984)**
>
> Since 1977 nitrite toxicity has been known to depend greatly on the salinity of the water: fish in seawater died at nitrite concentrations 50 to 100 times higher than in fresh water. The effect of chloride is so large that a test which does not report its chloride cannot be meaningfully compared with others. A recommendation of 1984 asked aquaculture to monitor the ratio of chloride to nitrite nitrogen, 17 for salmon and 8 for rough fish [16]. Farm guidance still adds a little salt to the water to take the toxicity out of the nitrite peak of a new biofilter [15].

What these sources do not give is the time course. A balance sizes the filter for a steady load and a lethal concentration describes a test at fixed exposure, while the farm changes its ration every day and its bacteria take weeks to follow. Integrated dynamic models of the fish and the treatment answer that question for a designed plant [5], and the precision farming program asks for the measurement, estimation and feedback that would act on it [6]. The simulator described next is a teaching plant of that kind.

12.3 The Farm as Modeled

The simulated farm is a generic land-based unit with no proprietary source, drawn in Figure 12.1 in the order the water travels. Water leaves an oxygen cone, passes the fish tank of volume $V_t = 50\,\mathrm{m}^3$, a drum filter that removes solids, a moving-bed biofilter of water volume $V_b = 10\,\mathrm{m}^3$ and a degasser over its sump, and a pump returns it to the cone at $Q_r = 100\,\mathrm{m}^3/\mathrm{h}$, so the tank turns over every 30 min. Makeup water Q_{ex} enters the tank and the same flow leaves with the solids. Sodium bicarbonate is dosed into the tank, and an oxygen gas valve of opening u feeds the cone from a supply at relative pressure P_s. The cohort is N fish of mean weight W and biomass B, fed by a feeder that runs four 30 min meals at 07:00, 11:00, 15:00 and 19:00. The model integrates in days with a step of 60 s, so every meal edge falls on a step.

Table 12.1 classifies the variables. Forty-one quantities are integrated: the biomass and count, two gut stages, oxygen, total ammonia nitrogen, nitrite, nitrate, inorganic carbon C_T and alkalinity in both the tank and the biofilter, the two populations, the gas flow, the water temperature, the fish's toxic load E, and twenty ledgers that integrate feed, nitrogen, oxygen, gas, water, bicarbonate and the fish lost, so that every balance can be checked exactly. Concentrations are in mg/L, with ammonia, nitrite and nitrate as nitrogen, flows in m^3/d, feed in kg/d and the populations in grams of volatile solids.

12.3.1 The Cohort and Its Ration

Growth at satiation follows the thermal-growth coefficient, the slope of the cube root of weight against degree days [10], and growth below satiation follows a concave ration curve, as shown in Equation (12.1).

$$
\begin{aligned}
g_{max} &= 3W^{2/3}\,\frac{TGC}{1000}\,(T - T_0), \qquad R_{sat} = FCR_{sat}\,g_{max} \\
e(\rho) &= e_{max}(1 - c\rho), \qquad M = \rho_m R_{sat}\,e_{max}(1 - c\rho_m) \\
\frac{\mathrm{d}B}{\mathrm{d}t} &= q_f\,e(\rho)\,p - \frac{NM}{1000} - (m+h)B \\
\frac{\mathrm{d}N}{\mathrm{d}t} &= -(m+h)N
\end{aligned}
\tag{12.1}
$$

where g_{max} and R_{sat} are per fish in grams per day, $\rho = \min(a, R/R_{sat,C})$ is the day's ration over the cohort's satiation $R_{sat,C} = NR_{sat}/1000$, capped by the appetite a, p is the feed processed, M the

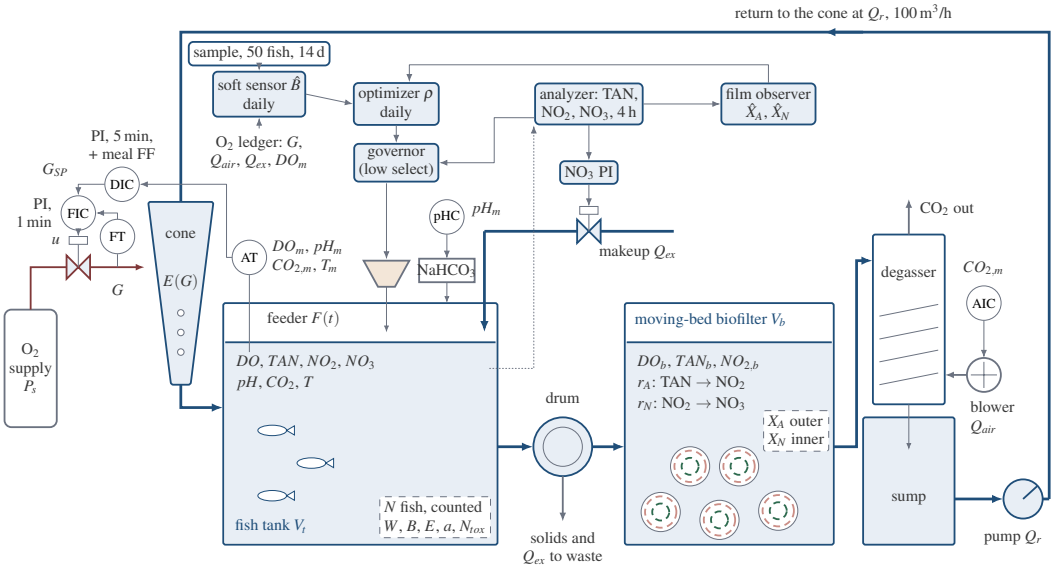


Figure 12.1 The recirculating farm as modeled, in the order of the app's Farm view. Gas at G through the valve u dissolves in the cone with absorption $E(G)$, the tank holds the cohort (N, W, B) and its water, the drum takes the solids and Q_{ex} to waste, the moving-bed biofilter holds its own water and the two populations (ammonia oxidizers X_A outer, nitrite oxidizers X_N inner), and the degasser strips CO_2 with blower air Q_{air} before the pump returns Q_r. The DO cascade (DIC over FIC), the pH loop (pHC) and the CO_2 loop (AIC) sit beside their actuators, and the supervisory layer (analyzer, governor, optimizer, soft sensor, film observer, nitrate loop) sits above the row. Process model only, not a construction drawing. Not to scale.

Table 12.1 Variables by Role, With Their Values at $t = 0$ of the Challenge

Variable	Role	At $t = 0$	Range or limit
DO	CV, tank oxygen	9.02 mg/L	SP 9, limit ≥ 6
pH	CV, tank	7.197	SP 7.2, limits 6.8–7.8
CO_2	CV, tank	9.80 mg/L	SP 10, limit ≤ 15
NO_3	CV, laboratory	59.0 mg N/L	SP 60
NO_2	constraint	0.106 mg N/L	limit ≤ 0.5
NH_3	constraint, un-ionized	0.0030 mg N/L	limit ≤ 0.025
u, G	MV, gas valve, flow	17.7 %, 0.286 kg/h	0–1.6 kg/h
u_{bic}	MV, bicarbonate	2.90 kg/d	0–150
Q_{air}	MV, blower	3.07 m³/min	1–20
Q_{ex}	MV, makeup	8.39 m³/d	2–60
R	MV, daily ration	13.44 kg/d	0–100
T_{env}, P_s	DV, surroundings, supply	15 °C, 1	8–20, 0.5–1.1
B, N, W	states, cohort	500 kg, 10,000, 50 g	truth (count known)
X_A, X_N	states, populations	481.9, 160.6 g	truth
TAN	state, tank	0.704 mg N/L	—
E, a	toxic load, appetite	0, 1	truth

maintenance, q_f the feed quality, m the background mortality and h the mortality the water causes. The constant $e_{max} = 1/\{FCR_{sat}[(1-c)-\rho_m(1-c\rho_m)]\}$ makes $\rho = 1$ grow exactly on the coefficient curve and convert at FCR_{sat}, and growth stops at ρ_m. With $c = 0.3$ the lowest conversion, 1.0068, comes at $\rho = 0.569$, which is 8.47 % better than satiation. Eaten feed F_{eat} passes two gut stages of $\tau_g = 3$ h, $\dot{g}_1 = F_{eat} - g_1/\tau_g$ and $\dot{g}_2 = (g_1 - g_2)/\tau_g$ with $p = g_2/\tau_g$, and the feed offered beyond the

appetite is wasted. After the 07:00 meal the processed feed peaks 2.93 h after the meal starts.

12.3.2 The Water Acting on the Fish

Nitrite and un-ionized ammonia act on the fish through one internal load, as shown in Equation (12.2).

$$\begin{aligned} x &= \max\Big(\frac{NO_2}{LC50_{NO_2}},\ \frac{NH_3}{LC50_{NH_3}}\Big), \qquad \tau_E \frac{\mathrm{d}E}{\mathrm{d}t} = x - E \\ s &= \max\Big(0,\ \frac{E - a_0}{1 - a_0}\Big), \qquad a = \max(0,\ 1 - s) \\ h &= s\,\frac{\ln 2}{4\,\mathrm{d}} \end{aligned} \tag{12.2}$$

where x is the water's toxicity in units of each toxicant's 96-hour LC50, $\tau_E = 1\,\mathrm{d}$ and $a_0 = 0.1$. The limits are a tenth of the LC50s, the no-effect level of [16], and τ_E gives 63 % of the load in a day and 86 % in two, in the range over which nitrite accumulates [16]. The larger toxic unit counts rather than the sum, because nitrite acts on the blood and ammonia on the gills and nerves [16, 18]. The LC50s, 5 and 0.25 mg N/L, are teaching values ten times the limits. The nitrite value is the trout LC50 of [8] raised by 0.29 mg N/L per mg/L of chloride [9] at about 16 mg/L of chloride, which the model assumes. The ammonia value is near the rainbow trout species mean of the national criteria, 82.88 mg TAN/L at pH 7, which is 0.226 mg N/L as un-ionized nitrogen at 15 °C [18]. Because the load takes a day to build, water held at one LC50 from clean kills 39.7 % of the fish in 96 h, so the law's own 96-hour LC50 is 1.324 times the declared one, and its 24-hour to 96-hour ratio of 7.60 is steeper than the median of 2.0.

12.3.3 Nitrogen, Oxygen and Carbon in Two Volumes

The ammonia excreted is the absorbed feed nitrogen less the nitrogen retained in the net growth of Equation (12.1), $\dot{n} = pN_f(1 - f_{fec}) - n_f\,(q_f e(\rho)p - NM/1000)$ with $N_f = 0.16\,PC$, and a share λ of the fecal and uneaten nitrogen leaches from the solids. At a ration of 0.9 of satiation and 150 g this is 0.0360 kg of ammonia nitrogen per kilogram of feed. The tank and the biofilter are each well mixed, as shown in Equation (12.3) for the ammonia, with nitrite, nitrate, oxygen, C_T and alkalinity balanced the same way.

$$\begin{aligned} V_t \frac{\mathrm{d}TAN}{\mathrm{d}t} &= Q_r(TAN_b - TAN) - Q_{ex}TAN \\ &\quad + \dot{n} + \lambda \dot{n}_s \\ V_b \frac{\mathrm{d}TAN_b}{\mathrm{d}t} &= Q_r(TAN - TAN_b) - r_A \\ V_b \frac{\mathrm{d}NO_{2,b}}{\mathrm{d}t} &= Q_r(NO_2 - NO_{2,b}) + r_A - r_N \\ V_t \frac{\mathrm{d}DO}{\mathrm{d}t} &= Q_r(DO_{ret} - DO) + Q_{ex}(DO_{mk} - DO) \\ &\quad - \dot{O}_f \end{aligned} \tag{12.3}$$

where r_A and r_N are the two oxidation rates in the biofilter's own water, $\dot{n}_s$ the nitrogen in feces and uneaten feed, and $\dot{O}_f$ the fish's oxygen uptake. During a startup the dosed ammonium chloride enters the tank with $\dot{n}$. The uptake is routine respiration plus the post-meal demand, $\dot{O}_f = \phi(DO)\,[Bs_0(W/100)^{bo}\,\theta_O^{T-15} \cdot 0.024 + 1000\,k_{SDA}\,p]$ in grams per day, with $\phi(DO) = 1 - e^{-DO/0.8}$ so that uptake falls only near anoxia. The filter's water loses $3.43r_A + 1.14r_N$ of oxygen and the heterotrophs on the fines $k_{het}\,p$, nitrification removes two equivalents of alkalinity per mole of nitrogen, excreted ammonia adds one, and the fish add carbon at a respiratory quotient of 0.9.

12.3.4 Two Nitrifier Populations

The biofilter's capacity is the two populations of Equation (12.4), with Monod terms in the filter's own substrate and oxygen, the uptake that grows with its population of Section 1.4.3.

$$\begin{aligned} r_A &= \frac{\mu_A}{Y_A}\,\frac{X_A}{1+X_A/X_{h,A}}\,\frac{TAN_b}{K_A+TAN_b}\,\frac{DO_b}{K_{OA}+DO_b} \\ r_N &= \frac{\mu_N}{Y_N}\,\frac{X_N}{1+X_N/X_{h,N}}\,\frac{NO_{2,b}}{K_N+NO_{2,b}} \\ &\quad\times\frac{DO_b}{K_{ON}+DO_b}\,\frac{K_I}{K_I+FA} \\ \frac{\mathrm{d}X_A}{\mathrm{d}t} &= Y_A r_A-(b_A+k_{det})X_A \\ \frac{\mathrm{d}X_N}{\mathrm{d}t} &= Y_N r_N-(b_N+k_{det})X_N \end{aligned} \tag{12.4}$$

where $\mu=\mu_{20}\theta^{T-20}$, $b=b_{20}\theta_b^{T-20}$, FA is the free ammonia in the filter, and X_h is the biomass at which a thickening film on a fixed carrier area is half active. At steady state $X=Yr/(b+k_{det})$, so a film is sized by its load, 1461 g of ammonia oxidizers and 487 g of nitrite oxidizers for 1.5 kg N/d at 15 °C. The nitrite oxidizers are slow and of high affinity ($\mu_N=0.6$ against $\mu_A=0.9$ per day, $K_N=0.08$ against $K_A=0.7$ mg N/L), the choice that leaves them less reserve than the first population and delays the nitrite.

12.3.5 Cone, Degasser and Chemistry

The cone and the degasser act on the returning water without holdup, as shown in Equation (12.5).

$$\begin{aligned} E(G) &= \max(0.2,\ E_0-E_1 G/G_{max}) \\ \tau_G\frac{\mathrm{d}G}{\mathrm{d}t} &= G_{max}\,u\sqrt{P_s}-G \\ \eta_k &= 1-e^{-k_k GL},\quad GL=1440\,Q_{air}/Q_r+GL_0 \\ DO_{ret} &= C_s+(DO_b-C_s)(1-\eta_O) \\ &\quad+24000\,E(G)\,G/Q_r \\ C_{T,ret} &= C_{T,b}-\eta_C\,(CO_{2,b}-CO_{2,eq})/44.01 \end{aligned} \tag{12.5}$$

where G is in kg/h, C_s is oxygen saturation in fresh water [19], 10.08 mg/L at 15 °C, and $CO_{2,eq}=1.204$ mg/L is carbon dioxide in equilibrium with air at 600 ppm. The absorption falls from 0.92 with a nearly closed valve to 0.70 fully open, and the blower's 1 to 20 m^3/min moves the stripping efficiency η_C from 0.185 to 0.883. The pH follows from the charge balance $ALK=C_T(\alpha_1+2\alpha_2)+K_w/[H^+]-[H^+]$ solved by eight Newton steps on $\ln[H^+]$, with the carbonic acid constants of [20] and the water and carbon dioxide constants as the PHREEQC database carries them, and the unionized share of the ammonia is $f_{NH_3}=1/(1+10^{pK_a-pH})$ with $pK_a=0.09018+2729.92/T$, 9.564 at 15 °C [17]. At pH 7.2 and 15 °C the share is 0.43 %, and at 7.8 it is 1.69 %, so the total ammonia that reaches the 0.025 mg N/L limit falls from 5.81 to 1.48 mg/L. The water temperature follows its surroundings, $\tau_T\dot{T}=T_{env}-T$ with $\tau_T=1.5$ d. A power failure stops the pump, the cone, the blower, the feeder and the makeup, while the fish keep breathing.

12.3.6 Measurements

The controllers read only declared instruments. The tank oxygen probe, the gas flow meter, the pH probe, a membrane carbon dioxide probe behind a 10 min lag and a temperature probe carry first-order colored noise of 0.05 mg/L, 1 %, 0.01, 0.2 mg/L and 0.05 °C. An analyzer draws a tank sample every 4 h and reports total ammonia, nitrite and nitrate 20 min later with errors of 0.03, 0.01

Table 12.2 Model Parameters. Hidden Ones Are Marked h, the Rest Are in the Parameters Panel

Symbol	Meaning	Value	Unit
TGC, T_0 (h)	growth coefficient, zero	2.0, 0	—, °C
FCR_{sat}, c, ρ_m	conversion, ration curve	1.1, 0.3, 0.1	—
m	background mortality	0.02	%/d
τ_g, k_{SDA}	gut stage, post-meal O_2	3, 0.12	h, kg/kg
s_0 (h), b_O (h), θ_O	routine O_2 at 100 g	150, −0.2, 1.07	mg/kg/h, —, —
PC, f_{fec} (h), n_f (h), λ (h)	feed N, retention	0.45, 0.15, 0.028, 0.1	—
RQ (h), DO_{mk} (h)	quotient, makeup O_2	0.9, 0.95	—, of C_s
$LC50_{NO_2}$, $LC50_{NH_3}$	96-h LC50s	5, 0.25	mg N/L
a_0 (h), τ_E (h)	no-effect share, load lag	0.1, 1	—, d
V_t, V_b, Q_r	volumes, recirculation	50, 10, 100	m^3, m^3/h
G_{max}, E_0, E_1, τ_G (h)	cone	1.6, 0.92, 0.22, 2	kg/h, —, —, min
k_C, k_O (h), GL_0 (h)	degasser	0.17, 0.17, 0.6	—
μ_A, K_A, Y_A (h), b_A	ammonia oxidizers	0.9, 0.7, 0.15, 0.12	1/d, mg N/L
μ_N, K_N, Y_N (h), b_N	nitrite oxidizers	0.6, 0.08, 0.05, 0.12	1/d, mg N/L
K_{OA}, K_{ON} (h)	oxygen half-saturation	0.6, 1.2	mg/L
θ_A, θ_N, θ_b, all (h)	temperature factors	1.072, 1.06, 1.029	—
k_{det}, K_I (h), X_h	detachment, FA, carrier	0.05, 0.5, 4000	1/d, mg N/L, g
k_{het}, K_{OH} (h)	heterotrophs	0.10, 0.5	kg/kg, mg/L
x_0	natural seed of new media	0.1	% of mature
ALK_{mk}, pH_{mk}, $NO_{3,mk}$ (h)	makeup water	100, 7.6, 1	mg/L $CaCO_3$, —, mg N/L
τ_T (h), p_{CO_2}	temperature lag, air	1.5, 600	d, ppm
$\hat{s}_0$, FCR_{book}	farm's respiration, book	165, 1.0	mg/kg/h, —
T_s, n_s, CV	sample period, size, spread	14, 50, 18	d, fish, %
ΔT, t_{out}	warming, power failure	4, 20	°C, min
κ_A, κ_N	disinfectant kills	60, 75	%
c_{fish}, c_{feed}, c_{O_2}	prices	3.3, 1.6, 0.2	$/kg
c_w, c_{bic}	makeup, bicarbonate	0.5, 0.5	$/$m^3$, $/kg

and 0.5 mg N/L, holding each report until the next. A daily test kit, drawn at 09:00 and read an hour later, is the alternative. Section 1.4.5 treats such sparse and late samples. Dead fish are counted, so N is known, while the biomass is measured only by netting 50 fish every 14 d, whose mean carries the spread of individual weights, 18 %, over $\sqrt{50}$, about 2.5 %. The populations, the biofilter's water, W, B, E, a and the count killed N_{tox} are truth, drawn dashed in the scene and prefixed in the export.

12.3.7 Parameters, Initialization and Numerics

Table 12.2 lists the parameters. All are teaching values chosen to make a consistent design point, except the chemistry and saturation constants, which are published, and the toxicity constants, which are teaching values checked against the trout data above. Each case starts at a midnight of a lined-out plant. The cohort is placed as stocked, the film is either the steady film for the starting load or, on new media, a thousandth of the film a 0.5 kg N/d load grows, and the water is then settled over three days of the case's feeding (one without feed for new media) with the cohort and the film held. At $t = 0$ of the challenge the tank holds 500 kg of 50-gram fish fed 13.44 kg/d, a load of 494.8 g N/d on a film of 481.9 and 160.6 g, and the fish use 3.89 kg/d of oxygen against 4.40 kg/d in the biofilter. Integration uses fourth-order Runge–Kutta steps of 60 s, the controller runs every minute, and a row is logged every 15 min, on the engine of Section 1.6.1. Ninety simulated days take 45.0 s in a browser at the fastest setting, 2 d/s.

12.3.8 Assumptions

One cohort of one species lives in one well-mixed tank, with no size spread, grading, disease or behavior, and the coefficient model is a fit over a size range rather than a bioenergetic model. The tank and the biofilter are two well-mixed volumes and the cone and degasser act without holdup, so the model says nothing about where in a tall column the oxygen dissolves. Nitrification is two populations with Monod kinetics and a carrier-area limit, with no diffusion inside the film, no denitrification and no heterotrophic growth, so a thick film's response is approximate. The solids leave at once, and the

leached ammonia brings no alkalinity, so nitrification removes 1.023 rather than 1.000 equivalents per equivalent excreted. The chemistry is fresh water with unit activity coefficients. The temperature follows its surroundings with no heat balance. The toxicity law is linear between the no-effect level and the LC50, which no single law can match against tests of differing duration, and oxygen and carbon dioxide act on the fish only through the uptake's fall near anoxia. The water-quality limits are teaching values for a salmonid. Real ones come from water-quality criteria for reuse systems [2], national criteria [18] and a species' own toxicity data, and a simulated pass is not a welfare or compliance assessment.

12.4 Two Clocks in One Tank

12.4.1 The Meal, in Hours

The fast clock is the meal. In case 1 the oxygen station is held at 0.50 kg/h, the settled daily mean of 0.492 kg/h. Each meal's digestion overlaps the next, so the lows fall through the day, 9.99, 8.38, 7.75 and 7.63 mg/L at 11:00, 14:28, 18:05 and 21:40, and overnight the oxygen climbs to 12.50 mg/L by 07:35. The total ammonia follows the same digestion, from 1.09 at midnight to 1.30 mg/L at 22:03. A fixed gas flow is right only on average.

> **TRY IT IN THE APP** **1 · The Meal: Oxygen Dips and Ammonia Climbs**
>
> Load the case and press run. Watch the tank oxygen trend through the day. After each of the four meals the oxygen dips, and each low sits below the one before, down to 7.63 mg/L after the last meal ends. Overnight, with no meal to digest and the gas flow unchanged, the oxygen climbs back to 12.50 mg/L.

12.4.2 The Biofilter, in Weeks

The slow clock is the film. On new media in case 2, dosed with ammonium chloride and no fish, the total ammonia rises to 8.09 mg N/L on day 8.8 before the ammonia oxidizers catch up, and the nitrite they make rises to 10.78 mg N/L on day 18.9, a lag of 10.1 days, as shown in Figure 12.2a. Figure 12.4 shows the app's Water view on day 30 of this case. The nitrite oxidizers start from the same trace but have nothing to eat until nitrite appears, and in the first days free ammonia of up to 0.313 mg N/L also slows them. The lag is not an artifact of one parameter set. Run on the plant equations alone with the dose at 0.12 kg N/d and 6 m^3/d of makeup, the declared values give 13.25 days, and a tenfold smaller seed, no free-ammonia inhibition, a slower first population, no detachment, and equal activated-sludge-like kinetics without the carrier limit give lags from 8.7 to 22.1 days, with the nitrite peak second in every run.

> **TRY IT IN THE APP** **2 · Starting the Biofilter on New Media: The Ammonia Hump, Then the Nitrite Spike**
>
> Load the case and press run. Watch the ammonia and nitrite trend: the ammonia hump comes first and tops out on day 8.8, and the nitrite that the ammonia oxidizers make reaches its own peak about ten days later, on day 18.9. The startup sequence stays in its dose stage until the laboratory reads both low for three days.

A mature film shows the same physics after a feed step. Case 3 feeds a third of the table for weeks, so both populations have grown to that load, and then returns to the full table on day 1. A step from a third to the full table in one day would not happen on a well-run farm, but it separates the biofilter's delay from everything else. The total ammonia peaks at 4.34 mg/L at midnight, while the nitrite passes its limit at 15:38 on day 1.65 and stays above it after the meals until day 7.05. With the fish's response switched off the nitrite peak, 4.69 mg N/L, comes 2.19 days after the step and the limit is broken for 84.6 h. As shipped, the fish that take up the nitrite eat less (an appetite of 64 % on day 3.22), which ends the peak sooner at 3.07 mg N/L, 1.35 days after the step, at a cost of

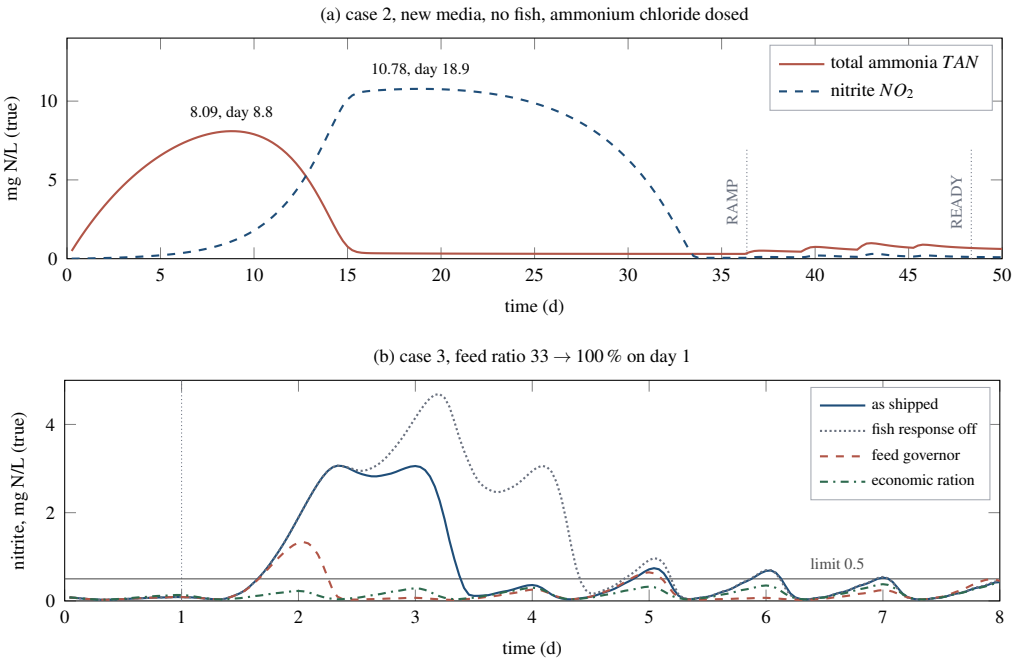


Figure 12.2 Nitrite after the ammonia, from the simulator's truth. (a) Case 2's startup on new media, total ammonia solid and nitrite dashed, with the sequence's RAMP and READY marked by dotted verticals. (b) Case 3's tripled ration on day 1 under four policies: as shipped solid, fish response off dotted, feed governor dashed and economic ration dash-dotted, against the 0.5 mg N/L teaching limit drawn as a thin line. Without the fish's response the peak comes 2.19 days after the step.

873 fish and 57.0 h over the limit, as shown in Figure 12.2b. The consequence for control is that a measurement of nitrite reports a decision made a day or more earlier.

Pause. Try Exercise 12.1 before reading on.

12.4.3 The Chemistry, in Units of pH

Case 4 holds a tank of 250-gram fish at pH 7.2, where the total ammonia ranges from 0.32 to 1.70 mg/L through day 0 and the un-ionized ammonia peaks at 0.0073 mg N/L. Raising the pH setpoint to 7.75 on day 1 barely moves the total, a daily high of 1.76 mg/L on day 2, but the un-ionized peak rises 3.73 times to 0.0273 mg N/L and first crosses the limit at 20:56 on day 1.87. The filter is unchanged and the toxicity is not.

Pause. Try Exercise 12.2 before reading on.

12.5 The Control Layers as Implemented

All five loops use the framework's velocity-form proportional integral algorithm of Section 1.6.2, Equations (1.12) to (1.15), run on measured values with no derivative term. This chapter writes the action $\alpha = \pm 1$ in place of Chapter 1's a, so $\alpha = +1$ is reverse acting and $\alpha = -1$ direct acting. Windup is prevented by back-calculation from the clamped output, with τ_I as the tracking time. A loop resting on a lower limit reports its request to the station floored at that limit.

12.5.1 The Oxygen Cascade

The outer loop runs every 5 min on the measured tank oxygen, reverse acting ($\alpha = +1$), with $K_c = 0.25$ kg/h per mg/L and $\tau_I = 30$ min, and its output plus a meal feedforward is the gas flow

setpoint,

$$G_{SP} = \mathrm{clip}_{[0,\,G_{max}]}\left(\mathrm{PI}_{DO} + \frac{k_{SDA}\,\hat{p}}{24\,E(G_m)}\right) \tag{12.6}$$

where $\hat{p}$ is the processed feed of the controller's own two-stage gut model driven by the feed it delivered, and $E(G_m)$ is the declared absorption at the metered flow. The inner loop runs every minute on the flow meter, reverse acting, with $K_c = 0.6$ per kg/h and $\tau_I = 2$ min and moves the valve between 0 and 100 %. Both loops track while the pumps are stopped. In case 1 switched to Auto on the second morning, the next day's oxygen stays between 8.88 and 9.10 mg/L, against 8.76 to 9.15 mg/L without the feedforward. The inner loop earns its place on the supply side. A sag to 60 % of the supply pressure for six hours takes the tank to 8.87 mg/L with the inner loop in Auto and 8.44 mg/L with it in Manual. Volume 3, Chapter 3 measures how fast such an inner loop has to be in a fermenter's oxygen cascade.

12.5.2 pH, Carbon Dioxide and Nitrate

Bicarbonate is dosed between 0 and 150 kg/d by the pH loop every 5 min, reverse acting, with $K_c = 200\,\mathrm{kg/d}$ per pH unit and $\tau_I = 2$ d. The carbon dioxide loop moves the blower between 1 and 20 $\mathrm{m^3/min}$ every 5 min, direct acting ($\alpha = -1$), since more air lowers the carbon dioxide, with $K_c = 1.5$ per mg/L and $\tau_I = 1$ h, toward a setpoint of 10 mg/L. The nitrate loop sets the makeup between 2 and 60 $\mathrm{m^3/d}$, direct acting ($\alpha = -1$), since more makeup dilutes the nitrate, with $K_c = 0.5$ per mg N/L and $\tau_I = 2$ d, and executes only when an analyzer report arrives, toward 60 mg N/L. A schedule replaces it when the switch is set.

12.5.3 Feeding: Table, Book, Governor and Optimizer

The feed table gives the daily ration in percent of body weight by weight class and temperature, printed once from the declared model at satiation, and the farm multiplies it by the biomass it believes, either the feed book (the stocking weight plus the feed fed over an assumed conversion of 1.0, reset at each sample) or the soft sensor. The feed governor is an override with a low select, of the kind Section 1.4.6 describes for limits approached through another variable. Each limit gives a fraction x of its limit, from the laboratory's total ammonia at the measured pH and temperature, from the nitrite, from the carbon dioxide probe, and for oxygen from the fall below its setpoint toward 6 mg/L. For ammonia and nitrite the governor reads the larger of the last report and the next one projected from the last two. Each limit cuts the ration as $f = 1 - (1 - f_{min})\,\mathrm{clip}_{[0,1]}((x-a)/(1-a))$ with $a = 0.7$ (0.5 for oxygen) and $f_{min} = 0.25$, the lowest cut wins, and the cut in force falls at once and rises by at most a quarter of the ration a day. It acts on a four-hourly analyzer, the slow measurement whose cost to a composition loop Volume 5, Chapter 3 measures.

The economic optimizer chooses the day's ration at midnight as a fraction ρ of the table ration R_{tab}, as shown in Equation (12.7).

$$\begin{aligned} \max_{0\le\rho\le 1}\ & c_{fish}\,\frac{\hat{N}\,g(\rho)}{1000} \\ & - \left(c_{feed} + c_{O_2}\tfrac{o_f}{E} + c_w w_f + c_{bic} b_f\right)\rho R_{tab} \\ \text{s.t. } & \dot{O}_r + 1.5\,o_f\,\rho R_{tab} \le 0.9\,O_{cap} \\ & 1.2\,y\,\rho R_{tab} \le \min(\hat{C}_A, \hat{C}_N) \\ & \dot{C}_0 + \dot{C}_1\,\rho R_{tab} \le \dot{C}_{cap}, \quad \rho R_{tab} \le 1.25\,R_{k-1} \end{aligned} \tag{12.7}$$

where $g(\rho)$ is the farm's growth model at the believed weight and the measured temperature, o_f, w_f and b_f are the oxygen, makeup water and bicarbonate per kilogram of feed, y the ammonia yield, $\dot{O}_r$ the routine oxygen, O_{cap} the cone and degasser capacity at full valve and full air, 1.5 and 1.2 declared peak factors on the oxygen and nitrogen loads, $\hat{C}_A$ and $\hat{C}_N$ the film observer's capacities at 60 % of each limit, the carbon dioxide row the degasser's capacity scaled from yesterday's operating

point, and the last row a ramp of 25 % a day. Every constraint is affine in ρ, so the feasible interval is exact, and golden section on the concave objective stops at 10^{-4} of the table or 40 iterations. With no feasible ration the fish fast for a day.

12.5.4 Two Estimators and a Sequence

The soft sensor closes a daily oxygen ledger from measurements and declared values only. The fish's oxygen is the cone's dissolved gas, $\sum E(G_m)G_m$, plus the degasser's transfer and the makeup's oxygen, less the accumulation in the tank and the biofilter's oxygen from the measured conversion, and the post-meal oxygen of the processed feed is then removed. The remaining routine respiration, divided by the declared per-kilogram rate at $\hat{s}_0 = 165\,\mathrm{mg/kg/h}$, gives a weight and with the count a biomass measurement $z = \hat{N}\hat{W}/1000 \cdot e^{\beta}$. The update is a fixed-gain observer,

$$\begin{aligned} \hat{B}^{-} &= \hat{B} + \frac{\hat{N} g(\hat{\rho})}{1000}, \qquad \hat{B} = \hat{B}^{-} + w(z - \hat{B}^{-}) \\ \beta &\leftarrow \beta + \lambda(\ln B_s - \ln z_{raw} - \beta) \end{aligned} \tag{12.8}$$

with $w = 0.3$, the sample B_s correcting the bias β with $\lambda = 0.8$ through the framework's bias updater (clamped to ± 0.5), the logarithmic update of Equation (1.8), and the estimate itself by 0.8 of the residual. Volume 5, Chapter 10 corrects a moisture soft sensor's bias the same way from a sampled and delayed laboratory. The cohort's own coefficient, 150, is hidden, so the sensor starts biased.

The film observer follows each population once a day. The ammonia made is the farm's yield times the processed feed plus any ammonium chloride the controller dosed, and the conversion $\hat{r}_A$ is that less the accumulation and the makeup's loss of total ammonia, and $\hat{r}_N$ the same on nitrite. Each estimate grows by $Y\hat{r} - (b + k_{det})\hat{X}$, and when the filter's water was above half-saturation but the film converted less than 85 % of what the estimate predicts, the estimate is scaled down by the shortfall, which is how a kill shows. The capacity it reports is what the estimated film converts at 60 % of each limit. A startup sequence on the shared stage sequencer doses 0.12 kg N/d of ammonium chloride until the laboratory reads total ammonia under 1 and nitrite under 0.5 mg N/L for three days, raises the dose by half every three days while the readings hold up to the 0.5 kg N/d that 10,000 fingerlings of 50 g make, and then reads READY.

12.5.5 Three Models

The plant is Section 12.3. The farm's model, used by the feedforward, the soft sensor, the film observer and the optimizer, is a copy of the declared values with the respiration coefficient set to 165 against the cohort's hidden 150, and it never reads the populations, the biofilter's water, the true pH or the weight. A grep-level test in the application fails if a controller or estimator file reads a truth field. A third, identified model exists only when a learner fits one to a case's trends.

12.6 Ten Cases, Measured

The cases appear in Table 12.3, and Section 1.6.4 describes how a case is run, scored and exported. The numbers below come from the engine run headless at each case's seed and advanced with the same calls the browser makes, with any setting a sentence names entered as that sentence says. Opening a case first, choosing it after another one, or restarting it with Reset produces identical runs. The shared defaults are the tunings above, the feed table on the feed book with no governor, the analyzer, and the mature film for the starting ration.

12.6.1 The Governor on a Sampled Analyzer

The governor can prevent what it sees coming. On a step from 55 to 100 % of the table it holds the nitrite under its limit, where the step without it breaks the limit for 26.5 h. On case 3's tripled ration it cuts the time over from 57.0 to 20.7 h with a peak of 1.34 mg N/L and one fish lost, because the analyzer reports every four hours, after the meals that make the nitrite have been eaten. The economic

Table 12.3 The Cases

#	Case and events	Teaching point
1	150-g fish, Manual gas flow, 3 d	the meal
2	New media, no fish, startup sequence, 50 d	ammonia, then nitrite
3	Feed ratio 33 to 100 % on day 1, 10 d	nitrite days later
4	250-g fish, pH setpoint to 7.75 on day 1	pH sets the toxicity
5	Soft sensor feeding, new feed lot on day 21, 60 d	a biomass from oxygen
6	Power failure at 12:00 for 20 min, 1 d	minutes of reserve
7	Governor and soft sensor, 30 d	the ration as economics
8	Disinfectant kills 60 and 75 % on day 1, 14 d	a film regrown
9	Challenge: 90 d, warming on day 40, upset on day 62	the whole farm
10	10,000 fish stocked on new media, 21 d	infeasible

ration under its ramp raises the feed 4.8 to 12.6 % a day, from 14.4 to 25.1 kg/d over nine days, and never breaks the limit, as shown in Figure 12.2b. The biofilter, not the fish, sets how fast the ration may rise.

12.6.2 Weighing Fish by Their Oxygen

In case 5 the table feeds on the soft sensor. Before the first sample the estimate reads up to 18.5 % low, because the farm's coefficient is 10 % above the cohort's, and on day 14 the table feeds 14.4 kg/d against the 16.4 kg/d the fish would eat. The sample on day 14.4 reads 674 kg, 0.6 % below the truth, and moves the bias. From day 15 the estimate stays within 5.5 % of the truth between midnights and within 4.1 % after the second sample. A new feed lot that grows 15 % less per kilogram arrives on day 21. The feed book cannot see it and is 8.2 % high on day 42 (10.2 % at its worst), while the soft sensor, which reads what the fish breathe, is 2.0 % high.

Pause. Try Exercise 12.3 before reading on.

12.6.3 Minutes, Margin and a Lost Film

In case 6 the oxygen above the 6 mg/L limit in the tank is about 149 g against an uptake of 0.529 kg/h at noon, and the oxygen reaches the limit 16.7 min after the power fails, falling to 5.34 mg/L. With 20,000 fish (4.98 t) it takes 8.7 min, 1.92 times faster, and falls to 2.15 mg/L, while 300-gram fish (3.01 t) take 14.6 min and 150-gram fish (1.51 t) stay above the limit with a low of 6.41 mg/L. Doubling the count halves the reserve time, and weight shortens it more slowly because heavier fish use less per kilogram.

Case 7 prices the ration over 30 days in Table 12.4. Feeding 80 % of the table converts better than the full table, 1.05 against 1.09, and earns less, $764 against $847, and the fish reach the full table's 234 g only on day 35.5. The optimizer feeds 0.95 of the table and earns $837, within 1.1 % of the full table on 45 kg less feed. It earns slightly less because it maximizes each day's margin and does not value that a heavier fish grows more tomorrow. Its value appears where a limit binds. In case 8 a disinfectant kills 60 % of the ammonia oxidizers and 75 % of the nitrite oxidizers on day 1 while 10,000 fish of 200 g keep eating. Without the governor the nitrite reaches 6.49 mg N/L on day 2.22, the appetite falls to 33 % and 1,843 fish die. The governor holds the peak to 1.09 mg N/L with no fish lost, and the optimizer with the soft sensor to 0.91 mg N/L, 2.5 h of the 6.4 h over the limit coming before the first report after the upset.

> **DID YOU KNOW? Better Conversion, Smaller Margin**
>
> From the simulator: over the 30 days of case 7, feeding 80 % of the table gains 677 kg and earns $764, against 824 kg and $847 on the full table (Table 12.4). Fish sold by weight pay for growth, not for a good conversion.

Table 12.4 Measured Results. Hours Over Any Limit, Nitrite Peak, Fish Killed by the Water, Biomass Gained and Margin

Case	Policy	Over (h)	NO_2 peak (mg N/L)	Killed	Result
3, 10 d	as shipped	57.0	3.07	873	—
	fish response off	84.6	4.69	0	—
	governor	20.7	1.34	1	—
	economic ration	0.0	0.41	0	—
7, 30 d	table 100 %	—	—	—	824 kg, $847
	table 80 %	—	—	—	677 kg, $764
	table 60 %	—	—	—	506 kg, $581
	economic, $\rho = 0.95$	—	—	—	793 kg, $837
8, 14 d	as shipped	84.1	6.49	1,843	1974 kg
	governor	13.4	1.09	0	2237 kg
	economic, soft sensor	6.4	0.91	0	2241 kg

Case 3's economic row is that run's peak over 8 days. Case 7's results are biomass gained and margin, case 8's the biomass on day 14.

12.6.4 The Infeasible Variant

Case 10 stocks 10,000 fingerlings directly on new media with the governor on. The nitrite passes its limit on day 5.8, the total ammonia reaches 14.8 mg/L on day 8.0 and the nitrite 15.8 mg N/L on day 14.4, and the fish go off their feed and die, half of them by day 9.2 and 9,937 of the 10,000 by day 21. A logged sweep of feed ratios from 10 to 100 % with makeup from 10 to 60 m^3/d, the governor off and on, breaks the limit in every run. The lowest peak, 0.517 mg N/L at 10 % of the table and 60 m^3/d, comes from the maintenance ration, on which the biomass falls from 500 to 490.1 kg in three weeks, and at 20 %, where the fish just hold their weight, the peak is 0.632 mg N/L. A lower pH through a higher carbon dioxide setpoint and the daily test kit do not change that. The economic feeding stays under, at 0.345 mg N/L, only by starving the cohort to 471.0 kg, and colder water at 12 °C would allow it. At the case's 15 °C no ration that grows the fish keeps nitrite under its limit, and the answer is case 2's sequence, which carries the stocking load by day 48.3.

12.6.5 The Challenge

The challenge grows 10,000 fingerlings of 50 g for 90 days through a 4 °C warming of the surroundings from day 40 and a disinfectant upset on day 62. It is judged on the truth at day 90: a harvest of at least 2.40 t, any water-quality limit broken for at most 8 h in total, and a feed conversion of at most 1.10, with every parameter and the stocking as shipped. Table 12.5 gives five policies at the case's seed and over eleven seeds. As shipped the nitrite after the upset peaks at 6.85 mg N/L on day 63.1 and stays over the limit for 69.8 h, the appetite falls to 42 % and 1,359 fish die. The cohort still weighs 2.58 t, in larger fish, on 2664 kg of feed, a conversion of 1.28 and a margin of $1,326. The suggested soft sensor, governor and economic ration harvest 2.54 t in 1,352 more fish on 2199 kg of feed, at 1.08 and $2,148, with 6.6 h over a limit and none lost. Of those hours, 2.8 h come before the first report that shows the upset and 3.8 h from meals already eaten passing through the weakened filter. Feeding past what the biofilter can carry does not buy fish. Figure 12.3 draws the five policies of Table 12.5 side by side.

12.6.6 Verification

Code verification, agreement with published relations and educational use are kept apart in this section. On the code side the nitrogen inventory over the shipped challenge, 191.8 kg of nitrogen in through feed and 132.8 kg out, closes to 3.6×10^{-12} kg, and the oxygen over its 795.8 kg of fish uptake to 9.4×10^{-13} kg. The same holds through case 10's mortality, case 2's dose and a power failure, a supply sag and a mortality event fired together, to below 3×10^{-12} kg. The one exception is case 6 with 20,000 fish, whose biofilter goes anoxic during the outage and leaves an oxygen residual of 1.05×10^{-6} kg through a round-off guard. Halving the step to 30 s with the instrument noise off

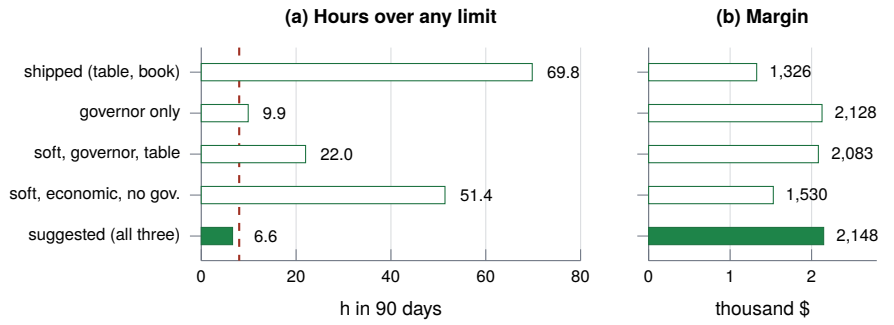


Figure 12.3 The five challenge policies of Table 12.5 at the case's seed: hours over any water-quality limit against the 8 h the challenge allows (dashed line), and the margin. The solid bar passes, white bars fail. Notice that the fewer the hours a policy spends over a limit, the larger its margin, and that the governor alone misses the allowance only narrowly.

Table 12.5 The Challenge Over 90 Days at the Case's Seed, and the Range Over Eleven Seeds

Policy	Harvest (t)	Over (h)	FCR	Killed	Margin ($)	Seeds passing
shipped (table, book)	2.58	69.8	1.28	1,359	1,326	0/11
governor only	2.55	9.9	1.08	0	2,128	2/11
soft, governor, table	2.47	22.0	1.07	0	2,083	4/11
soft, economic, no gov.	2.47	51.4	1.21	1,172	1,530	0/11
suggested (all three)	2.54	6.6	1.08	0	2,148	11/11
shipped, 11 seeds	2.57–2.58	69.6–69.9	1.28–1.30			
suggested, 11 seeds	2.45–2.54	6.0–6.9	1.07–1.08	0		

moves the shipped harvest by 0.004 % and the suggested one by 0.396 %, and case 3's nitrite peak, 3.052 mg N/L, not at all. At every midnight of the suggested challenge the optimizer's ration agrees with the closed-form maximum of its concave quadratic objective within 2.0×10^{-5}, 65 times interior, 15 times on the nitrite row and 10 on the oxygen row. The application's own suite passes 44 of 44 checks.

Against published relations, the carbonic acid, water and carbon dioxide constants equal those of the PHREEQC database, which carries the expressions of [20], to four decimals at 10, 15 and 25 °C, and three equilibria through the app's solver and an independent bisection agree to four decimals in pH. The un-ionized ammonia share is within 0.4 % of the form the national criteria print [18] at three conditions. The satiated cohort follows the declared coefficient curve, 96.3, 164.9 and 260.2 g at 30, 60 and 90 days, within 0.447 %. These checks confirm that the model is built as declared. They do not validate it against a farm, and no learning gain is claimed.

12.7 Technology in Practice

The simulated farm runs five loops, a feed governor, an economic ration and two estimators on a handful of instruments. Figure 12.5 draws the oxygen cascade and the nitrogen path with the devices a land-based farm would install.

What Measures It

The fast loops read probes in the tank for oxygen, pH, carbon dioxide and temperature, each with colored noise. The nitrogen that sets the ration is not measured in the tank at all. The chapter's analyzer draws a sample every 4 h and reports total ammonia, nitrite and nitrate 20 min later, and the

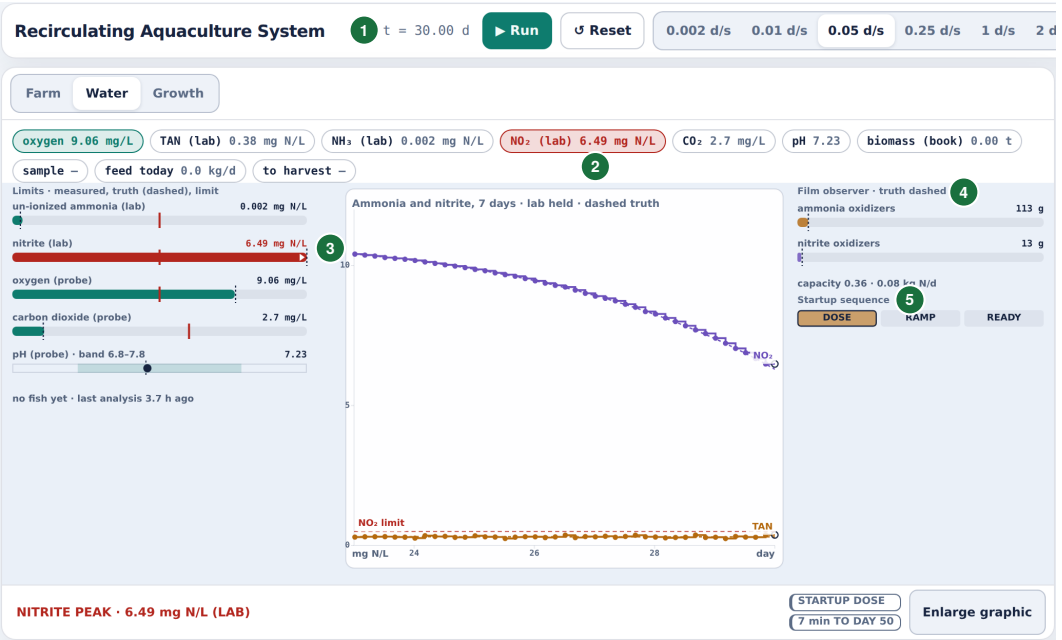


Figure 12.4 Case 2's biofilter startup on day 30 (1), on the Water view. The laboratory (2) holds total ammonia at 0.38 and nitrite at 6.49 mg N/L from a report 3.7 h old, against the dashed truth of 0.31 and 6.30 on the limit bars (3), and the status strip reads the nitrite peak. The film observer (4) believes 113 and 13 g of ammonia and nitrite oxidizers, against a truth of 112 and 16 g, and the startup sequence (5) is still dosing. The readings show tank oxygen at 9.06 mg/L, pH 7.23 and carbon dioxide at 2.7 mg/L.

biomass is known only from netting 50 fish every 14 d. No loop or estimator reads the simulated truth.

> **TECHNOLOGY** **Sensor: Optical Dissolved Oxygen Probe**
>
> Dissolved oxygen is measured mainly by two methods, the Clark electrode and fluorescence quenching [21]. In the optical probe a thin layer of material fluoresces red under green or blue light, and oxygen that reaches the layer quenches the glow, so less oxygen gives a stronger and longer fluorescence [22]. In a circular tank the probe sits near the outlet, never close to where oxygen is injected or oxygen-rich water enters, and it must be wiped clean and calibrated from time to time [15]. Water quality sensors are prone to biological growth on their surfaces, which makes online measurement less reliable [21]. The chapter's probe reads the true tank oxygen with colored noise of 0.05 mg/L.

What Moves It

The oxygen station sets the gas into the cone, a dosing pump adds bicarbonate, the blower drives air through the degasser, a valve admits the makeup water and the feeder delivers the day's ration in four meals. The ration is the actuator that matters most, because the load it puts on the biofilter returns days later as nitrite, while the other four act within minutes or hours.

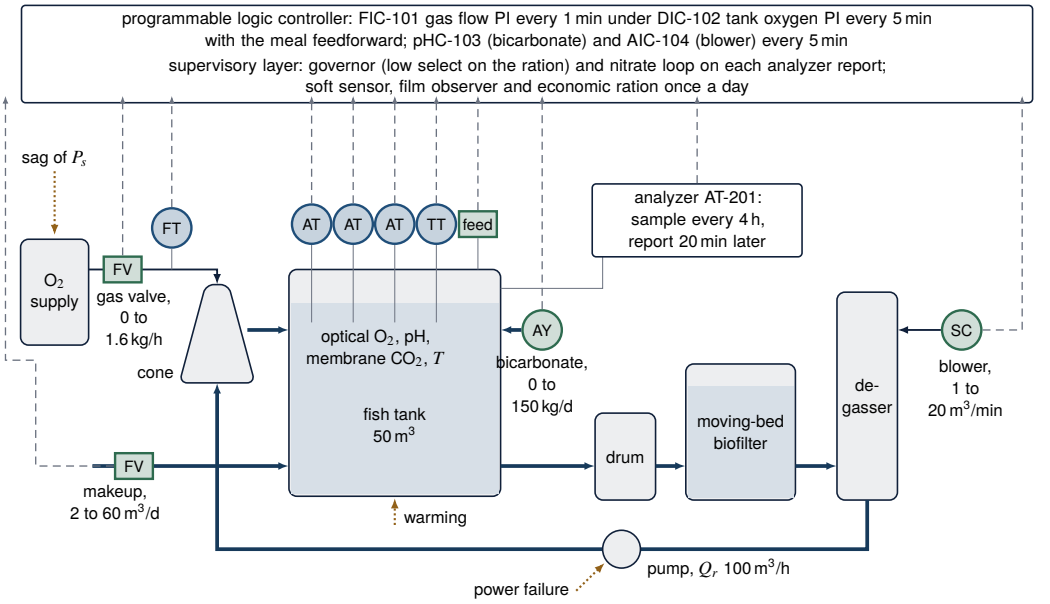


Figure 12.5 The chapter's loops with the devices a land-based farm would install. Thick blue lines carry the water, thin dark lines the oxygen and the air, dashed gray lines the signals and dotted amber arrows the disturbances. Notice that the probes on the tank are read continuously, while the nitrogen that limits the ration reaches the controller only through an analyzer sample every four hours.

TECHNOLOGY Final Element: Oxygen Cone on a Gas Valve

Pure oxygen reaches a farm as liquid in tanks or comes from an oxygen generator on site. In a cone a pump forces water and the gas together at typically around 1.4 bar, which drives the oxygen into the water, at the price of the electricity the pump uses. A cone treats only part of the circulating water, while a low-pressure oxygen platform treats the main flow. A central control system can raise the oxygen dose at each feeding as the fish's consumption rises [15]. In the chapter the valve sets a gas flow of up to 1.6 kg/h, an inner loop holds that flow, and the cone's absorption falls from 0.92 to 0.70 as the valve opens.

Where the Controller Runs

On a land-based farm a central control system monitors and controls oxygen, temperature, pH, water levels and motor functions, starts a corrective action when a value leaves its preset band, raises an alarm when that fails, and can coordinate the automatic feeding with a higher oxygen dose during the meals. Its operators set its alarms and emergency levels on programmable logic controllers and computers [15]. A programmable logic controller holds instructions for logic, timing, counting and three-mode PID control, and supervisory and lower loops run continuously with cycle times from milliseconds to minutes [23]. The chapter's inner gas-flow loop runs every minute and the outer oxygen, pH and carbon dioxide loops every 5 min, so the platform's scan is never the limit. In a function block program the oxygen cascade is two PI blocks with the meal feedforward summed into the outer block's output. The governor is a low select on the ration, and the economic ration and the two estimators run once a day in a supervisory layer above the loops.

What Is Modernizing It

Three lines of work aim at the delays this chapter measures. The first is the sensor. A review describes optical dissolved oxygen probes as needing no calibration and responding fast, and expects probes based on fluorescence without a membrane to be a focus of future research [21]. The second is

the soft sensor. Models that take online temperature, salinity, pH and oxygen as auxiliary variables have been built to estimate the ammonia nitrogen [21], as the chapter's soft sensor estimates the biomass from the oxygen balance. The third is learned feeding control. A reinforcement-learning feeding controller for recirculating systems has been compressed to run on a small edge device in real time [24], and precision fish farming sets feed delivery among the first problems that control engineering should take on [6]. None of them removes the days a nitrifier film needs to grow.

> **IN PRACTICE** **Minutes Count When the Power Fails**
>
> Technical failures can easily cause substantial losses, so the control system calls the staff when a major failure is about to occur, and a reaction time under 20 minutes is recommended even where automatic backups exist. Pure oxygen as a backup is the first safety precaution: if the electricity fails, a magnetic valve opens and stored oxygen flows to diffusers in every tank, at a flow set beforehand so that the store lasts until the failure is corrected. A fuel-driven generator should restart the main pumps within an hour or so, because the ammonia builds up when the water stops passing over the biofilter. The routine includes alarm tests and a check of the emergency oxygen in every tank [15].

12.8 Limitations

The sizes, kinetics and prices are teaching values that make one consistent farm without being fitted to any. The two-population film with a carrier limit was chosen because equal activated-sludge-like kinetics leave a mature film three to four times its load in reserve, so its delay after a feed step is a property of the chosen constants and not a prediction for a real filter, while the startup's ammonia-then-nitrite order survives every variant tried. The toxicity law is a declared teaching law whose 96-hour mortality at the LC50 is 39.7 % rather than half, and it ignores chronic effects below the limits, disease and the effect of carbon dioxide on growth. One cohort, two mixed volumes and a temperature without a heat balance leave out grading, stratification, total gas pressure and the farm's heating.

The soft sensor is a fixed-gain observer whose weights are declared rather than computed from noise statistics, and the film observer finds a film larger than it believes only by growth. The optimizer maximizes one day's margin, uses declared peak factors of 1.5 and 1.2 for the meal, and does not plan the cohort to harvest, which is why it trails the full table in case 7. The governor reads a four-hourly analyzer and cannot prevent the nitrite of meals already eaten. The disturbances arrive as scheduled steps.

> **OPEN CHALLENGE** **Measuring Nitrogen Where the Fish Are**
>
> Ammonia, nitrite and nitrate on a fish farm are usually measured only by online analyzers that move the laboratory's chemistry into a detection box. Water quality sensors suffer from biological growth on their surfaces, poor reliability and a short life, and many of the studies behind intelligent farms are still done in a laboratory [21]. Precision fish farming asks for control-engineering principles that move production from an experience-based to a knowledge-based regime, with biomass monitoring and feed delivery among its first cases [6]. Progress would look like reliable nitrogen sensing in the tank itself, so that a feed governor could act on the meal rather than on a sample hours old.

12.9 Conclusion

A recirculating farm is a plant whose most important actuator, the nitrifying film, has to be grown, and the simulator of this chapter integrates the fish, the film and the water's chemistry together with every controller restricted to the farm's own instruments. The fish answer a meal in hours, with oxygen lows falling to 7.63 mg/L under a fixed gas flow, while a new film shows its ammonia peak on day 8.8 and its nitrite peak on day 18.9, and a tripled ration returns as nitrite 1.35 days later. A governor on a four-hourly analyzer cuts case 3's hours over the nitrite limit from 57.0 to 20.7, a soft sensor on the oxygen balance follows the biomass within 5.5 % once a sample has removed its bias, and the three

together pass the ninety-day challenge on eleven of eleven seeds where the shipped table fails on all eleven. Stocking on new media is infeasible at any ration that grows the fish. The model is a declared teaching model and its film and toxicity constants are chosen, not fitted. A fuller model would add diffusion inside the film, a size-structured cohort, a heat balance and a ration planned over the whole cohort. The simulator shows that on a fish farm the ration is the load, and the biofilter decides how fast it may change.

KEY POINTS

- A recirculating farm cleans its water with bacteria that grow only on the load they carry, so a larger ration returns as nitrite days later.
- In case 3 the nitrite of a tripled ration peaks at 3.07 $\mathrm{mg\,N/L}$ 1.35 days after the step and stays above its limit for 57.0 h.
- A governor on a four-hourly analyzer cuts those hours to 20.7, and it cannot prevent the nitrite of meals already eaten.
- A soft sensor on the oxygen balance follows the biomass within 5.5 % once a sample has removed its bias, and with the governor and the economic ration it passes the challenge on eleven of eleven seeds.

Where else this applies. Volume 9, Chapter 10 treats nitrifiers as the plant, Chapter 11 a culture as the actuator, and Volume 3, Chapter 3 an oxygen cascade like this farm’s.

References

[1] S. T. Summerfelt, B. J. Vinci, and R. H. Piedrahita, “Oxygenation and carbon dioxide control in water reuse systems,” *Aquacultural Engineering*, vol. 22, no. 1–2, pp. 87–108, 2000.

[2] J. Colt, “Water quality requirements for reuse systems,” *Aquacultural Engineering*, vol. 34, no. 3, pp. 143–156, 2006.

[3] E. H. Eding, A. Kamstra, J. A. J. Verreth, E. A. Huisman, and A. Klapwijk, “Design and operation of nitrifying trickling filters in recirculating aquaculture: A review,” *Aquacultural Engineering*, vol. 34, no. 3, pp. 234–260, 2006.

[4] T. M. Losordo and A. O. Hobbs, “Using computer spreadsheets for water flow and biofilter sizing in recirculating aquaculture production systems,” *Aquacultural Engineering*, vol. 23, no. 1–3, pp. 95–102, 2000.

[5] T. E. I. Wik, B. T. Lindén, and P. I. Wramner, “Integrated dynamic aquaculture and wastewater treatment modelling for recirculating aquaculture systems,” *Aquaculture*, vol. 287, no. 3–4, pp. 361–370, 2009.

[6] M. Føre, K. Frank, T. Norton, E. Svendsen, J. A. Alfredsen, T. Dempster, H. Eguiraun, W. Watson, A. Stahl, L. M. Sunde, C. Schellewald, K. R. Skøien, M. O. Alver, and D. Berckmans, “Precision fish farming: A new framework to improve production in aquaculture,” *Biosystems Engineering*, vol. 173, pp. 176–193, 2018.

[7] FAO, “In brief to the state of world fisheries and aquaculture 2024: Blue transformation in action,” Food and Agriculture Organization of the United Nations, Rome, Tech. Rep., 2024.

[8] R. C. Russo, C. E. Smith, and R. V. Thurston, “Acute toxicity of nitrite to rainbow trout (*salmo gairdneri*),” *Journal of the Fisheries Research Board of Canada*, vol. 31, no. 10, pp. 1653–1655, 1974.

[9] W. M. Lewis and D. P. Morris, “Toxicity of nitrite to fish: A review,” *Transactions of the American Fisheries Society*, vol. 115, no. 2, pp. 183–195, 1986.

[10] G. K. Iwama and A. F. Tautz, “A simple growth model for salmonids in hatcheries,” *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 38, no. 6, pp. 649–656, 1981.

[11] P. B. Liao and R. D. Mayo, “Intensified fish culture combining water reconditioning with pollution abatement,” *Aquaculture*, vol. 3, no. 1, pp. 61–85, 1974.

[12] P. G. Lee, “Process control and artificial intelligence software for aquaculture,” *Aquacultural Engineering*, vol. 23, no. 1–3, pp. 13–36, 2000.

[13] C. I. M. Martins, E. H. Eding, M. C. J. Verdegem, L. T. N. Heinsbroek, O. Schneider, J. P. Blancheton, E. R. d’Orbcastel, and J. A. J. Verreth, “New developments in recirculating aquaculture systems in Europe: A perspective on environmental sustainability,” *Aquacultural Engineering*, vol. 43, no. 3, pp. 83–93, 2010.

[14] R. E. Speece, “U-tube oxygenation for economical saturation of fish hatchery water,” *Transactions of the American Fisheries Society*, vol. 98, no. 4, pp. 789–795, 1969.

[15] J. Bregnballe, *A Guide to Recirculation Aquaculture: An Introduction to the New Environmentally Friendly and Highly Productive Closed Fish Farming Systems*. Copenhagen: Food and Agriculture Organization of the United Nations and EUROFISH International Organisation, 2015, iSBN 978-92-5-108776-3, https://www.fao.org/3/i4626e/i4626e.pdf.

[16] H. Kroupova, J. Machova, and Z. Svobodova, "Nitrite influence on fish: a review," *Veterinární medicína*, vol. 50, no. 11, pp. 461–471, 2005.

[17] K. Emerson, R. C. Russo, R. E. Lund, and R. V. Thurston, "Aqueous ammonia equilibrium calculations: Effect of pH and temperature," *Journal of the Fisheries Research Board of Canada*, vol. 32, no. 12, pp. 2379–2383, 1975.

[18] U.S. Environmental Protection Agency, "Aquatic life ambient water quality criteria for ammonia – freshwater 2013," U.S. Environmental Protection Agency, Office of Water, Office of Science and Technology, Washington, DC, Tech. Rep. EPA-822-R-13-001, Apr. 2013.

[19] B. B. Benson and D. Krause, "The concentration and isotopic fractionation of oxygen dissolved in freshwater and seawater in equilibrium with the atmosphere," *Limnology and Oceanography*, vol. 29, no. 3, pp. 620–632, 1984.

[20] L. N. Plummer and E. Busenberg, "The solubilities of calcite, aragonite and vaterite in CO_2-H_2O solutions between 0 and 90°C, and an evaluation of the aqueous model for the system $CaCO_3$-CO_2-H_2O," *Geochimica et Cosmochimica Acta*, vol. 46, no. 6, pp. 1011–1040, 1982.

[21] C. Wang, Z. Li, T. Wang, X. Xu, X. Zhang, and D. Li, "Intelligent fish farm—the future of aquaculture," *Aquaculture International*, vol. 29, no. 6, pp. 2681–2711, 2021.

[22] T. R. Kuphaldt, *Lessons in Industrial Instrumentation*, version 2.33 ed. Open textbook, Creative Commons Attribution 4.0 license, ibiblio.org/kuphaldt/socratic/sinst, 2024.

[23] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule, and M. Thompson, "Guide to operational technology (OT) security," National Institute of Standards and Technology, Tech. Rep. NIST Special Publication 800-82r3, 2023.

[24] W. M. Elmessery, M. Y. Shams, T. A. El-Hafeez, P. Szűcs, M. H. Eid, M. Alhumedi, A. F. Ahmed, and A. E. Elwakeel, "Lightweight deep deterministic policy gradient for edge computing in recirculating aquaculture systems: real-time feeding control with reduced computational requirements," *Scientific Reports*, vol. 15, p. 37960, 2025.

12.10 Exercises

12.1 ○ Concept check · 2 min

In case 3 the feed ratio steps from 33 to 100 % on day 1. The true total ammonia peaks at 4.34 mg/L that midnight, but the true nitrite peaks only 1.35 days after the step. Why does the nitrite come so much later?

(a) The analyzer reports every four hours, so the nitrite peak shows on the trend later than it occurs

(b) Nitrite forms slowly in the tank's own water from the ammonia, away from the biofilter

(c) The fish excrete nitrite only after two gut stages of about three hours each

(d) The nitrite oxidizers have less reserve and grow more slowly, so they must grow on the new load first

12.2 ○ Concept check · 1 min

In case 4 the pH setpoint rises from 7.2 to 7.75 on day 1. The daily high of total ammonia moves only from 1.70 to 1.76 mg/L. What happens to the toxicity of the water?

(a) It rises about 3.7 times, because the un-ionized share of the same total grows with pH

(b) It falls, because the bicarbonate dosed to raise the pH binds part of the ammonia

(c) It stays the same, because the limit applies to the total ammonia and the total barely moved

(d) It rises about tenfold, because each step of the pH multiplies the un-ionized share by ten

12.3 ○ Concept check · 2 min

In case 5 the table feeds on the soft sensor, and before the first sample on day 14 the estimate of the biomass reads up to 18.5 % low. What causes the low reading?

(a) The oxygen probe's colored noise averages to a negative error over each day
(b) The farm's model gives each kilogram of fish a higher routine oxygen use than this cohort has
(c) The feed book resets the soft sensor to the stocking weight at every midnight
(d) The ledger leaves out the biofilter's oxygen, which hides part of the fish's use

12.4 Theory and calculation · 5 min

The ammonium constant is $pK_a = 0.09018 + 2729.92/T$ with T in kelvin, and the un-ionized share of the total ammonia is $f = 1/(1 + 10^{pK_a - pH})$.

(a) Compute pK_a at 15 °C. *Checkpoint:* 9.564.

(b) Compute f at pH 7.2 and 7.8, and the total ammonia at which the un-ionized ammonia reaches its limit of 0.025 mg N/L at each pH. *Checkpoint:* 0.43 % and 1.69 %; 5.81 and 1.48 mg/L.

(c) Compute the ratio of f at pH 7.75 to f at 7.2, and compare it with the 3.73-fold rise of the un-ionized peak in case 4. *Checkpoint:* 3.51; the total also rose, from 1.70 to 1.76 mg/L.

12.5 Theory and calculation · 8 min

Use the growth law of Equation (12.1) per fish, with the eaten feed equal to the ration, ρR_{sat}, and $q_f = 1$. The feed conversion is the feed over the growth.

(a) Show that the conversion is $FCR(\rho) = \rho / \{e_{max}[\rho - c\rho^2 - k]\}$ with $k = \rho_m(1 - c\rho_m)$. *Checkpoint:* Growth $= R_{sat} e_{max}(\rho - c\rho^2 - k)$.

(b) Find the ρ that minimizes the conversion for $c = 0.3$ and $\rho_m = 0.1$. *Checkpoint:* $\rho^* = \sqrt{k/c} = 0.569$.

(c) With $FCR_{sat} = 1.1$, compute e_{max}, the lowest conversion and how much better it is than the conversion at satiation. *Checkpoint:* 1.508; 1.0068; 8.47 %.

12.6 Theory and calculation · 5 min

In case 6 the power fails at 12:00. The tank holds 50 m^3 at a true oxygen of 8.98 mg/L, and the fish take up 0.529 kg/h. The pumps stop, so no water reaches the cone.

(a) Compute the oxygen in the tank above the 6 mg/L limit, in grams. *Checkpoint:* 149 g.

(b) Estimate the minutes to the limit at a constant uptake. *Checkpoint:* 16.9 min.

(c) Case 6 stocks 10,000 fish. Estimate the minutes for 20,000 fish of the same weight, starting from the same oxygen. *Checkpoint:* About 8.5 min.

12.7 App challenge · 9 min

Load case 3, The feed step: nitrite returns days later, and change no setting. Run the ten days at 0.25 d/s (40 s) with the default seed until the run stops at day 10.

(a) Read the Limits broken tile (hours and worst nitrite) and the fish killed by the water on the Harvest mass tile. *Checkpoint:* 57.0 h, 3.07 mg N/L, 873 fish, within 0.2 h, 0.02 and 2.

(b) Set Feed governor to On, press Reset and run the ten days again. Read the same values and the feed conversion (FCR) of both runs. *Checkpoint:* 20.7 h, 1.34 mg N/L, 1 fish; FCR 3.29 as shipped and 1.05 with the governor.

(c) The governor cuts the ration as the nitrite nears its limit. Explain why the limit is still broken for most of a day. *Checkpoint:* The four-hourly analyzer reports the nitrite after the meals that make it have been eaten.

12.8 App challenge · 6 min

Load case 6, Power failure: minutes to the oxygen limit, and change no setting. Run one day at 0.25 d/s with the default seed (about 4 s).

(a) Read the minutes from the failure to 6 mg/L in the note, and the worst oxygen on the Limits broken tile. *Checkpoint:* 16.7 min and 5.34 mg/L, within 0.1 and 0.02.

(b) Set Fish stocked to 20,000, press Reset and run the day again. Read the same two values, and the ratio of the two times. *Checkpoint:* 8.7 min and 2.15 mg/L; 1.92.

(c) The second run's Limits broken tile also reports about 16 h over a limit and a few hundred fish killed. Which limit caused that, and what does it say about a stocking density chosen for its oxygen reserve alone? *Checkpoint:* Nitrite after the meals; the biofilter, not only the oxygen, sets the density.

12.9 App challenge · 10 min

Load the Challenge, ninety days to harvest through a summer and an upset, and change no setting. Run at 2 d/s (45 s) with the default seed to day 90 and read the Challenge target tile.

(a) Record the harvest, the hours over a limit, the feed conversion, the fish killed and the margin. *Checkpoint:* 2.58 t, 69.8 h, 1.28, 1,359 fish, $1,326; not met.

(b) Press Reveal suggested tuning, then Reset, and run the ninety days again. Record the same five values. *Checkpoint:* 2.54 t, 6.6 h, 1.08, none, $2,148; met.

(c) The shipped run harvests more fish mass and fails. Explain both facts, and say which part of the suggested settings acts on the upset of day 62. *Checkpoint:* The shipped fish are fewer and larger; the governor and the economic ration cut the feed after the upset.

12.10 Application area · 3 min

A review of nitrite toxicity reports that each milligram per liter of chloride raises the trout's 96-hour LC50 by 0.29 to 2.0 mg N/L, and soft-water tests gave 0.19 to 0.39 mg N/L. The model assumes about 16 mg/L of chloride.

(a) Compute the LC50 range at 16 mg/L of chloride with the low slope, and with the high slope. *Checkpoint:* 4.83 to 5.03, and 32.19 to 32.39 mg N/L.

(b) The model uses an LC50 of 5 mg N/L and a limit of a tenth of it. Say why a farm adds salt, and why the limit uses the low slope. *Checkpoint:* Chloride protects against nitrite; the low slope is the cautious end.

12.11 Application area · 4 min

A farm manager asks how fast a ration may rise on a mature biofilter at 15 °C. Take the nitrite oxidizers' growth at its upper bound: substrate and oxygen saturating and no carrier limit, so the population grows at $\mu_N - (b_N + k_{det})$ with $\mu_N = 0.6 \times 1.06^{T-20}$, $b_N = 0.12 \times 1.029^{T-20}$ and $k_{det} = 0.05$ per day.

(a) Compute this net growth rate and the factor by which the population can grow in one day. *Checkpoint:* 0.294 d^{-1}; 1.34.

(b) Compute the shortest time for the population to triple, and compare it with the economic ration's ramp of at most 25 % a day. *Checkpoint:* 3.73 d; the ramp stays below the bound.

12.12 Review the agent · 5 min

A new farm runs case 2's startup on new media. On day 9 an AI assistant writes: "The total ammonia peaked at 8.09 mg N/L on day 8.8 and is now falling, so the biofilter is working. Stock the 10,000 fingerlings today and turn the feed governor on. If the water gets worse, the governor will cut the ration in time." Find the flaw and say how you would check it in the app.

CHAPTER 13

Greenhouse Climate Simulator for Temperature Integration Under a Humidity Limit



The controller moves 1 roof vent drive 2 thermal screen drive 3 boiler firing 4 fog valve 5 carbon dioxide dosing valve

The controller reads 1 measuring box, air temperature and humidity 2 carbon dioxide sample point 3 weather station

What disturbs it 1 sun, wind and outside cold

Two glasshouse spans of tomatoes with a roof vent, a thermal screen at the gutter, heating pipes along the rows, overhead fog and a carbon dioxide line from the plant room: notice that every handle acts on the one body of air that the measuring box samples. Illustration, not to scale.

IN PLAIN TERMS A Daily Heat Budget Under an Hourly Limit

The main idea. A tomato crop grows by the average warmth of its whole day, not by the warmth of each hour. So a glasshouse can let the free afternoon sun do some of the heating and burn less fuel at night.

The challenge. The air must also stay dry enough every hour, or water settles on the leaves. Opening a window or heating to dry the air uses up the saving, and the same window lets out the gas the plants feed on.

Why it matters. A glasshouse heated for most of the year saves fuel only if it also keeps its leaves dry.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the fact on an energy target (Section 13.1), the two history segments (Section 13.2), the two try-it segments and the fact on the dark day (Section 13.6), and the key points
Study	50 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	20 min	the first page, the cases and the graded week (Section 13.6), Technology in Practice (Section 13.7), Limitations (Section 13.8) and the open challenges

The control question A tomato crop answers the mean temperature of its day, so a glasshouse can spend its heat where it is cheapest, while the humidity under a closed screen must hold hour by hour. How much does a planner over the weather forecast save against fixed setpoints, and what does the humidity limit cost once one air exchange carries heat, water and carbon dioxide together?

Learning objectives

- Explain why a daily-mean recipe makes the heating setpoint a budget and the humidity limit a constraint.
- Compute the screen's insulation, its gap and the air exchange time of the house from the model's coefficients.
- Describe the pipe cascade with external reset and the three stages of the humidity override.
- Compare fixed setpoints with the receding-horizon planner on gas, daily mean and humid minutes in the app.
- Decide which combination of planner, override and cap meets the graded week, and price a point of humidity limit.

Prerequisites Sections 1.6.2 (velocity form and action), 1.6.1 and 1.6.4 (engine and scenarios) and 1.4.6 (overrides on limits). Chapter 11 for an economic supervisor above regulatory loops.

The app https://apmonitor.com/apps/greenhouse-climate/

See also Volume 9, Chapter 9 for a comfort band spent against a tariff, Volume 9, Chapter 6 for heat stored ahead of demand and Volume 9, Chapter 7 for a forecast whose error grows with lead time. Volume 4, Chapter 2 for a protective override and Volume 4, Chapter 4 for selectors on limits measured late.

13.1 Introduction

A glasshouse that grows tomatoes in a temperate climate is heated for most of the year, and its crop does not need every hour to be warm. Development and yield of a glasshouse tomato crop were found to depend mainly on the 24-hour average temperature, and a control program that let the air range between 14 and 26 °C around the same mean grew the crop no differently from a fixed regime [1]. The heating setpoint of such a house is therefore a budget for the day. A degree that the afternoon sun supplies for nothing can stand in for a degree the boiler would buy at night, provided the day's mean comes out right, and with a weather forecast the trade becomes a planning problem. Humidity does not trade that way. Water settles on the leaves in the hour the air saturates, whatever the afternoon was like, so the humidity limit has to hold hour by hour, and the three things that hold it, a gap in the screen, an open vent and more heat, all spend the budget.

A common reading of a climate computer is that it has done its job once its setpoints hold. On fixed setpoints of 19.5 by day and 17 °C at night it holds both, apart from a dawn dip of 1.4 K with the boiler at its limit, and the day's mean comes out at 19.02 °C on a bright Monday with 19.7 MJ/m^2 of sun and 18.57 °C on a dull Tuesday with 11.4 MJ/m^2, against a recipe of 19 °C. A planner that spends the same recipe over the forecast heats the bright Monday on 865 m^3 of gas instead of 1040 m^3, 17 % less, for a mean of 18.91 °C. The thermal screen that saves 622 of the 1188 m^3 a night would otherwise burn keeps the house above 85 % relative humidity for all twelve hours of that night, and opening every vent at noon cools the house by 3.2 K, dries it by 17 points of humidity and takes 85 ppm of carbon dioxide (CO_2) out within 15 min. Each handle on the climate moves all three of its variables.

The class of problem reaches well beyond glass. A product that responds to the integral of a variable, under hard limits on that variable and on others that share its actuators, is planned over a prediction and re-planned as measurements arrive, which is the receding-horizon structure of model predictive control, already surveyed in theory and in industrial practice by 1989 [2]. Horticulture made the structure explicit as a hierarchy in which an upper layer manages the crop and a lower one the climate, with predictions of the weather and of the auction price among its inputs [3], and a receding-horizon controller of a solar greenhouse was designed to keep temperature and humidity

within limits while satisfying a temperature integral [4]. A building held inside a comfort band, as in Volume 9, Chapter 9, or a store charged against a tariff has the same shape, an amount that can be moved between hours bound to a limit that cannot, and a transfer function from boiler to air contains neither.

> **+ DID YOU KNOW? A Target of 65 Percent**
>
> Under agreements between the Dutch grower association and the government, glasshouse horticulture aimed to cut its energy efficiency index by 65 % between 1980 and 2010. Climate regimes such as temperature integration were designed to reach savings with the equipment growers already had [5].

However, optimal control of a greenhouse needs a quantitative model of how its crop production responds to the controls and the weather [3], while the climate computers growers buy come with hundreds of settings and courses on how to adjust them but no account of when or why [6], so there is no open plant on which the budget, the limit and the forecast can be set against one another while the gas meter keeps count. This chapter supplies one. It balances temperature integration against a humidity limit and counts the cost, on a browser glasshouse whose ledgers of energy, water, CO_2 and assimilates close to round-off and whose planner sees only the forecast and what the climate computer has measured. The climate and crop model reduces the Vanthoor models [7, 8], the vent law follows de Jong [9], the stomatal resistance follows Stanghellini [10], and temperature integration is established practice. The chapter gives a 41-state model in enough detail to rebuild, a climate computer that reads only its instruments, and a planner written as a linear program of 165 to 286 variables whose cost curves come from a quasi-steady model of each hour. Its results are comparisons on one plant. Among them, the planner lowers the cost of the graded week by \$656 to \$725 against fixed setpoints under the same humidity override at every limit from 78 to 90 %, each point of limit costs between \$76 and \$113, and in that week the CO_2 cap the challenge suggests changes nothing once the planner and the override run.

Section 13.2 follows the domain from hand-set vents to a planned daily mean. Section 13.3 gives the model of the house, the crop, the supply and the instruments, with the time scales its equations imply. Section 13.4 describes the climate computer: the pipe cascade with external reset, the vents under a three-stage humidity override, the screen, the CO_2 loop and the fog. Section 13.5 sets out the planner's linear program, Section 13.6 reports the nine cases and the graded week, and Sections 13.8 and 13.9 state the limitations and the conclusions, after Section 13.7 names the devices and the controller platform of a real glasshouse. The exercises of Section 13.10 run the cases in the app.

13.2 From Hand-Set Vents to a Planned Daily Mean

Control of the glasshouse climate began as the grower's own work. A review of heating and ventilation control in 1968 found many English growers still without automatically fired boilers or controlled heating, judged automatic ventilation hard to justify on the labor it saved, and reported that the automatic systems of the day held the dry-bulb temperature reasonably precisely while the grower still chose the temperature level from the weather and the state of the plants and corrected what those systems did to the humidity [11]. The same review found each factor of the climate controlled in isolation, because growers added equipment piecemeal as plant research advanced and capital allowed. Heat loss then became a subject of its own. A symposium on energy saving in greenhouses heard in 1977 that thermal screens drawn over the crop at night cut the thermal transmittance of glasshouse compartments by 28 to 43 % at zero wind speed, depending on the film, the largest saving coming from an aluminized polyester not yet durable enough for commercial use [12]. The first climate computers reached growers in the same decade, as the segment below tells.

HISTORY A Computer at the Growers' Fair (1974)

A computer was first shown to Dutch growers as the heart of a greenhouse climate control system at the annual horticultural fair of 1974 in Bleiswijk. Research at that moment was still aimed mostly at analog control, and seventeen electronic firms were selling analog controllers to growers. The first systems sold so well that competitors at the fair started at once to develop digital climate computers of their own. Within two years systems from five producers were on sale, and by the 1980 report eleven firms, nine of them Dutch, competed for the market [13].

Research took a system view. A paper of 1978 described greenhouse culture as three levels, the instantaneous climate, short-term plant growth and long-term crop development, each controlled through relatively simple models, with the computer essential at the first level and the third, where predictions of cost and return seemed attractive, left to the grower's final decision [14]. Theses at the agricultural university in Wageningen then supplied the physics, among them the ventilation of large multi-span houses through windows on the lee and windward sides, driven by the wind and by the temperature difference and measured at full scale [9], and a model of the transpiration of a tomato crop offered as the quantitative basis for controlling the air's humidity [10]. A thesis of 1994 on CO_2 measured in greenhouse fruit vegetable crops that each additional 100 µmol/mol raised the light-use efficiency by about 10 to 15 %, and that intermittent supply raised the average concentration only a little under normal ventilation [15], which is the coupling between the vents and the dosing line.

Temperature integration grew from the energy question, as the segment below tells.

HISTORY Temperature Integration (1988 to 2003)

Experiments had shown that the development and yield of a glasshouse tomato depend mainly on the 24-hour average temperature. A control program of 1988 let the air range between 14 and 26 °C, not limited to a 24-hour period, and December-sown tomatoes grew no differently from a fixed regime with the same mean [1]. A thesis of 2003 nested the integration in windows of 24 hours and six days, predicted up to 9 % less energy than regular integration for a year-round tomato crop, and found the potential limited by humidity control on fixed setpoints. A humidity regime built on the processes of cut chrysanthemum saved 18 % of the energy of a fixed 80 % setpoint in simulation, the two combined more than 33 %, and the only addition was a computer acting as a setpoint generator above the existing climate computer [5].

Model-based optimal control of the crop took the same path. A thesis of 1994 found the optimal management systems of the literature presented as two-layer hierarchies and set out to include humidity explicitly and to cope with errors in the model and in the predictions of weather and price [3]. A receding-horizon optimal controller meant to hold a solar greenhouse within temperature and humidity limits and a temperature integral used 77 % less gas, in simulations of a simplified version, than a conventional greenhouse under optimal control [4], and a monograph collected the methods of the optimal control of greenhouse cultivation [6]. Energy balances of a reference tomato house in the Netherlands put the savings of a higher humidity setpoint, a stepwise screen gap and temperature integration at about 5 % each, against 23 % for a better-insulated cover [16]. The humidity limit and the temperature budget stayed in tension throughout, and the simulator of this chapter puts a price on that tension. Staffing is a further pressure. The greenhouse industry finds it hard to staff the management of its crops, and in an international competition five teams grew cucumbers remotely on setpoints chosen by their own artificial intelligence, one of them outperforming the manually grown reference [17].

13.3 One Hectare Under Glass

The simulated house is one hectare of Venlo glasshouse growing tomatoes at a leaf area index of 3, modeled per square meter of floor as a single well-mixed air node of mean height $h = 5\,\mathrm{m}$ under a single-glass cover, as shown in Figure 13.1. A boiler of 150 $\mathrm{W/m^2}$ heats water in pipes along the crop

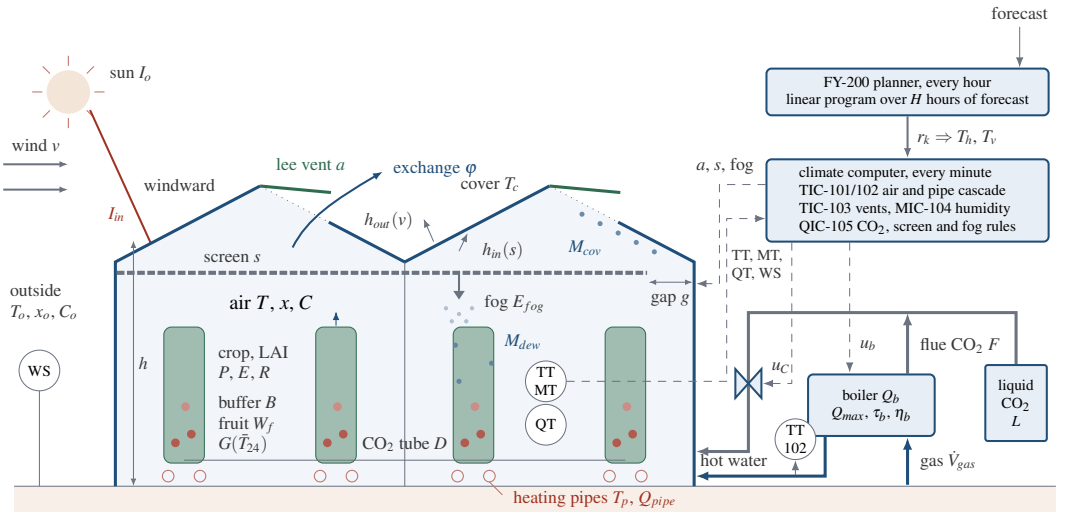


Figure 13.1 The simulated glasshouse in cross-section, two of its Venlo spans with the wind v from the left, beside the plant room and the climate computer. The air node at T, x and C, of mean height h, exchanges φ with the outside air at T_o, x_o and C_o through the vents of aperture a (lee side first) and the leaks, and loses heat through the cover at T_c by $h_{in}(s)$ inside and $h_{out}(v)$ outside, where water condenses as M_{cov}. The screen closes to s at the gutter with a humidity gap g. Sunlight I_o enters as I_{in}, the crop transpires E, photosynthesizes P, respires R, fills the buffer B and grows fruit W_f at a rate $G(\bar{T}_{24})$ set by its trailing daily mean, and dew M_{dew} settles on it when the air saturates. The boiler fires at u_b to deliver Q_b into pipes at T_p that give Q_{pipe} to the air, and the dosing valve u_C draws the dose D from the flue gas F first and the liquid store L for the rest. The climate computer reads the measuring box (TT, MT), the CO_2 analyzer (QT), the pipe sensor (TT-102) and the weather station (WS), and the planner above it writes the hourly temperatures r_k. Not to scale.

rows, roof vents open on the lee side first and then on the windward side, a thermal screen draws across at the gutter, a dosing line of 130 kg/(ha h) brings CO_2 from the boiler's flue gas and from a liquid store, and high-pressure fog evaporates up to 150 g/(m^2 h). The ten physical states are the air, cover and pipe temperatures T, T_c and T_p, the boiler's heat Q_b, the vapor concentration x, the CO_2 concentration C, the screen closure s, the vent aperture a, the assimilate buffer B and the fruit dry matter grown W_f. Thirty-one more states integrate every flux into a ledger, so the energy, vapor, CO_2 and assimilate balances and the costs are exact integrals of the model. Table 13.1 gives each signal's role and its value at a settled start.

13.3.1 Heat in the Air, the Cover and the Pipes

The air node gains the heat of the pipes, the sun it absorbs and the latent heat of any dew, and loses heat to the cover, to the exchange and to the water the crop and the fog evaporate, as shown in Equation (13.1).

$$C_{air}\frac{\mathrm{d}T}{\mathrm{d}t} = Q_{pipe} + \alpha_s I_{in} + \lambda M_{dew} - h_{in}(s)\,(T - T_c) - \rho c_p\,\varphi\,(T - T_o) - \lambda\,(E + E_{fog}) \quad (13.1)$$

where every flux is per square meter of floor, t is in seconds (the application integrates in hours and applies the factor of 3600 in one place), C_{air} lumps the air, the crop and the light structure, α_s is the share of the transmitted sun the house absorbs, λ the latent heat of water and ρc_p the volumetric

Table 13.1 Role and Starting Value of Each Signal, From the Settled Start of Case 2

Signal	Role	Start	Range
T	state, CV, TT-101	17.00 °C	band 16–26
T_c	state, cover	9.97 °C	—
T_p, Q_b	states, CV of TIC-102	42.8 °C, 50.0 W/m^2	15–90, 0–150
x	state, MT-101 reads RH	82.5 % RH	limit 85, 90
C	state, CV, QT-101	978 ppm	SP to 800
s, a	states, drives	98.9 %, 0	0–100 %
B, W_f	states, crop	4 g/m^2, 0	—
—	31 ledger states	0	—
$T_{p,SP}$	MV of TIC-101	—	15–90 °C
u_b	MV of TIC-102, firing	—	0–100 %
a_T, a_H	MVs of TIC-103, MIC-104	—	0–100, 0–50
s_{cmd}, fog	MVs of the rules	—	0–100, on/off
u_C	MV of QIC-105, dosing	—	to the supply
T_h, T_v	heating, vent setpoints	17.0, 19.0	—
r_k	planner's hourly temperature	—	16–26
T_o, I_o, v	DVs, weather station	6.11 °C, 0, 2.48 m/s	a week
—	DVs, offsets	0	±10 K, 0–150 %
—	DVs, events	—	7 kinds
p_{gas}, p_C	DVs, prices	0.40 \$/m^3, 0.15 \$/kg	—

heat capacity of air. The cover warms from the air inside, from the water condensing on it and from the share α_c of the sun it absorbs, and it loses to the outside by wind and radiation lumped in one coefficient, as shown in Equation (13.2).

$$C_{cov}\frac{\mathrm{d}T_c}{\mathrm{d}t} = h_{in}(s)(T - T_c) + \lambda M_{cov} + \alpha_c I_o - h_{out}(v)(T_c - T_o) \tag{13.2}$$

where $h_{out} = 7 + 2v$ in W/(m^2 K) with the wind v in m/s. The screen is a resistance R_s in the inside path, and a gap costs more than its area because warm air escapes through it, as shown in Equation (13.3).

$$\begin{aligned} h_{in}(s) &= \left[\frac{1}{h_{in,0}} + \left(1 - (1-s)^{0.7}\right)R_s\right]^{-1} \\ h_{in,0} &= \left[\frac{1}{U_{open}} - \frac{1}{h_{out}(4)}\right]^{-1}, \qquad R_s = \frac{1}{U_{scr}} - \frac{1}{U_{open}} \end{aligned} \tag{13.3}$$

so the declared cover coefficients of 7.5 with the screen open and 4.0 W/(m^2 K) under it at a 4 m/s wind give $h_{in,0} = 15\,\mathrm{W/(m^2\,K)}$ and $R_s = 0.117\,\mathrm{m^2\,K/W}$, and a 10 % gap keeps 80 % of the screen's insulation. The sun reaching the crop is $I_{in} = \tau_c\,(1 - s\,(1 - \tau_s))\,I_o$. The boiler delivers its firing behind a lag, and the pipes give heat to the air by natural convection and radiation lumped together in Equation (13.4).

$$\begin{aligned} \tau_b\frac{\mathrm{d}Q_b}{\mathrm{d}t} &= u_b\,Q_{max} - Q_b, \qquad C_{pipe}\frac{\mathrm{d}T_p}{\mathrm{d}t} = Q_b - Q_{pipe} \\ Q_{pipe} &= k_p\,|T_p - T|^{1.25}\,\mathrm{sgn}(T_p - T), \qquad \dot{V}_{gas} = \frac{Q_b}{\eta_b\,\mathrm{LHV}} \end{aligned} \tag{13.4}$$

where $u_b \in [0, 1]$ is the firing command, which a boiler outage overrides with zero heat whatever TIC-102 asks, and k_p is set so that pipes at 80 °C give 150 W/m^2 to air at 18 °C. Burning the gas makes 1.78 kg of CO_2 per cubic meter, so a firing boiler offers its flue gas to the dosing line at $f_g Q_b$ with f_g equivalent to 2.20 kg/(ha h) per 1 W/m^2, up to $F_{max} = 50\,\mathrm{kg/(ha\,h)}$, which it reaches at 22.7 W/m^2. A boiler at full fire makes 330 kg/(ha h), and at night, when nothing is dosed, all of it goes up the stack.

13.3.2 Water, Carbon Dioxide and the Exchange

The exchange φ, in cubic meters of air per square meter of floor per second, enters the vapor and the CO_2 balance alike in Equation (13.5).

$$\begin{aligned} h\frac{\mathrm{d}x}{\mathrm{d}t} &= 10^3\,(E + E_{fog} - M_{cov} - M_{dew}) - \varphi\,(x - x_o) \\ h\,\kappa\frac{\mathrm{d}C}{\mathrm{d}t} &= D - \mu_C P + R - \varphi\,\kappa\,(C - C_o) \end{aligned} \tag{13.5}$$

where x is in g/m^3, the water fluxes in kg/(m^2 s), C in ppm with $\kappa = 1.829\,\mathrm{mg/m^3}$ per ppm, D the dose and R the crop's respiration in mg/(m^2 s), and $\mu_C P$ the CO_2 taken up by a gross photosynthesis P in µmol/(m^2 s), with $\mu_C = 44.01 \times 10^{-3}$ mg/µmol. Condensation on the cover and dew on the crop follow Equation (13.6).

$$\begin{aligned} M_{cov} &= 6.4 \times 10^{-9} \cdot 4 \cdot \pi(s)\left(e(x,T) - e_s(T_c)\right)^+ \\ M_{dew} &= \frac{h}{1.2 \times 10^5}\left(x - x_s(T)\right)^+ \end{aligned} \tag{13.6}$$

in which the vapor pressure e and the saturation pressure e_s come from Tetens' formula in Pa, the coefficient of four is a heat exchange coefficient in W/(m^2 K) in the form the application takes from the Vanthoor model [7], x_s the saturation concentration and $(\cdot)^+ = \max(0,\cdot)$. Dew is a numerical stand-in that removes any supersaturation within about two minutes. The exchange is the vents' law, reduced from the measurements of Venlo windows [9], plus the leaks, both scaled by what the screen lets through, as shown in Equation (13.7).

$$\begin{aligned} \varphi &= \varphi_v + \varphi_l \\ \varphi_v &= \frac{\pi(s)\,\varphi_{max}}{3600}\left(0.4 f(a_l) + 0.6 f(a_w)\right) \\ &\quad \times \frac{1}{4}\sqrt{v^2 + 0.5\,(T - T_o)^+} \\ \varphi_l &= \pi(s)\,\frac{n_l h}{3600}\,\frac{v + 0.5}{4.5} \\ f(a) &= \frac{1 - \mathrm{e}^{-3a}}{1 - \mathrm{e}^{-3}} \\ a_l &= \min(1, 2a), \qquad a_w = (2a - 1)^+ \\ \pi(s) &= p_s + (1 - p_s)(1 - s)^{0.4} \end{aligned} \tag{13.7}$$

where the lee window opens over the first half of the aperture and the windward one over the second, $\varphi_{max} = 60\,\mathrm{m^3/(m^2\,h)}$ is the exchange at full aperture in a 4 m/s wind, 24 m^3/(m^2 h) of it through the lee windows alone, the buoyancy term carries a calm day, $n_l = 0.5$ air changes an hour is the leakage at 4 m/s, and a closed screen passes $p_s = 15\,\%$ of the exchange. A 10 % gap passes 49 %.

13.3.3 The Crop

Equation (13.8) gives the transpiration, the Penman–Monteith combination with the canopy at the air temperature and Stanghellini's stomatal resistance [10].

$$\begin{aligned} \lambda E &= \frac{\Delta R_n + \rho c_p\left(e_s(T) - e\right)/r_a}{\Delta + \gamma(1 + r_s/r_a)}, \qquad r_a = \frac{r_b}{2\,\mathrm{LAI}} \\ r_s &= \frac{r_{min}}{\mathrm{LAI}}\,\frac{I_l + 4.3}{I_l + 0.54}\left(1 + 2\,\mathrm{e}^{-I_l/2}\right) f_C(C) \end{aligned} \tag{13.8}$$

with Δ the slope of the saturation curve, $\gamma = 66.5\,\mathrm{Pa/K}$, $S_c = 0.85\,(1 - \mathrm{e}^{-0.7\,\mathrm{LAI}})\,I_{in}$ is the sunlight the canopy absorbs, $R_n = S_c + f_p Q^{+}_{pipe}$ adds the pipes' radiation, $I_l = S_c/\mathrm{LAI}$, the stomata close three times further in the dark, and $f_C = 1 + 6.08 \times 10^{-7}(\tilde{C} - 200)^2$ while S_c exceeds $5\,\mathrm{W/m^2}$, with $\tilde{C}$ the CO_2 clipped to 200 to 1100 ppm, and $f_C = 1$ otherwise. Gross photosynthesis is a big-leaf light response whose ceiling saturates in CO_2 and whose slope falls toward the compensation point, as shown in Equation (13.9).

$$\begin{aligned} P &= \sigma(B)\,P_c\Big[1 - \exp\Big(-\frac{\varepsilon\,\Phi}{P_c}\Big)\Big] \\ P_c &= P_m f_T(T)\,\frac{C^*}{C^* + K_C}, \qquad \varepsilon = \varepsilon_0\,\frac{C^* - \Gamma}{C^* + 2\Gamma} \\ C^* &= \max(C, \Gamma), \qquad \Gamma = 2T + 5 \\ \Phi &= 2.3 \cdot 0.93\,(1 - \mathrm{e}^{-0.7\,\mathrm{LAI}})\,I_{in} \end{aligned} \tag{13.9}$$

where the absorbed photosynthetically active radiation Φ is in $\mu\mathrm{mol/(m^2\,s)}$, $f_T(T) = \mathrm{clip}\big((T - 6)/10\big)\,\mathrm{clip}\big((38 - T)/8\big)$ with clip limiting to $[0, 1]$, and $\sigma(B) = 1 - \mathrm{clip}\big((B - 20)/10\big)$ stops the uptake as a full buffer runs out of sinks. At $500\,\mathrm{W/m^2}$ outside and 21 °C the canopy photosynthesizes 23.6 % faster at 800 ppm than at the outdoor 420 ppm. The assimilates fill the buffer, and growth draws on it at a rate set by the trailing 24-hour mean $\bar{T}_{24}$, as shown in Equation (13.10).

$$\begin{aligned} \frac{\mathrm{d}B}{\mathrm{d}t} &= m_H P - R_m - G, \qquad \frac{\mathrm{d}W_f}{\mathrm{d}t} = f_f\,\eta\,G \\ R_m &= \frac{r_m W_c}{86400}\,2^{(T-25)/10}\,\frac{B}{B + 0.2} \\ G &= \frac{g_{19}}{86400}\,\mathrm{clip}_{16/9}\Big(\frac{\bar{T}_{24} - 10}{9}\Big)\,\frac{B}{B + 2} \\ &\quad \times \mathrm{clip}\Big(\frac{T - 8}{4}\Big)\,\mathrm{clip}\Big(\frac{36 - T}{4}\Big) \end{aligned} \tag{13.10}$$

where B is in $\mathrm{g/m^2}$ of assimilates, $m_H = 30.03 \times 10^{-6}\,\mathrm{g/\mu mol}$, $\mathrm{clip}_{16/9}$ limits to $[0, 16/9]$, and the respiration in Equation (13.5) is $R = (R_m + (1 - \eta)G)\,\mu_C/m_H$ in $\mathrm{mg/(m^2\,s)}$, maintenance plus the share of growth not converted to dry matter. The demand is $18\,\mathrm{g/(m^2\,d)}$ at a mean of 19 °C, 12 at 16 and 24 at 22, so the crop answers the mean, not the hour. The mean $\bar{T}_{24}$ is the average of the last 288 five-minute means of the true air temperature.

🕮 **Pause.** Try Exercise 13.1 before reading on.

13.3.4 Supply, Costs, Weather and Instruments

The dosing line takes flue gas first and buys liquid CO_2 for the rest, $D = \min(u_C D_{max}, F + L)$ with $F = \min(F_{max}, f_g Q_b)$ while the boiler fires and $L = D_{max}$ when liquid is allowed, of which $\min(D, F)$ is free. The cost integrates $p_{gas}\dot{V}_{gas}$, p_C times the liquid dosed, and electricity at 0.15 $/kWh for $0.5\,\mathrm{W/m^2}$ of base load, $0.6\,\mathrm{W/m^2}$ of pumps once the boiler delivers $20\,\mathrm{W/m^2}$, $1.5\,\mathrm{W/m^2}$ while the fog runs and $0.2\,\mathrm{W/m^2}$ for the dosing fan at full rate.

The weather is a synthetic spring week at 52° N in mid-April. Each day of a table of nine sets a dawn minimum and an afternoon maximum joined by half-cosines at 05:00 and 15:00, a mean sky transmissivity with seeded smooth variation, a wind and a dew point below the dawn minimum, and the sun is clear-sky radiation from the solar elevation scaled by the transmissivity, near $700\,\mathrm{W/m^2}$ at a clear noon. A front blends one day's air mass into the next over about an hour and a half with a band of rain. The forecast issued at hour t_0 for hour t is the truth plus an error of seeded smooth noise in t whose size grows as $\min(t - t_0, 48)/24$ times 1.2 K, 0.18 of transmissivity and 1.2 m/s at the declared size, so successive forecasts of one hour converge as it nears. A missed front is a forecast that places it

Table 13.2 Model Constants by Group, Marked V When the Show Model Dialog Offers a Slider, H When Hidden From It and C When Written Into the Code

Symbol	Meaning	Value	Class
House			
h	mean air height	5.0 m	H
C_{air}, C_{cov}	air with crop and structure, cover	30, 10 kJ/(m^2 K)	V, H
U_{open}, U_{scr}	cover, screen open and closed	7.5, 4.0 W/(m^2 K)	V
τ_c, τ_s	light transmission, cover and screen	0.78, 0.80	V
p_s	exchange a closed screen passes	0.15	V
α_s, α_c	sun absorbed in the house, cover	0.90, 0.05	H
n_l, φ_{max}	leakage, vents at full	0.5 h^{-1}, 60 m^3/(m^2 h)	V
—	vent and screen drive lags	2, 3 min	H
Heating and CO_2			
Q_{max}, η_b, τ_b	boiler capacity, efficiency, lag	150 W/m^2, 0.92, 5 min	V, V, H
C_{pipe}, k_p	pipe water and steel, transfer	12, 0.862	V, H
$T_{p,max}$, f_p	highest pipe, pipe radiation share	90 °C, 0.2	V, H
LHV	natural gas, lower heating value	31.65 MJ/m^3	C
D_{max}, F_{max}	line capacity, flue share	130, 50 kg/(ha h)	V
C_o	outdoor CO_2	420 ppm	V
Crop			
LAI	leaf area index	3	V
P_m, K_C	photosynthesis ceiling, half point	55 µmol/(m^2 s), 400 ppm	V
ε_0	light-use efficiency	0.08 mol/mol	V
r_m, W_c	maintenance at 25 °C, standing crop	0.01 d^{-1}, 800 g/m^2	V, H
g_{19}	growth demand at 19 °C	18 g/(m^2 d)	V
f_f, η	fruit share, conversion	0.6, 0.7	H
r_{min}, r_b	stomatal, boundary-layer resistances	82, 275 s/m	H
Fog, prices and constants			
—	fog capacity	150 g/(m^2 h)	V
p_{gas}, p_C	gas, liquid CO_2	0.40 $/m^3, 0.15 $/kg	control
ρc_p, λ, γ	air, latent heat, psychrometer	1206, 2.45 MJ/kg, 66.5	C

k_p in W/(m^2 K$^{1.25}$), C_{pipe} in kJ/(m^2 K), ρc_p in J/(m^3 K), γ in Pa/K.

Δ hours late until it is three hours away and blends toward the truth from six hours away. The climate computer reads the air temperature and humidity in an aspirated box behind a 1 min first-order lag with 0.05 K and 0.5 points of noise, the CO_2 behind the 2 min lag of its sample line with 5 ppm of noise, the pipe water with 0.1 K, the weather station with 0.1 K, 2 % of radiation, 0.2 m/s and one point of humidity, and the drives' own positions and the boiler's meter without error, the noise given as standard deviations and drawn, like the weather's, from a generator seeded at 20261009.

13.3.5 Starting State and Numerics

Every case starts at midnight with the air at its night setpoint and a trailing mean of 19 °C, and the rest of the house is solved at rest for the weather of that instant. The cover, the vapor, the CO_2 and the pipe temperature that hold the air still are found by a Gauss–Seidel sweep with a bisection for the vapor, the screen is set by its rules, and where the screen is closed and the override is on, its humidity gap is placed at the fixed point of its proportional rule. Case 2 starts this way at 17.00 °C and 82.5 % with the screen 98.9 % closed, a gap of 1.06 %, 978 ppm of CO_2 gathered from the night's respiration and the boiler at 50.0 W/m^2. A fixed 30 s step of the fourth-order Runge–Kutta method advances the 41 states, the climate computer runs every minute, the planner every simulated hour, and the historian logs a row every 5 min, 2017 rows for a week, on the engine of Section 1.6.1. A week runs in 28 s at the top speed. Halving the step changes the cost of the graded week with the suggested settings by 0.0003 %, its gas by 0.0007 % and no day's mean by more than 0.000 03 K, and over that week the seven balances close to 3.6×10^{-14} of their throughput, 1.3×10^{-14} as shipped. The application passes its 35 verification checks.

13.3.6 What the Equations Imply

The exchange sets the clock of the house. At a 4 m/s wind and a 5 K difference, h/φ is 120 min with the vents shut, 10.6 min with the lee windows open and 4.5 min with every window open, while under a closed screen the shut house exchanges its air once in 800 min. Water and CO_2 follow that clock, which is why a closed screen traps the crop's night transpiration and a wide roof empties the CO_2 in minutes. The air node against the cover and the exchange answers in 60 min shut, 121 min screened and 17 min with the vents full, the cover in 5.6 min, and the pipes, whose transfer coefficient is $\mathrm{d}Q_{pipe}/\mathrm{d}T_p = 2.3$ to 3.0 W/(m^2 K) at 20 to 60 K above the air, in 88 to 67 min, so heat bought now keeps arriving for an hour after the boiler stops. The crop answers a day.

13.3.7 Assumptions

One air node. The house has no vertical gradients, no separate canopy, soil or structure temperature and no cold air falling through a screen gap, and the space above the screen is lumped with the cover. The lumped capacity makes the house answer a little faster and more evenly than a real one.

A reduced crop. The crop is one big leaf at a fixed leaf area index, after the Vanthoor tomato model [8] reduced to a buffer and a growth rate on the daily mean. Fruit set, truss development, leaf picking, quality and the crop's tolerance of several days off its mean are absent, and the integration window is the calendar day.

Reduced vents, screen and cover. One law serves the lee windows and one the windward, the screen's exchange and gap are concave fits rather than a buoyant flow through the gap, and the cover loses heat to the outdoor air temperature, so a clear night's colder sky is absent.

A synthetic week and a declared forecast. The weather is generated, fronts are blends between air masses and the forecast error is declared, not learned, and a real forecast fails in more ways than a front placed late.

An idealized climate computer. Sensors are aspirated and seeded, actuators do what they are told within their drive times, and the boiler fails only when the outage is fired. Every limit, price and coefficient is a teaching value, not a calibration of any house.

13.4 The Climate Computer as Implemented

The climate computer executes every minute from its instruments alone, in the order of the instruments, the temperature strategy, the heating cascade, the vents with the humidity override, the screen, the CO_2 loop and the fog. Five loops run the framework's velocity-form proportional-integral-derivative (PID) law of Equation (13.11), none with derivative action and TIC-101 with its integral replaced by the external reset of Section 13.4.1.

$$\begin{aligned} e_k &= \gamma_a\,(\mathrm{SP}_k - \mathrm{PV}_k), \qquad P_k = K_c\,e_k \\ S_k &= S_{k-1} + (P_k - P_{k-1}), \qquad v_k = u_0 + S_k + I_{k-1} \\ u_k &= \mathrm{clamp}(v_k, u_{min}, u_{max}) \\ I_k &= I_{k-1} + \left[\frac{K_c}{\tau_I}\,e_k + \frac{u_k - v_k}{\tau_I}\right]\Delta t_c \end{aligned} \tag{13.11}$$

which is the law of Section 1.6.2 with no derivative term, written with γ_a for the action a, $u_0 = 0$ for the bias u_b, $\Delta t_c = 1$ min for the control period and the integral time as the tracking constant. The limit acts on the output alone and the anti-windup on the integral alone, and a return from MANUAL to AUTO causes no jump. The loops and their tunings are listed in Table 13.3, where $\gamma_a = +1$ is reverse acting and -1 direct acting, the convention of Section 1.3. The vent loops are direct acting because an open vent lowers both the air temperature and the humidity.

13.4.1 Temperature Strategy and the Heating Cascade

On fixed setpoints the heating setpoint is $T_h = T_{night} + (T_{day} - T_{night})\,\mathrm{clip}(\theta - 6)\,\big(1 - \mathrm{clip}(\theta - 19)\big)$ at clock hour θ, ramped over an hour at 06:00 and 19:00, and the vents work at $T_v = T_h + \Delta T_v$ with

Table 13.3 The Loops as Executed, Each on the Law of Equation (13.11) With No Derivative and a 1 min Period

Loop	CV → MV	K_c	τ_I	Limits	γ_a
TIC-101	$T \to T_{p,SP}$	8	60 min	15–90 °C	+1
TIC-102	$T_p \to u_b$	8	15 min	0–100 %	+1
TIC-103	$T \to a_T$	15	15 min	0–100 %	−1
MIC-104	RH → a_H	4	20 min	0–50 %	−1
QIC-105	$C \to u_C$	0.3	10 min	0–supply	+1

K_c in K/K, %/K, %/K, % per point and %/ppm. TIC-101 uses external reset.

$\Delta T_v = 2\,\mathrm{K}$, so between the two the house floats on the sun. With the planner on, the setpoints come from Section 13.5. The air loop TIC-101 does not fire the boiler. It asks the pipe loop TIC-102 for a water temperature, and its reset is a lag of the pipe temperature the slave achieves rather than of its own output, the external reset feedback of Equation (13.12).

$$\begin{aligned} I_k &= I_{k-1} + \frac{\Delta t_c}{\tau_{I,T}}\left(\tilde{T}_{p,k} - I_{k-1}\right) \\ T_{p,SP,k} &= \mathrm{clamp}\left(K_{c,T}\left(T_{h,k} - \hat{T}_k\right) + I_k,\ 15,\ 90\right) \end{aligned} \tag{13.12}$$

where $\hat{T}$ is the measuring box and $\tilde{T}_p$ the measured pipe temperature clipped to 15 to 90 °C. While the pipes follow, the loop is an ordinary PI controller. While the boiler is at its limit, out, or slower than asked, the reset follows the pipes there are, so the master cannot wind up while the water takes an hour to heat. Volume 3, Chapter 4 uses the same external reset feedback behind a high select. With the slave in MANUAL the master tracks the measured pipe temperature, so the cascade closes without a bump.

🕮 **Pause.** Try Exercise 13.3 before reading on.

13.4.2 Vents, the Humidity Override and the Screen

TIC-103 opens the vents as the measured air rises above T_v, and MIC-104 opens the lee windows, up to half the aperture, as the measured humidity rises above a limit RH_{sp} that is $\mathrm{RH}_{lim} = 85\,\%$ in the dark and rises linearly to $\mathrm{RH}_{day} = 90\,\%$ as the measured radiation outside goes from 50 to 150 W/m^2. A high select drives the vents with $a = \max(a_T, a_H)$, and the loop not selected has its integral set so that its output is the command plus its own proportional term, $I_T = a$ while MIC-104 is selected and $I_H = \min(a, 50)$ otherwise, so either can take over without a bump. The override has three stages. While the screen is closed, a gap opens in it in proportion to the humidity from three points below the limit, $g = \mathrm{clamp}\left(K_g(\widehat{\mathrm{RH}} - \mathrm{RH}_{lim} + 3),\ 0,\ g_{max}\right)$ with $K_g = 2$ and $g_{max} = 10\,\%$. The lee vent is the second stage, and the third is heat, as shown in Equation (13.13).

$$\begin{aligned} \Delta T_{h,k} = \mathrm{clamp}\Big(&\Delta T_{h,k-1} + 2\Delta t_c\big[\sigma_H\,\mathrm{clip}(\widehat{\mathrm{RH}} - \mathrm{RH}_{sp}) \\ &- \mathrm{clip}(\mathrm{RH}_{sp} - 1 - \widehat{\mathrm{RH}})\big],\ 0,\ 5\Big) \end{aligned} \tag{13.13}$$

where Δt_c is in hours here, $\sigma_H = \mathrm{clip}\left((a_H - 45)/5\right)$ engages the stage as the lee vent nears its half-open limit, and ΔT_h raises both T_h and T_v by up to 5 K at up to 2 K/h, letting the humidity outrank the recipe. With the override off, or the heating or the vents in MANUAL, the term falls at 2 K/h. The screen closes when the radiation outside is under $I_{scr} = 10\,\mathrm{W/m^2}$ and the air outside under $T_{scr} = 14\,°\mathrm{C}$, opens when the radiation passes $5I_{scr}$ or the air outside passes $T_{scr} + 2\,\mathrm{K}$, moves at most once in 20 min, opens in proportion to the light from 5 to 15 times I_{scr} while it is cold outside, and never opens faster than over half an hour.

13.4.3 Carbon Dioxide and Fog

QIC-105 doses toward a setpoint C_{sp} that rises with the measured radiation from the outdoor 420 ppm above 20 W/m^2 to $C_{SP} = 800$ ppm at 200 W/m^2, and it doses nothing below 20 W/m^2. With the cap on and the vents open past $a_{cap} = 50\,\%$, the setpoint falls toward 50 ppm above outdoors over the next 30 points of aperture, $C_{sp} \leftarrow C_{sp} - (C_{sp} - \min(C_{sp}, C_o + 50))\,\mathrm{clip}((a - a_{cap})/0.3)$. The output is limited to what the line can deliver at that moment, flue gas while the boiler fires and liquid when it is allowed. The fog runs while the measured air is 2 K above T_v with the vents past 90 % and the humidity under 80 %, stops 1 K lower or three points wetter, and switches at most once in 5 min.

13.5 The Planner: A Receding-Horizon Linear Program

Every simulated hour, and whenever a price, the recipe or a limit changes, the planner FY-200 chooses one planned temperature r_k for each of the next $H = 24$ hours of a fresh forecast, uses only the first, and solves again an hour later from the hours already measured. Its house model is quasi-steady for each hour of the forecast, as shown in Equation (13.14).

$$\begin{aligned} q(r) &= \big(U(s,v) + \rho c_p \varphi_l\big)(r - T_o) + \lambda E - \alpha_s I_{in} \\ Q &= q(r) + \rho c_p\,(\varphi_H - \varphi_l)^+ (r - T_o)^+ \quad \text{if } q \geq 0 \\ \varphi_V &= -q(r)/\big(\rho c_p\,(r - T_o)\big) \quad \text{if } q < 0 \\ c_k(r) &= 3.6 \times 10^7 \Big[p_{gas} \frac{Q}{\eta_b\,\mathrm{LHV}} + 10^{-6}\, p_C\,(D_k - F)^+ \Big] \\ D_k &= \min\big(D_{max},\ \mu_C P + (\varphi_l + \varphi_V)\,\kappa\,(C_{sp} - C_o)^+\big) \end{aligned} \tag{13.14}$$

where the screen s is the one the rules would set for the hour, $U(s,v) = [1/h_{in}(s) + 1/h_{out}(v)]^{-1}$ the cover's overall coefficient, φ_l the leakage, E the transpiration at the humidity limit, φ_H the exchange the limit would need to remove the crop's water, capped at what the gap and the lee windows can pass, φ_V the venting that throws away the sun's surplus, D_{max} the line's 130 kg/(ha h), and F the flue gas a firing boiler makes, all per square meter and second, so that the factor of 3.6×10^7 gives dollars per hour for the hectare when D_k and F are in mg/(m^2 s). Each curve is sampled every 0.5 K between the lowest temperature full vents can reach and the band's top, 26 °C in daylight and 20 °C in the dark, and replaced by its lower convex hull, so it becomes a sum of segments of rising slope. The program is Equation (13.15).

$$\begin{aligned} \min\ & \sum_j c_j \delta_j + 10^5 \sum_d (\sigma_d^+ + \sigma_d^-) \\ & + \sum_k \left[u_k^+ + u_k^- + 500\,(w_k^+ + w_k^-) + 0.05\,(z_k^+ + z_k^-) \right] \\ \text{s.t. } & r_k = L_k + \sum_{j \in k} \delta_j, \qquad 0 \leq \delta_j \leq \ell_j \\ & \textstyle\sum_{k \in d} \omega_k r_k + \sigma_d^+ - \sigma_d^- \in [N_d - 24\delta_T,\ N_d + 24\delta_T] \\ & r_k - r_{k-1} = (u_k^+ - u_k^-) + (w_k^+ - w_k^-) \\ & 0 \leq u_k^{\pm} \leq \rho_r \\ & r_k - \hat{r}_k = z_k^+ - z_k^- \end{aligned} \tag{13.15}$$

where segment j of hour k has width ℓ_j and slope c_j (weighted by $1 + 0.001k$ so that ties go to the nearer, surer hour), L_k is the hour's lowest reachable temperature, ω_k is one except for the part of the present hour already gone, N_d is the day's sum $24\bar{T}$ less the hours already measured, held loosely for a day the horizon only reaches part of, $\delta_T = 0.2$ K is the tolerance, $\rho_r = 3$ K/h the ramp with r_{-1} the present plan, and $\hat{r}_k$ the previous plan's value for the same hour. Every variable is nonnegative. Because the slopes rise, the cheap segments fill first and no integer variable is needed. A shortfall on a

Table 13.4 The Nine Cases and the Graded Week

#	What differs from the reference	What it shows
1	cap off, $\Delta T_v = 4$ K, every vent open 12:00 to 13:00	one exchange, three variables
2	none, Monday and Tuesday	a setpoint is not the recipe
3	override off, two nights	the screen traps the water
4	cap off, a bright Wednesday	dosing into open vents
5	planner on, case 2's days	temperature integration
6	planner on, a front 9 h early	the forecast and the band
7	planner on, a dark saturated day	no setting wins
8	planner on, gas $\times 3$ for a day	price against the mean
9	a heat wave, 12 K warmer	fog as the last handle
10	the week, 20/18 °C, override and cap off	graded challenge

day's mean is priced at \$100 000 per kelvin-hour, so it is reported rather than taken quietly, a change between hours costs \$1 per kelvin so nearly equal hours are planned level, and the hold term of \$0.05 per kelvin keeps a plan in force against a fourth-digit change of an input. The bounded-variable simplex of the framework's shared linear programming module solves it, and over the graded week with the suggested settings each of the 168 solves had 165 to 286 variables and 25 to 49 rows, ended optimal in 54 to 269 pivots and bound flips, and the application logs the count in place of a solve time, which varies from run to run. The plan becomes setpoints by its hour's free temperature T_f, where the sun alone holds the house, as heating to r venting at $r + \Delta T_v$ when $r \geq T_f + 0.3$, heating at $r - \Delta T_v$ and venting at r when $r \leq T_f - 0.3$, and the band split around r in between, and the climate computer moves to each hour's setpoints at the ramp rather than in one step.

The plant, the planner and the loops each have a model of their own. The plant's is Equations (13.1) to (13.10). The planner's internal model is Equation (13.14) on the forecast, which omits the dawn and dusk transients, the override's third stage and the fog, and prices the CO_2 that holds the setpoint by light up to the line's capacity but without the cap. Summed over the present hours it plans, it expects \$4498 of gas and CO_2 where the house buys \$5351, 16 % less, and \$1040 against \$1282 over case 5. The loops are tuned on no identified model, and the application identifies none.

13.6 Nine Cases and a Spring Week

Table 13.4 lists the cases. Each starts from a reference line with the fixed setpoints of Section 13.4, the humidity override on at 85 %, the cap on, both CO_2 sources and gas at 0.40 \$/m^3, and alters it as its row states. Headless engine runs at the default seed, with every changed setting made before Run as a learner makes it, supplied every number below, and Table 13.5 collects what each case scored. A day counts as dark below 50 W/m^2 outside and as bright above 14 MJ/m^2, and its bright hours are those above 400 W/m^2. One seed draws both the synthetic weather and the sensor noise, so a case replays the same weather in every run, and the spreads quoted over other seeds of the noise keep that weather fixed. A Reset restarts a case from midnight with the learner's settings kept, as Section 1.6.1 describes, and a setting changed before Run gives the same run as the same setting carried through a Reset, the settled start included.

13.6.1 One Exchange, Three Variables

Case 1 runs a bright Monday with the vents 4 K above a 19.5 °C heating setpoint and the cap off. At 11:57 the house is at 23.50 °C, 83.4 % and 584 ppm, its vents 51 % open and its air changing 6.4 times an hour, and the line doses its full 130 kg/h, of which 96 kg/h already leaves by the vents and leaks. At noon the vent loop goes to MANUAL with every window open, the exchange more than doubles to 13.7 air changes an hour, and by 12:15 the air is at 20.30 °C, the humidity at 66.6 % and the CO_2 at 499 ppm. The boiler is asked to fire from 12:21, the air reaches the heating setpoint at 12:24, and at 12:57 the boiler burns 51 W/m^2 against the open roof with the air at 19.24 °C. The hour

Table 13.5 Results by Case. Mean, Gas and Cost Are the First Day's, and RH Is the Minutes Above 85 % in That Day's Dark Hours

Case and configuration	Mean (°C)	Gas (m^3)	Cost ($)	RH (min)	CO_2 (ppm)
2 fixed, Monday	19.02	1040	651	64	585
2 fixed, Tuesday	18.57	1547	787	89	—
3 screen, override off[a]	—	566	—	720	—
3 screen never closed[a]	—	1188	—	130	—
3 screen, override on[a]	—	781	—	108	—
4 cap off	19.14	850	581	172	550
4 cap on	19.14	850	557	172	530
5 planner, Monday	18.91	865	577	68	632
5 planner, Tuesday	18.79	1501	774	92	—
6 front missed	18.68	1743	883	81	—
6 perfect forecast	18.79	1624	866	43	—
7 override on	20.69	1965	900	601	—
7 override off	18.80	456	283	931	—
8 gas ×3	18.80	1456	1933	85	—
8 normal price[b]	18.79	1521	785	84	—
9 fog on	22.24	197	251	162	470
9 fog off	22.97	197	234	162	469

[a]Gas and RH over the second night, 18:00 to 06:00, all minutes above 85 %.
[b]The scripted price spike removed, a counterfactual run.
CO_2 is the mean over the bright hours of a bright day.

costs 41.6 m^3 of gas ($17) that goes straight out, and the vents and leaks carry off 103 of the 130 kg dosed, 79 % against 74 % in the hour before noon at half aperture, while gross photosynthesis falls 5.4 %. Back in AUTO the house is at 22.30 °C, 85.1 % and 578 ppm by 13:30. The heating loop sees only its own temperature and fights the vent with gas.

🕮 **Pause.** Try Exercise 13.2 before reading on.

13.6.2 A Setpoint Is Not the Recipe

Case 2 runs Monday and Tuesday on 19.5 °C from 06:00 to 19:00 and 17 °C at night. On bright Monday the sun carries the air to 21.7 °C by 09:19, the vents open from 21.5 °C and hold it there, standing 56 % open at 12:30 with 96 kg/h of CO_2 leaving, and the day's mean is 19.02 °C on 1040 m^3. On dull Tuesday there is no sun to float on and the mean falls to 18.57 °C, 0.43 K short, on 1547 m^3. The boiler sits at its limit for 1.68 h on Monday and 3.57 h on Tuesday, 186 of those 315 minutes between 05:00 and 08:00, when the screen opens over a cold house and the setpoint ramps up, and the air falls to 15.58 °C, 1.42 K under the night setpoint, before the pipes catch up. The setpoints are held on both days and the recipe on one.

13.6.3 The Screen and the Water It Traps

Case 3 runs from Monday 00:00 to Tuesday 06:00 with the override off. Over the second night, 18:00 to 06:00, the closed screen holds the gas to 566 m^3 against 1188 m^3 with the screen never closing, and it keeps the house above 85 % for all 720 minutes, reaching saturation and wetting the leaves for 26 min. With the override on, the gap opens to its 10 % limit and MIC-104 asks for up to 42.1 % of lee vent, which keeps the night under 85 % for all but 108 min at a peak of 88.5 % on 781 m^3, 215 m^3 more than the closed screen and 407 m^3 less than no screen. The third stage never acts that night. The screen saves more than half of a night's gas and the override returns a third of the saving to keep the leaves dry.

> **TRY IT IN THE APP 3 · The Screen at Night**
>
> Load the case and press run to Tuesday 06:00. Watch the humidity badge through the second night: under the closed screen the house stays above 85 % for all 720 minutes. Reload the scenario, turn the Humidity override on and run again: the gap and the lee vent keep the night under 85 % for all but 108 min.

13.6.4 Dosing Into Open Vents

Case 4 runs a bright Wednesday with the cap off. For 8.02 h the vents are more than half open while the line doses, and of the 1043 kg it doses in those hours, 784 kg leave by the vents. Over the day the line doses 1460 kg and the bright hours average 550 ppm. With the cap on the day doses 1305 kg, the CO_2 bill falls from \$210 to \$186, the bright hours average 530 ppm and the crop banks 240 kg of assimilates against 243. Widening the vent band to 4 K instead keeps 592 ppm through the bright hours, for a day 0.88 K over the recipe at 19.88 °C. The cap saves CO_2 only where the vents stand wide, and anything that keeps the vents closer is also a CO_2 strategy.

13.6.5 Temperature Integration

Case 5 runs case 2's days with the planner on. On Monday it lets the middle of the day run warm on the sun, planning 24.7 °C from 11:00 to 16:00 with the air peaking at 24.89 °C at 11:18, and holds the night at the floor of 16 °C, for a mean of 18.91 °C on 865 m^3, against 19.02 °C on 1040 m^3 for the fixed setpoints. The closer vents also keep 632 ppm through the bright hours against 585. On dull Tuesday there is less sun to bank, and the planner spends 1501 m^3 against 1547 and holds 18.79 °C where the fixed setpoints let the day fall to 18.57 °C. Over the two days it uses 8.5 % less gas, 2366 m^3 against 2587, for \$1351 against \$1438 and 406 kg of assimilates against 400. The band constrains the plan and not the air. The plan sits at the 16 °C floor for 15.0 h of the two days, the air below it for 11.6 h of those, but below 15.5 °C for only 1.08 h, all of it between 05:00 and 09:00, when the screen opens over a cold house and the boiler spends 88 minutes of the two dawns at its limit, and the lowest air is 14.63 °C at 05:57. Figure 13.2 shows the planner at the end of Monday.

> **TRY IT IN THE APP 5 · Temperature Integration**
>
> Load the case and press run. Watch the air and the violet plan in the Climate view through bright Monday: the plan lets the middle of the day run warm, near 24.7 °C, and holds the night at the 16 °C floor. The day ends at a mean of 18.91 °C on 865 m^3 of gas, against 1040 m^3 for the fixed setpoints of case 2.

13.6.6 A Forecast That Misses a Front

Case 6 runs Friday with a cold front at 12:00 that the forecast places at 21:00 until it is three hours away. Trusting a bright afternoon, the planner holds the night at 16.0 °C. From 06:00 the forecast catches up, and with the front coming the planner raises the morning's heating setpoint to 22.5 °C, unscreened, to bank the mean before the cold, and the air reaches 21.6 °C by noon. The day ends at 18.68 °C, 0.12 K short of the 18.8 °C the planner aims for at the cheap edge of its tolerance, on 1743 m^3 and \$883. On a perfect forecast the planner heats the screened night to its 20 °C ceiling instead and makes 18.79 °C on 1624 m^3 and \$866, so the miss costs 119 m^3, \$17 and a tenth of a kelvin. The band's floor kept the night at 16 °C however bright the afternoon looked. Widening the band from 17 to 24 to 15 to 28 °C cuts Monday's gas in case 5 from 969 to 794 m^3, while what the missed front of case 6 costs falls from \$22 to \$17, so in this case the band's floor, not its width, sets what a bad forecast spends.

13.6.7 A Dark Day No Setting Wins

Case 7 runs Thursday under overcast and drizzle, about 14 °C outside, saturated and calm, with 2.5 MJ/m^2 of sun, the planner on and the CO_2 setpoint asking for its full 800 ppm from 50 W/m^2. Saturated air at 14 °C holds 98 % of the water the 85 % limit allows at 17 °C, so venting barely dries the house. The override holds the 90 % line, never crossing it, but not the limit, which the house

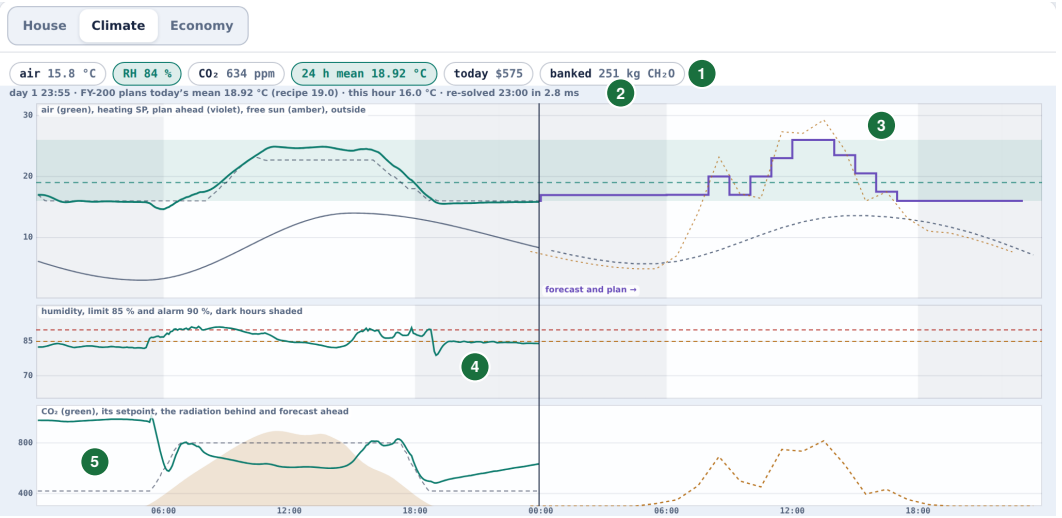


Figure 13.2 Temperature integration in case 5 at day 1 23:55, in the Climate view with no control changed. (1) The badges read 15.8 °C, 84 %, 634 ppm, a trailing mean of 18.92 °C, $575 today and 251 kg banked. (2) The summary line reads that FY-200, re-solved at 23:00, plans today's mean at 18.92 °C against the 19.0 °C recipe and 16.0 °C for this hour, beside the browser's own solve time. Monday's air climbs to about 25 °C around noon and sits just under the 16 °C floor through the night, and (3) Tuesday's plan, the stepped solid line in violet, holds 17.0 °C until 08:00, reaches 26.0 °C at 12:00 and 13:00 under a free-sun forecast, dotted in amber, that peaks at 29.3 °C, and is back at 16.0 °C from 17:00. (4) The humidity strip holds the night at the 85 % limit, and (5) the CO_2 strip shows what the night's respiration gathers, near 1000 ppm, falling as the screen opens and about 630 ppm through the bright hours.

exceeds for 17.6 h with the heating raised the full 5 K for 9.53 h and the lee windows at half. That heat sends the mean to 20.69 °C, 1.7 K over the recipe, and from 16:00 the planner reports that the day's mean cannot come down to 19 °C inside the band. The CO_2 is held within 22 ppm of its setpoint only by dosing 985 kg into vents that carry 96 % of it out, for 19 kg of assimilates, and the day costs $900. With the override off it costs $283 and keeps the mean at 18.80 °C, while the house sits above 90 % for 21.8 h and wets the leaves for 16.5 min. The override ranks the humidity above the recipe and the CO_2, and the planner says what that costs instead of widening its band.

> ✦ **DID YOU KNOW? Drying a Saturated Day Buys Little**
>
> From the simulator: on the dark, saturated Thursday of case 7 the humidity override raises the day's gas from 456 to 1965 m^3 and its cost from $283 to $900, and the house still spends 601 dark minutes above 85 %, against 931 without it (Table 13.5).

13.6.8 Gas Prices and Fog

Case 8 runs a dull Tuesday with the gas price tripled from midnight. The planner moves heat out of the day and into the screened night, 19.50 °C before dawn where it averaged 18.44 at the normal price and 20.37 °C on average from 09:00 to 17:00 instead of 21.86, the normal-price figures coming from a counterfactual run with the scripted spike removed. The boiler fires less in daylight, the flue gas falls from 295 to 276 kg and the line buys 994 kg of liquid CO_2 instead of 928. The day costs $1933 against $785 at the normal price, and since the unchanged plan would have cost $2001, re-planning wins back $68 of the $1217 the spike adds ($52 to $74 over five other seeds of the sensor noise), while

Table 13.6 The Graded Week in Every Combination of the Planner (P), the Override at 83 % (O) and the Cap (C)

Changes	Cost ($)	Gas (m^3)	Worst (K)	RH (min)	CO_2 (ppm)	Fails
none	4292	6447	1.00	775	578	mean, RH, CO_2
C	4271	6447	1.00	775	559	mean, RH, CO_2
P	3780	5306	0.22	715	617	RH
P and C	3778	5306	0.22	715	615	RH
O	6312	10500	0.83	92	573	mean, CO_2, cost
O and C	6290	10500	0.82	92	553	mean, CO_2, cost
P and O, or all	5594	9184	0.33	78	611	none
all at 85 %	5398	8780	0.33	192	611	RH

Worst is the largest miss of a day's mean, RH the most humid day's dark minutes, CO_2 the lowest bright day. Gas costs 0.40 \$/m^3, 1.20 \$/m^3 on Wednesday.

a recipe of 18.5 °C would save \$125. Case 9 runs a cloudless Monday 12 K warmer with the vents wide for 15.92 h. The fog runs for 7.90 h from 09:16 and takes the peak from 29.63 °C to 27.78 °C, while the humidity at the peak rises from 64 to 78 % with 36 % outside, and the crop banks 213 kg against 207 for \$18 more electricity.

13.6.9 The Graded Week

The challenge runs the spring week, bright, dull, bright, dark, a front, bright and bright, with the gas price tripled for Wednesday and Friday's front arriving nine hours before the forecast says. The target is every day's mean within 0.5 K of 19 °C, no more than 150 min above 85 % in any day's dark hours, at least 600 ppm through the bright hours of every bright day and at least 1200 kg of assimilates, for at most \$6100. Shipped on 20 and 18 °C with no planner, no override and no cap, the week costs \$4292 and fails three parts, six days off the recipe with the worst by 1.00 K, a day with 775 min of humid dark hours and a bright day at 578 ppm. Table 13.6 adds the three suggested changes in every combination.

The planner and the override pass together, at \$5594 with every day within 0.33 K, at most 77.5 min of humid dark hours, 611 ppm on the worst bright day and 1227 kg banked, margins of 0.17 K, 72.5 min, 11 ppm, 27 kg and \$506 on the five targets, the same week whether the settings are made before Run or carried through a Reset. That is more than the shipped \$4292 because dry nights cost gas. Each fails alone. The planner alone gives the cheapest week, \$3780, but leaves 715 min of humid dark hours in its worst day, and the override alone costs \$6312 and leaves three days off the recipe and a bright day at 573 ppm. The same pattern holds over five other seeds of the sensor noise, with the passing week at \$5595 to \$5606. Once the planner and the override run, the cap changes nothing in this week. It engages for 6.55 h, but the measured CO_2 stays below the lowered setpoint with the line at its limit, so the two weeks are identical state for state. With the planner alone the cap lowers the worst bright day from 617 to 615 ppm, and without the planner to 553 ppm. The planner is the CO_2 strategy here, because its warm afternoons keep the vents closed. Figure 13.3 draws the rows of Table 13.6 against the cost and humidity targets.

The humidity limit is the price setter, as shown in Figure 13.4. With the planner, moving the limit from 83 to 85 % saves \$196 a week and lets the worst night rise from 77.5 to 192 min, and each point of limit from 78 to 90 % is worth between \$76 and \$113. At every limit the planner's week is \$656 to \$725 cheaper than the same override on fixed setpoints, and only the planner's weeks at 80 to 84 % pass, since at 78 % the dry nights cost \$6140. The horizon has to reach a day ahead. With horizons of 24 to 48 h the suggested week costs \$5594 to \$5597 and passes, while at 18 h Wednesday's bright hours hold 598 ppm and fail the CO_2 target. At 12 and 6 h a plan that sees only part of the day holds Wednesday's bright hours cooler, venting at 23.1 and 21.3 °C against 23.5 °C, so the vents stand 55 and 72 % open against 52 and the bright hours hold 596 ppm and 535 ppm. Volume 9, Chapter 7 meets the same question of how far a forecast should reach on a slower plant. A perfect forecast

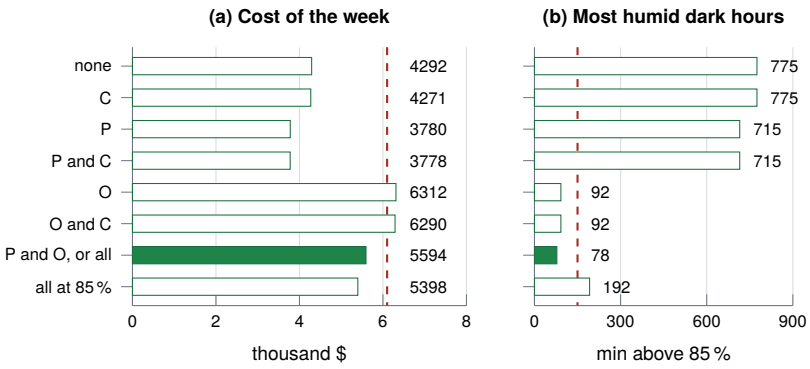


Figure 13.3 The graded week of Table 13.6 in every combination of the planner (P), the override at 83 % (O) and the cap (C): the week's cost against the $6100 the challenge allows and the most humid day's dark minutes against its 150 min (dashed lines). The solid bar passes every target, white bars fail at least one. Notice that the planner alone is the cheapest week and the override alone the driest, and that only the two together sit inside both lines.

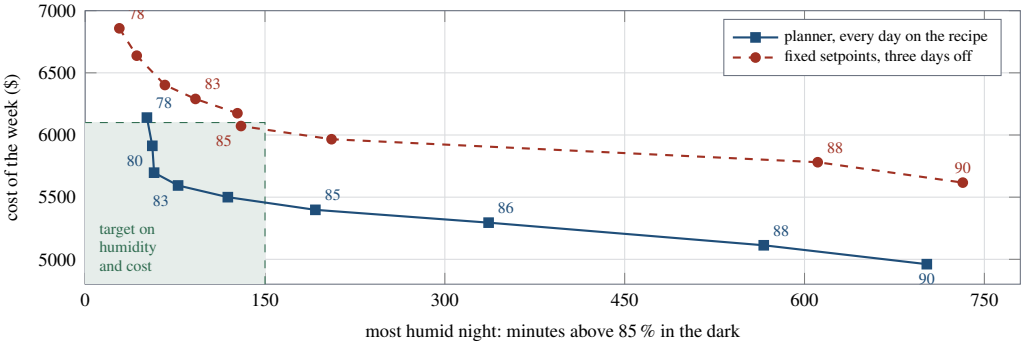


Figure 13.4 The price of a humidity limit over the graded week, with the override and the cap on and the limit at 78 to 90 % as labeled, with the planner (solid line, squares) and on fixed setpoints (dashed line, circles). With the planner every day stays within 0.35 K of the recipe, and on fixed 20 and 18 °C setpoints three days miss it by up to 0.84 K and the worst bright day holds 551 to 554 ppm, so the shaded box covers two of the five targets.

makes the suggested week $30 dearer, at $5625, with its worst day 0.21 K off rather than 0.33 and Friday, the day of the front, only $2 cheaper, and an error three times the declared size makes it $28 dearer.

The external reset of TIC-101 earns its place. Replacing it, in a copy of the source, with the shared law's own back-calculation on the same tuning raises the overshoot of an unramped 3 K step of the night setpoint at 01:00 in case 2 from 0.21 to 1.26 K, and that of the step that opens case 6 on a perfect forecast, where the plan starts at 20 °C over the settled 17 °C night, from 0.22 to 0.41 K at the shipped ramp of 3 K/h and from 0.31 to 0.98 K at the fastest planned ramp of 8 K/h, because the master's integral, which external reset ties to the pipe temperature achieved, otherwise grows through the hour the pipes take to heat and then has to unwind.

13.7 Technology in Practice

The climate computer of Section 13.4 reads a handful of instruments and moves five kinds of handle, and the planner of Section 13.5 sits above it. Figure 13.5 draws them as a glasshouse would install

them.

What Measures It

The loops that matter most read one measuring box in the crop for the air temperature and the humidity, a CO_2 analyzer behind its sample line, a sensor in the pipe water and a weather station outside. CO_2 sensors of the nondispersive infrared type measure the absorption of infrared light by the CO_2 molecules and correlate it with the concentration [18]. In the chapter the box reads behind a 1 min lag and the analyzer behind the 2 min lag of its sample line.

> **TECHNOLOGY Sensor: Air Temperature and Humidity**
>
> The standard measurements of a glasshouse climate computer include the air temperature and the humidity deficit inside, the heating pipe temperature, the screen positions and an outside weather station [17]. Temperature elements are thermistors, sensitive over a limited range, or platinum resistance detectors, accurate and linear over a wider one. Humidity sensors use a capacitive or resistive element with a hygroscopic material whose capacitance or resistance changes with the humidity. Accuracies of 0.5 °C and two points of relative humidity are reported for sensors in greenhouse systems, which still need calibration and maintenance, and dew is a known source of interference with a humidity sensor [18]. The chapter's measuring box reads the true air behind a 1 min lag with 0.05 K and 0.5 points of noise.

What Moves It

The heating loop fires a boiler into pipes along the rows, the vent loops open roof windows lee side first, a drive draws the screen at the gutter, a valve sets the dose of CO_2 and the fog runs on and off. A test facility equipped like a commercial house had the same set: continuous roof ventilation with an opening of 0.3 m^2 per square meter of floor, movable screens, a fogging system and a CO_2 supply [17]. Every handle acts on the same air.

> **TECHNOLOGY Final Element: Boiler and Heating Pipe Rail**
>
> The compartments of a test facility, equipped with actuators available in commercial high-tech greenhouses, had two pipe heating systems, a rail pipe on the floor with a peak capacity of 180 W/m^2 and a pipe at crop height with 30 W/m^2, each on its own setpoint, and growers set a minimum rail pipe temperature among their climate setpoints [17]. The pipe water is a store of heat between the burner and the air. In the chapter the boiler of 150 W/m^2 fires behind a 5 min lag into pipes held at 15 to 90 °C by TIC-102, and an outage shows on the trend as a firing request at its limit while the pipe temperature falls.

Where the Controller Runs

In the competition of Section 13.2 the teams' setpoints reached a standard commercial process computer, and setpoints and measurements were exchanged every five minutes [17]. Such controllers hold logic, timing, counting and three-mode PID control, and supervisory and lower loops run continuously with cycle times from milliseconds to minutes [19]. The chapter's climate computer executes every minute and the planner every hour, so the scan of the platform is never the limit. In a function block program the heating cascade is two PI blocks with the master's reset fed from the measured pipe temperature, the vents take a high select of two PI blocks, and the planner is a supervisory layer that writes the heating and vent setpoints.

What Is Modernizing It

Three lines of work bear on this chapter. The first puts learned models inside predictive control. A neural network that predicts the inside temperature from 15 inputs served an optimizer of six roof windows on a strawberry house, and in a six-day field test its root mean square temperature error was 2.45 °C against 3.01 °C for conventional ventilation control [20]. The second is autonomous growing. One team of the competition ran three layers: a crop model set average climate targets, a

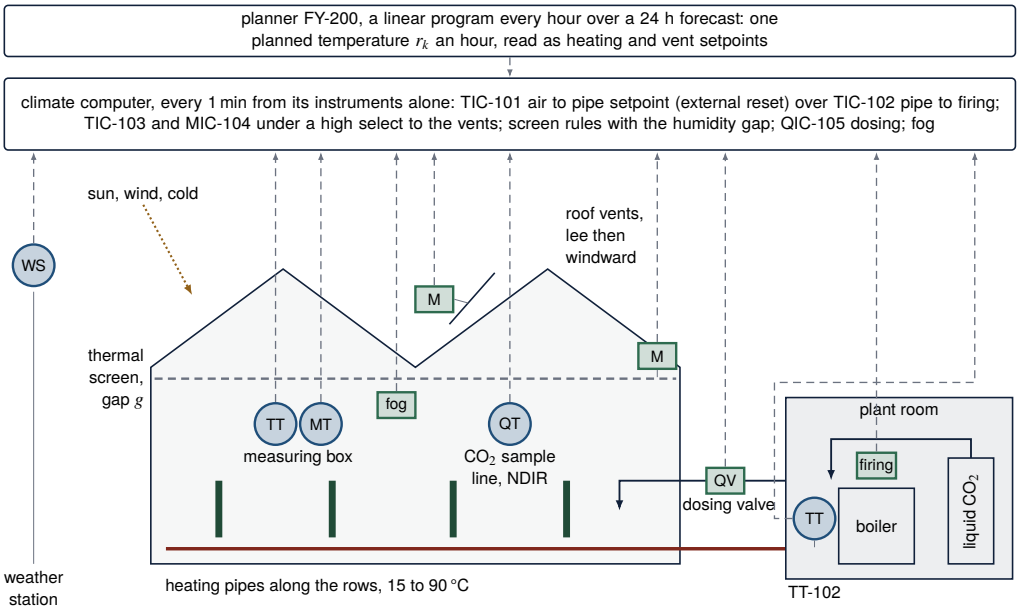


Figure 13.5 The chapter's loops with the devices a commercial glasshouse would install. The thick line is the hot water of the pipes, thin dark arrows the carbon dioxide, dashed gray lines the signals and the dotted amber arrow the weather. Notice that every handle (the boiler, the vents, the screen, the fog and the dosing valve) acts on the one body of air that the measuring box samples, and that the planner writes setpoints, never a valve position.

middle layer turned them with the weather forecast into ideal climate trajectories, and a bottom layer set the setpoints [17], the hierarchy of this chapter's planner over its climate computer. The third is dense sensing. Networks of temperature, humidity, CO_2 and light sensors capture the variation of the microclimate across the house, and Kalman filtering and the wavelet transform reduce their measurement errors [18]. None of them removes the coupling of one air exchange that carries heat, water and CO_2 together.

> ⚠ **IN PRACTICE** **Drying a House With Heat and Vents**
>
> Humidity is held for the health of the crop: appropriate relative humidity prevents fungal infections and controls transpiration [21]. In the remote growing competition of 2018, a team that found small spots of fungal disease in its cucumber compartment was advised to lower the humidity. It stopped the misting, opened the vents on both the lee and the wind side, and set the minimum temperature of the pipe rail to 40 °C, which held the air at its night and day temperatures while the house ventilated. The heating energy rose steeply [17]. Heating and venting at once is the third stage of this chapter's override.

13.8 Limitations

The house is one air node with a big-leaf crop, a reduced vent law and a lumped screen, so it shows the couplings between temperature, humidity and CO_2 but not the gradients inside a real house, the cold air that falls through a screen gap or a clear sky's radiation. The crop answers the calendar day's mean with no tolerance of a mean held over several days, which real temperature integration exploits [1, 5], and fruit set and quality are absent. The weather is synthetic and the forecast error declared.

OPEN CHALLENGE Sensing the Crop Instead of the Air

In the remote growing competition, the extra sensors the teams chose showed no clear link with yield or net profit. Robust sensors giving direct information on the crop's performance, rather than indirect information on its air and root zone, might improve such control substantially [17]. A review adds that the calibration, maintenance and interoperability of sensors remain significant concerns [18]. Progress would be a measured crop state, such as the leaf wetness duration and the transpiration that a crop-based humidity regime set targets for [5], held by the planner in place of the air's humidity.

The planner's model is quasi-steady per hour and prices CO_2 without the cap, so it expects 16 % less than the house buys over the suggested week, and it holds each day's mean separately. There is no heat buffer, no combined heat and power unit and no artificial light, which are how real growers move gas in time and make CO_2 on demand. Volume 9, Chapter 6 prices heat from such a unit and a stored tank. Every threshold and price is a teaching value, and none of the numbers is a measurement of a commercial glasshouse.

OPEN CHALLENGE One Box for a Hectare

A greenhouse is traditionally monitored by sensors placed with no guarantee that they represent the house, and most methods proposed for placing them do not approach the problem from a control perspective. One study took the mean of 56 paired temperature and humidity sensors as the reference and found that eight well-chosen sensors, aggregated by a learned formula, estimated it with a correlation of 0.999 [21]. The chapter's house is one air node read by one box. Progress would be a placement chosen for the loop it serves.

13.9 Conclusion

A glasshouse spends a temperature budget against a humidity limit through one air exchange that also carries its CO_2, and the simulator of this chapter measures each side of the trade from ledgers that close to 3.6×10^{-14}. Fixed setpoints of 19.5 and 17 °C make the 19 °C recipe on a bright day and miss it by 0.43 K on a dull one, the screen saves 622 of 1188 m^3 on a night while keeping the house above 85 % for twelve hours, and a vent opened at noon costs 41.6 m^3 of gas and 79 % of the CO_2 dosed. The receding-horizon linear program over the forecast heats the bright Monday on 17 % less gas and holds the dull Tuesday 0.22 K nearer the recipe, a missed front costs it \$17, and on a dark saturated day it reports a mean 1.7 K over the recipe rather than widen its band. Over the graded week the planner and the override pass together at \$5594, each fails alone, the cap changes nothing once both run, and the planner is \$656 to \$725 cheaper than fixed setpoints at every humidity limit from 78 to 90 %. The house is lumped, its crop reduced and its weather synthetic. A fuller model would add a canopy temperature, a screen gap with its own buoyant flow, integration over several days, a heat buffer and lighting. The simulator makes temperature integration a measurable trade against the humidity limit that every glasshouse has to hold.

KEY POINTS

- A tomato crop answers the mean temperature of its day, so the heating setpoint is a budget while the humidity limit holds hour by hour.
- The planner heats the bright Monday on 865 m^3 of gas instead of 1040, 17 % less.
- The thermal screen saves 622 of the 1188 m^3 a night would burn and keeps the house above 85 % for all twelve hours.
- Over the graded week the planner and the override pass together at \$5594, each fails alone, and the cap changes nothing once both run.

Where else this applies. Volume 9, Chapter 11 plans pumps over the same 24-hour horizon.

References

[1] A. N. M. de Koning, "More efficient use of base load heating with a temperature integrating control programme. Effect on development, growth and production of tomato," *Acta Horticulturae*, no. 229, pp. 233–238, 1988.

[2] C. E. García, D. M. Prett, and M. Morari, "Model predictive control: Theory and practice—a survey," *Automatica*, vol. 25, no. 3, pp. 335–348, 1989.

[3] E. J. van Henten, "Greenhouse climate management: An optimal control approach," Ph.D. dissertation, Agricultural University, Wageningen, The Netherlands, 1994.

[4] R. J. C. van Ooteghem, L. G. van Willigenburg, and G. van Straten, "Receding horizon optimal control of a solar greenhouse," *Acta Horticulturae*, no. 691, pp. 797–806, 2005.

[5] O. Körner, "Crop based climate regimes for energy saving in greenhouse cultivation," Ph.D. dissertation, Wageningen University, Wageningen, The Netherlands, 2003.

[6] G. van Straten, G. van Willigenburg, E. van Henten, and R. van Ooteghem, *Optimal Control of Greenhouse Cultivation*. CRC Press, 2010.

[7] B. H. E. Vanthoor, C. Stanghellini, E. J. van Henten, and P. H. B. de Visser, "A methodology for model-based greenhouse design: Part 1, a greenhouse climate model for a broad range of designs and climates," *Biosystems Engineering*, vol. 110, no. 4, pp. 363–377, 2011.

[8] B. H. E. Vanthoor, P. H. B. de Visser, C. Stanghellini, and E. J. van Henten, "A methodology for model-based greenhouse design: Part 2, description and validation of a tomato yield model," *Biosystems Engineering*, vol. 110, no. 4, pp. 378–395, 2011.

[9] T. de Jong, "Natural ventilation of large multi-span greenhouses," Ph.D. dissertation, Agricultural University, Wageningen, The Netherlands, 1990.

[10] C. Stanghellini, "Transpiration of greenhouse crops: An aid to climate management," Ph.D. dissertation, Agricultural University, Wageningen, The Netherlands, 1987.

[11] K. W. Winspear, "Control of heating and ventilation in glasshouses," *Acta Horticulturae*, no. 6, pp. 62–78, 1968.

[12] B. J. Bailey, "Thermal screens for reducing heat losses from glasshouses," *Acta Horticulturae*, no. 70, pp. 26–34, 1977.

[13] I. Th. H. Gieling, "Commercial greenhouse computersystems: Inventarisation of implemented research, carried out by commercial firms," *Acta Horticulturae*, no. 106, pp. 59–66, 1980.

[14] A. J. Udink ten Cate, G. P. A. Bot, and J. J. van Dixhoorn, "Computer control of greenhouse climates," *Acta Horticulturae*, no. 87, pp. 265–272, 1978.

[15] E. M. Nederhoff, "Effects of CO_2 concentration on photosynthesis, transpiration and production of greenhouse fruit vegetable crops," Ph.D. dissertation, Agricultural University, Wageningen, The Netherlands, 1994.

[16] A. Elings, F. L. K. Kempkes, R. C. Kaarsemaker, M. N. A. Ruijs, N. J. van de Braak, and T. A. Dueck, "The energy balance and energy-saving measures in greenhouse tomato cultivation," *Acta Horticulturae*, no. 691, pp. 67–74, 2005.

[17] S. Hemming, F. de Zwart, A. Elings, I. Righini, and A. Petropoulou, "Remote control of greenhouse vegetable production with artificial intelligence—greenhouse climate, irrigation, and crop production," *Sensors*, vol. 19, no. 8, p. 1807, 2019.

[18] E. Bicamumakuba, M. N. Reza, H. Jin, Samsuzzaman, K.-H. Lee, and S.-O. Chung, "Multi-sensor monitoring, intelligent control, and data processing for smart greenhouse environment management," *Sensors*, vol. 25, no. 19, p. 6134, 2025.

[19] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule, and M. Thompson, "Guide to operational technology (OT) security," National Institute of Standards and Technology, Tech. Rep. NIST Special Publication 800-82r3, 2023.

[20] D.-H. Jung, H.-J. Kim, J. Y. Kim, T. S. Lee, and S. H. Park, "Model predictive control via output feedback neural network for improved multi-window greenhouse ventilation control," *Sensors*, vol. 20, no. 6, p. 1756, 2020.

[21] O. S. Ajani, E. Aboyeji, R. Mallipeddi, D. Dooyum Uyeh, Y. Ha, and T. Park, "A genetic programming-based optimal sensor placement for greenhouse monitoring and control," *Frontiers in Plant Science*, vol. 14, p. 1152036, 2023.

13.10 Exercises

13.1 Concept check · 2 min

The planner of Section 13.5 lets a bright afternoon run warm and holds the night at the floor of its band. Why can this meet the same 19 °C recipe on less gas than fixed setpoints?

(a) Growth rises with the hourly temperature, so the warm afternoon adds yield that pays for the gas

(b) A warm house loses less heat through its cover, so a warm afternoon costs less than a warm night

(c) Growth follows the trailing daily mean, so a degree the sun supplies can replace one the boiler buys at night

(d) The humidity limit is looser in daylight, so the night can run cooler without wetting the leaves

13.2 Concept check · 1 min

In case 1 every vent opens wide at noon for an hour. What happens to the air temperature, the humidity and the CO_2 in the first quarter of an hour?

(a) The air cools and the humidity rises, while the CO_2 holds because the line keeps dosing

(b) The air cools and dries, while the CO_2 rises because cooler leaves take up less of it

(c) Only the air temperature moves, because the vent loop acts on temperature alone

(d) All three fall together, because one exchange carries heat, water and CO_2 out

13.3 Concept check · 2 min

At dawn the boiler of the glasshouse sits at its limit for most of an hour while the air is below its heating setpoint. Why does the air loop TIC-101 not wind up?

(a) Its reset is a lag of the measured pipe temperature, so it follows the pipes the slave achieves

(b) Its output is clamped at 15 to 90 °C, and the clamp alone stops the integral from growing

(c) The pipe loop TIC-102 is fast enough that the pipes reach every request within a minute

(d) The heating setpoint ramps over an hour, so the error never grows large enough to wind up

13.4 Theory and calculation · 6 min

The cover has $U_{open} = 7.5\,\mathrm{W/(m^2\,K)}$ with the screen open and $U_{scr} = 4.0\,\mathrm{W/(m^2\,K)}$ under it, both at a 4 m/s wind, where $h_{out} = 7 + 2v$. A closed screen passes $p_s = 0.15$ of the exchange.

(a) Use Equation (13.3) to find $h_{in,0}$ and R_s. *Checkpoint:* 15 W/(m² K), 0.117 m² K/W

(b) The screen closes to $s = 0.90$, a 10 % gap. What share of the closed screen's resistance R_s does it keep, and what is the overall cover coefficient $U = [1/h_{in}(s) + 1/h_{out}]^{-1}$ at 4 m/s? Give the coefficient to two decimals. *Checkpoint:* 0.800, 4.41 W/(m² K)

(c) What share of the open-house exchange does the screen pass with the 10 % gap, from $\pi(s)$ of Equation (13.7)? *Checkpoint:* 0.488

13.5 Theory and calculation · 8 min

Take the house of Section 13.3 with $h = 5$ m, a 4 m/s wind and the air 5 K warmer than outside, so that the factor $\frac{1}{4}\sqrt{v^2 + 0.5\,(T - T_o)^+}$ of Equation (13.7) is 1.075. Note $f(1) = 1$ and $f(0) = 0$.

(a) With the vents shut and the screen open, find the leakage φ_l in m³/(m² h) and the time h/φ in minutes. *Checkpoint:* 2.5 m³/(m² h), 120 min

(b) With the aperture at $a = 0.5$, so that only the lee windows are fully open, find φ and h/φ. *Checkpoint:* 28.3 m³/(m² h), 10.6 min

(c) With every window open ($a = 1$), and with the vents shut under a fully closed screen, find h/φ in each case. *Checkpoint:* 4.5 min, 800 min

13.6 Theory and calculation · 6 min

Fixed setpoints hold 19.5 °C by day and 17 °C at night, the heating setpoint ramping over an hour from 06:00 and from 19:00 as in Section 13.4.1. The recipe is a 19 °C mean.

(a) Find the 24-hour mean of the heating setpoint, to two decimals. *Checkpoint:* 18.35 °C

(b) Find the growth demand $g_{19}\,\mathrm{clip}_{16/9}\big((\bar{T}_{24} - 10)/9\big)$ of Equation (13.10) at that mean, with $g_{19} = 18\,\mathrm{g/(m^2\,d)}$. *Checkpoint:* 16.7 g/(m² d)

(c) Case 2's bright Monday ends at a mean of 19.02 °C. How many kelvin-hours above the setpoint's own mean did the sun supply, to the nearest whole number? *Checkpoint:* 16 K h

13.7 App challenge · 8 min

Run scenario 2 (Heating and venting on fixed setpoints) with its default settings at the default seed to its end, Tuesday midnight (48 h), at 6 h/s, then export the CSV file. Repeat with scenario 5 (Temperature integration).

(a) For each scenario, read Monday's mean and gas from the CSV row at 24 h (columns `T_day_C` and `metric_gas_m3`). Give the means to 0.01 K and the gas to the nearest $\mathrm{m^3}$. *Checkpoint:* 19.02 °C and 1040 $\mathrm{m^3}$, 18.91 °C and 865 $\mathrm{m^3}$

(b) At the end of each run, read Tuesday's mean from the *24 h mean* badge and the two-day gas from the *Gas* tile. What share of Tuesday's gas does the planner save, and what share of Monday's? *Checkpoint:* 3 %, 17 %

(c) Explain the difference between the two days in one or two sentences.

13.8 App challenge · 7 min

Run scenario 7 (The humid dark day) at the default seed with its default settings to midnight (24 h) at 2 h/s. Read the *Cost* tile, the *24 h mean* badge and the worst night on the *Humid minutes at night* tile. Then press Reset, turn the *Humidity override* off before pressing Run, and run to midnight again.

(a) Give the three readings for both runs (cost to the dollar, mean to 0.01 K, minutes to the nearest whole minute). *Checkpoint:* \$900, 20.69 °C, 601 min; \$283, 18.80 °C, 931 min

(b) Which target does each run give up, and why can no setting hold all of them on this day?

13.9 App challenge · 10 min

Run the challenge (a spring week) at the default seed to its end (168 h) at 6 h/s, three times, each time making the changes before pressing Run: (1) as shipped, (2) with the planner on, the Humidity override on with a humidity limit of 83 % and the CO_2 cap on, and (3) as in (2) with the limit at 85 %. For each run read the *Cost*, *Days off the recipe*, *Humid minutes at night* and *Bright-day CO_2* tiles.

(a) Which parts of the target does run (1) fail, and by how much? *Checkpoint:* mean (6 of 7, worst 1.00 K), 775 min, 578 ppm

(b) Does run (2) meet the target, and at what cost? *Checkpoint:* yes, \$5594

(c) What does run (3) save, and which part does it fail? *Checkpoint:* \$196, humidity at 192 min

13.10 Application area · 3 min

The crop of this chapter answers the calendar day's mean, with no tolerance of a mean held over several days. Real temperature integration has used windows of 24 hours and six days. Suppose the dull Tuesday of case 2 ends 0.43 K short of the recipe. What would a six-day window allow the grower to do about it, and what must the grower still check?

13.11 Application area · 4 min

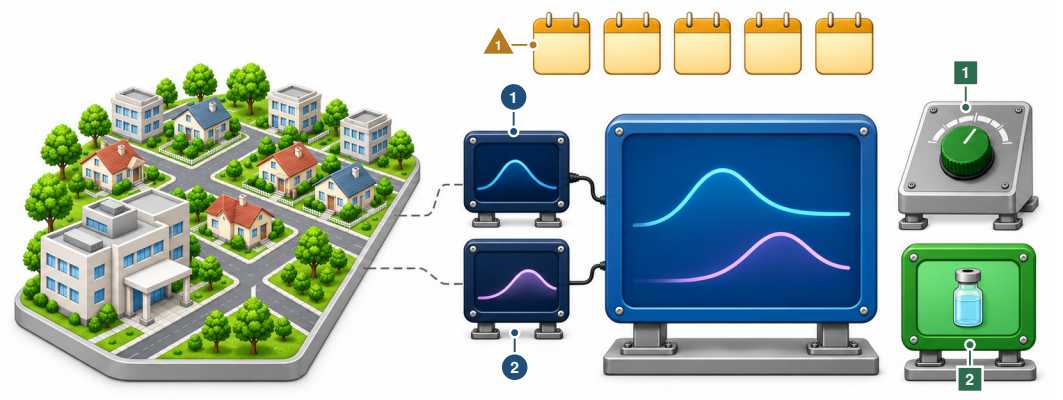
In case 8 the gas price triples for a whole day. Re-planning wins back $68 of the $1217 the spike adds, while lowering the recipe to 18.5 °C would save $125. Explain to a grower why re-planning wins back so little, and name two pieces of equipment that real houses use to move gas in time.

13.12 Review the agent · 5 min

An AI assistant reviews the graded week and writes: “Each point of humidity limit is worth $76 to $113 a week. Raise the limit from 83 to 90 %: the week will cost about $4961 instead of $5594, and since the override still holds the house at its limit, the humidity target stays met.” Find the flaw in the claim and say how you would check it in the app.

CHAPTER 14

Estimating and Steering an Epidemic From Week-Old, Partial Counts



The controller moves 1 the intervention dial, standing for every measure that cuts contact
2 vaccine doses a day (advanced)

The controller reads 1 reported cases, a week late and only a share of them
2 hospital census, the beds in use, counted daily

What disturbs it 1 a week between an onset and its report ▲ a testing change that moves the count, not the epidemic ▲ a variant or a holiday, unannounced

A teaching city publishes its case reports and its hospital census to the screens long after the infections they describe, while the dial acts on the city of today: notice the row of calendar days between what happens and what is seen. Illustration, not to scale.

IN PLAIN TERMS **Steering an Epidemic by Late, Partial Counts**

The main idea. A city learns about an epidemic from two counts. One is the cases its tests find, a week late and only a share of them. The other is the hospital beds in use. Measures that cut contact slow the spread at once, but the counts show it weeks later.

The challenge. A change in testing can halve the case count with no change in the epidemic. A more contagious strain can also arrive unannounced. Beds in use must stay below what hospitals hold, so whoever sets the measures must judge the present from old numbers.

Why it matters. Acting on old numbers means acting late, and each day of waiting fills more beds.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the history segment on 1918 (Section 14.1), the history segment on 1927 (Section 14.2), the fact on a count given as a range (Section 14.3), the two try-it segments and the fact on a retuned threshold (Section 14.7), and the key points
Study	45 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	20 min	the first page, the cases and the graded 180 days (Section 14.7), Technology in Practice (Section 14.9), Limitations (Section 14.10) and the open challenge

The control question A city counts its cases a week after the onsets they sample, finds only a share of them, and fills its hospital beds two weeks after the infections. What can an estimator recover from those counts, and which policy keeps the census under the beds when testing changes and a variant arrives unannounced?

Learning objectives

- Derive the early growth rate, the peak and the final size of the uncontrolled city, and add up the delays between the dial and the beds.
- Explain why a relay on the reported count cycles and a PI loop on the census overshoots its target.
- Compare an EKF, a particle filter and an MHE on one augmented state, and tell where the census decides the result and where a bound does.
- Set up the economic MPC and the scenario tree, and count their moves, slacks and shared days.
- Judge whether a scenario tree buys anything a hand-set planning margin does not.

Prerequisites Sections 1.4.3 (epidemic growth rate), 1.4.5 (sparse and late samples), 1.6.2 (velocity form), 1.6.3 (MPC and MHE) and 1.6.4 (scenarios). Chapter 11 for a loop on a late laboratory result.

The app https://apmonitor.com/apps/epidemic-control/

See also Volume 5, Chapter 4 for multirate estimation on a late analyzer and laboratory, Volume 5, Chapter 13 for an extended Kalman filter that applies late fixes at the instant they describe. Volume 9, Chapter 7 for a scenario tree with non-anticipativity on an ensemble forecast and Volume 9, Chapter 6 for a predictive controller that sends heat ahead of a delay.



14.1 Introduction

About 140 people are infected in a teaching city of one million when the simulator's first case begins. Left alone, the epidemic infects 94 % of the city and fills 7,201 hospital beds on day 73, against a capacity of 1,500. The reported 7-day average of new cases passes 300 a day on day 38, and a dial set to 70 % the next day turns the true infections at once. The reported average keeps rising for 20 more days, to 1,794 a day on day 59, and the census peaks at 1,435 beds on day 62. Waiting a day longer costs about 182 beds at the peak, and waiting two days breaches the capacity for 23 days.

> ⧖ HISTORY **Measures in Seventeen Cities (1918)**
>
> A study published in 2007 gathered the dates on which 17 cities of the United States brought in 19 classes of public health measures during the influenza pandemic of 1918, among them the closing of schools, churches and theaters. Cities that put several measures in place at an early phase of their epidemic had peak death rates about 50 % lower than cities that did not, and less steep epidemic curves. The difference in total excess deaths was smaller, about 20 %, and less certain statistically. Few cities kept their measures longer than six weeks, and the findings supported the view that spread resumes once measures are relaxed [1]. When federal planning guidance for community measures was drafted in 2007, evidence of this kind, historical analyses and observations rather than controlled studies, was most of what there was [2].

The delays are not the only obstacle. A reported case is a sample of the onsets of a week before, scaled by an ascertainment that no one measures, and a change in testing moves the count without moving the epidemic. Counts that arrive late and partial are the data of every surveillance system, and the statistics of a reporting delay were worked out for events that occur before they are reported [3]. Process plants have the same shape. A laboratory assay reports a stream's composition hours after the sample, an analyzer drifts against a slower unbiased measurement, and a constraint binds on a quantity no actuator moves directly. The kraft digester of Volume 5, Chapter 4 grades its cook on exactly such an analyzer and laboratory. The estimators that treat such plants, the extended Kalman filter (EKF), the particle filter (PF) and the moving horizon estimator (MHE), have been set against each other on process examples where the states must stay physical [4], and the constrained MHE has a stability analysis of its own [5]. The belief this city overturns is that a constrained estimator

wins whenever the data change their meaning. In this model the census decides that question, and the constraint matters only where the data fall far outside what the model expects.

A single plant on which all of this can be watched at once is rare. Work from the control community fits a compartmental model to a real outbreak and designs estimation, control or both for it [6–8], and the epidemiological literature estimates the reproduction number from reported incidence without a controller in the loop [9]. This chapter sets three estimators and four policies against one declared city, with its truth logged beside every count, and measures what each can and cannot do. The compartmental model, the three estimators and model predictive control (MPC) are established methods. Four features of the implementation carry the lessons: an augmented state with a chain of onsets that lets one model and one noise description serve all three estimators, an MHE solved daily by Gauss–Newton steps that are each a quadratic program, a scenario-tree MPC solved by successive linearization, and a measurement boundary that keeps every estimator and controller away from the truth. Two findings on this teaching model shape its lessons: the census anchors every estimator through a testing cut, and the scenario tree's protection against a variant is a margin that a hand-set margin also buys, for less. Ten cases, each graded by a completion rule, carry the experiments.

Section 14.2 traces the problem's history. Section 14.3 declares the city and the two counts it publishes, and Section 14.4 derives its growth rate, its final size and the three delays between the dial and the beds. Section 14.5 sets out the three estimators on one augmented state, and Section 14.6 the relay, the PI loop, the economic MPC and the scenario tree. Section 14.7 measures the ten cases, Section 14.8 says which cases carry the lesson, Section 14.9 names the surveillance streams, the measures and the people who decide in a real system, and Sections 14.10 and 14.11 state the limitations and the conclusions. The exercises in Section 14.12 work through the cases.

14.2 Compartments, Undercounts and Planned Measures

The model the simulator integrates descends from a paper of 1927, as the history box below tells. Seven decades later a review of the field set out the threshold theorems for such models in terms of the basic reproduction number R_0, the contact number and the replacement number, for the classic SIR models and for SEIR models with an exposed class [10]. The early growth of an epidemic and its reproduction number are related through the distribution of the generation interval, and the equation that applies depends on that shape [11].

> ⧖ HISTORY **One Formula for the Final Size (1927)**
>
> A paper of 1927 by Kermack and McKendrick set out the problem in its first paragraph: one or more infected people enter a community more or less susceptible to a disease, the disease spreads by contact, and each infected person is in the end removed by recovery or by death [12]. Its formula for the final size of an epidemic rested on a simple model with exponentially distributed infectious periods. A paper of 2006 reviewed the later work and showed the formula unchanged by a latent stage, by any number of infectious stages or by a stage of isolation, though it fails with certain kinds of uneven mixing. The size of an epidemic of a new respiratory pathogen can therefore be estimated as soon as its basic reproduction number can be approximated [13], and the simulator, with its latent stage, checks its integrator against the 1927 relation.

Counting came next as a problem in its own right. A count of events that have occurred but are not yet reported can be corrected for the delay with a model of how reports arrive [3]. Only a share of infections is ever reported, and under-reporting in epidemics became a modeling problem of its own [14]. A framework with software for estimating a reproduction number that varies in time followed in 2013 [9].

State estimation reached epidemic models from signal processing and from the process industries. The bootstrap particle filter of 1993 carried a cloud of samples through a nonlinear model and weighted each by the data [15]. Regularized filters added progressive correction, which applies a sharp likelihood in stages so that the cloud does not collapse [16], and filtering was turned into maximum

Table 14.1 Each Variable's Role and Its Value When Case 1 Begins

Variable	Role	Start	Range
H	constrained state (true census)	0 beds	$\le H_{\max} = 1500$
$\tilde{H}_k$	measurement, census count	0 beds	daily
C_k	measurement, reported cases	0 a day	daily
u	MV, intervention dial	0	0–1
D	MV, vaccine doses (advanced)	0	0–5000 a day
m	DV, variant multiplier	1	1–2.5
h	DV, holiday multiplier	1	1–2
$\bar{\rho}$	DV, testing level	0.30	cut $\times 0.5$
outage	DV, reports held	none	1–10 d
S, E, I, R	states	$N-140$, 100, 40, 0	people
u_e	state, effect of the dial	0	0–1
c	ledger, cost	\$0	\$M
β, ρ	hidden parameters	0.50/d, 0.30	truth
$\hat{x}, \hat{R}_t$	estimates	prior	per estimator

likelihood estimation for partially observed nonlinear stochastic systems, with a disease's dynamics as the worked application [17]. In process control, moving horizon estimation with constraints and an arrival cost received a stability analysis in 2003 [5], and a critical evaluation of the two estimators on process examples argued for the constrained MHE where the EKF's estimates can leave the physical region [4].

Planning the measures is the newest layer. Economic MPC optimizes an economic cost rather than tracking a setpoint [18], which suits a dial whose cost is lost output. Multi-stage MPC represents uncertainty by a tree of scenarios whose decisions may differ only once the branches can be told apart, and it was applied to a semi-batch polymerization reactor under uncertainty [19]. The reservoir of Volume 9, Chapter 7 plans its release on a scenario tree built from an ensemble of inflow forecasts. More recently the process systems and control communities applied modeling, state estimation and optimal control to an epidemic [6], robust and optimal predictive control to the same kind of outbreak [7], and an optimal predictive strategy to distancing policies [8]. These studies are context only. The simulator models no real place, period or pathogen, and every value in it is a declared teaching value.

14.3 A Declared City and What It Publishes

The simulated city is one well-mixed population of $N = 10^6$ people in days, as shown in Figure 14.1. Everything inside the dashed frame is teaching truth, logged and drawn for the learner but read by no estimator or controller. Two counts leave the city once a day, the reported cases C_k and the census count $\tilde{H}_k$, and they are the only path from the city to the estimators. The intervention dial u is the manipulated variable, chosen once a day by the policy in force. Each variable's role is given in Table 14.1.

14.3.1 Compartments, Transmission and the Dial

Susceptible people are infected at the incidence f, pass a latent period in E, are infectious in I, and leave either for the hospital or for the removed class, as shown in Equation (14.1).

$$\begin{aligned} f &= \beta(1-u_e)SI/N, \qquad \dot{S} = -f - DS/N \\ \dot{E} &= f - \sigma E, \qquad \dot{I} = \sigma E - \gamma I \\ \dot{H} &= p_h \gamma I - H/L_h \\ \dot{R} &= (1-p_h)\gamma I + H/L_h + DS/N \\ \tau_u \dot{u}_e &= u - u_e, \qquad \dot{c} = c_u u^2 \end{aligned} \tag{14.1}$$

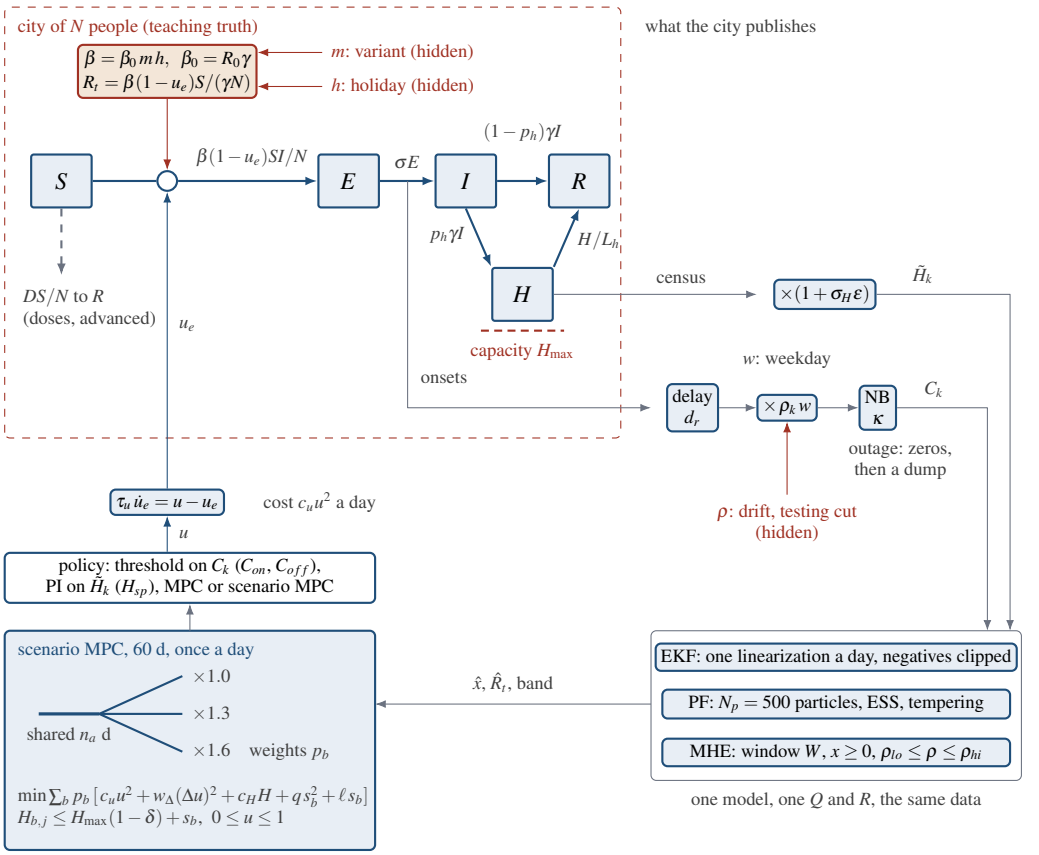


Figure 14.1 The teaching city as modeled, with what it publishes and who reads it. Inside the dashed frame (teaching truth) the compartments S, E, I, H and R exchange people at the rates shown, the transmission β carries the hidden variant and holiday multipliers m and h, and the census H stands against the capacity $H_{\max}$. The onsets σE reach the reports C_k after the delay d_r, the ascertainment ρ_k, the weekday factor w and negative-binomial noise of dispersion κ, and the census reaches $\tilde{H}_k$ with relative noise σ_H. Three estimators read only those two counts. A policy, the scenario-tree MPC among them, sets the dial u, whose effect u_e lags by τ_u. Not to scale.

where $1/\sigma$ is the latent period, $1/\gamma$ the infectious period, p_h the share of those leaving I who are admitted, L_h the stay, D the vaccine doses a day, u_e the dial's effect and c the cost ledger in millions of dollars. The doses reach whoever comes, so only the share S/N still susceptible gains immunity. The five compartments sum to N at every step. Transmission and the reproduction number are given by Equation (14.2).

$$\beta(t) = \beta_0\, m(t)\, h(t), \quad \beta_0 = R_0\gamma, \quad R_t = \frac{\beta(1-u_e)S}{\gamma N} \tag{14.2}$$

where m is the variant multiplier and h the holiday multiplier, both hidden. The dial removes a share of transmission once its effect has caught up, and its cost is quadratic, \$50 million a day at $u = 1$, so a moderate dial held long costs less than a hard one held briefly.

14.3.2 Reports, Ascertainment and the Census

The day's report counts a sample of the onsets σE that occurred during one day, d_r days earlier, as shown in Equation (14.3).

$$\mu_k = \rho_k \, w_{\mathrm{dow}(k)} \int_{k-d_r-1}^{k-d_r} \sigma E \, dt, \qquad C_k \sim \mathrm{NB}(\mu_k, \kappa) \tag{14.3}$$

where w is a declared weekday pattern (1.15, 1.10, 1.05, 1.00, 1.00, 0.85 and 0.85, mean 1) and the negative binomial has mean μ_k and variance $\mu_k + \mu_k^2/\kappa$. It is drawn as a Poisson count whose mean is μ_k/κ times a gamma deviate of shape κ, the Poisson by inversion below a mean of 30 and by a rounded normal above. The ascertainment drifts about a testing level $\bar{\rho}$ as a discrete Ornstein–Uhlenbeck deviation with a 30-day correlation, as shown in Equation (14.4).

$$\begin{aligned} \rho_k &= \mathrm{clip}(\bar{\rho} + \delta_k,\ 0.05,\ 0.95) \\ \delta_k &= a\,\delta_{k-1} + s_\rho \sqrt{1-a^2}\,\varepsilon_k, \qquad a = \mathrm{e}^{-1/30} \end{aligned} \tag{14.4}$$

where $s_\rho = 0.03$ and ε_k is a standard normal deviate. A testing cut multiplies $\bar{\rho}$ by $f_t = 0.5$ at once, clipped to 0.05 to 0.9. The census is counted once a day with $\sigma_H = 0.01$, as shown in Equation (14.5).

$$\tilde{H}_k = \max\left(0,\ \mathrm{round}[H(k)(1+\sigma_H \varepsilon'_k)]\right) \tag{14.5}$$

A reporting outage of d_{out} days publishes zeros and holds the day's reports, then adds the held reports to the next day's, the catch-up dump. Under zero noise the published report equals Equation (14.3) over the window $[k-8, k-7]$ to a relative 6.0×10^{-14}, and shifting the window a day either side misses by 14 to 16 %, which confirms the 7-day delay.

> **+ DID YOU KNOW? A Count Given as a Range**
>
> Even after an epidemic, its size is an estimate. Guidelines of the Centers for Disease Control and Prevention (CDC) put the people in the United States infected with the 2009 pandemic influenza virus between April 2009 and April 2010 at an estimated 43 to 89 million, and the deaths at about 12,000 [2].

14.3.3 Events, Starting State and Integration

Every hard event (the variant, a testing change, the outage and its dump, the holiday, a breach of the capacity) happens exactly once, at the boundary between two steps, and leaves a mark in the log, and the right-hand side of Equation (14.1) stays pure. The run starts with $E = 100$, $I = 40$, $S = N - 140$ and no one in hospital. The engine integrates in days with fourth-order Runge–Kutta at $\Delta t = 0.05$ d, runs the estimators and the policy once a day, and logs every 6 h. The onsets for Equation (14.3) are taken from a delay line of the cumulative onset ledger. Within one step the controller executes first, then the integration, the events, the outputs and the log, so the city publishes day k at $t = k$ and the controller uses it at the start of the next step. A log row at a whole day therefore shows that day's counts and the previous day's estimates, which an exported column (`est_day`) records. Every noise stream is seeded from the case, as Section 1.6.1 describes for every app. The parameters are listed in Table 14.2.

14.3.4 What the Model Leaves Out

The city has no age groups, households, regions or travel, and immunity does not wane. The latent and infectious stays are exponential, and real ones are distributions with a delay before infectiousness. Admission comes as people leave I and every patient stays 10 days, with no intensive care and no deaths. The reporting delay is fixed at 7 days, whereas a real one is a distribution that changes with testing capacity. The dial stands for every measure at once, with one lag and a quadratic cost. The estimators are told p_h, the census noise and the weekday pattern exactly, and that assumption carries

Table 14.2 Parameters, Defaults and Standing (A Adjustable in the Panel, H Hidden From the Estimators and Policies, F Not Adjustable in the Panel). All Are Teaching Values

Symbol	Meaning	Default (range)	
N	population	10^6	F
R_0	basic reproduction number	3 (1.2–6)	A, H
$1/\sigma$, $1/\gamma$	latent and infectious periods	4, 6 d	A
p_h, L_h	hospitalized share, stay	0.03, 10 d	A
τ_u	lag of the dial's effect	4 d (1–14)	A
d_r	onset-to-report delay	7 d (2–14)	A
ρ_0, s_ρ	ascertainment at start, drift sd	0.30, 0.03	A, H
κ, σ_H	report dispersion, census noise	20, 0.01	A
m_v, f_t	variant multiplier, testing cut	1.5, 0.5	A, H
h, d_{out}	holiday ($\times$, 7 d), outage	1.3, 3 d	A, H
c_u	cost of the full dial	\$50M/d	A
$H_{\max}$, δ	capacity, planning margin	1500, 5 %	A
C_{on}, C_{off}, u_{on}	threshold lines and dial	300, 100/d, 0.8	A
H_{sp}, K_c, τ_I	PI target, gain, reset	900, 0.002/bed, 20 d	A
W, ρ_{lo}, ρ_{hi}	MHE window and band	21 d, 0.1, 0.8	A
q_β, σ_ρ, q_E	random walks of β, ρ, E	0.03, 0.01, 0.03	A, F
N_p	particles	500	A
m_b, p_b, n_a	tree branches, weights, shared days	1, 1.3, 1.6, $\frac{1}{3}$, 7	A, F
w_Δ, c_H	move penalty, cost of illness	\$5M, 0	A
q, ℓ	slack prices	0.1, \$10M/bed	F

more of the estimation results than any other (Section 14.7). Each of these choices keeps the lesson on the delays and the hidden fraction, and each breaks down where a real epidemic's heterogeneity matters.

14.4 Growth, Final Size and Three Delays

With the dial at zero the model has two closed forms. Because $\mathrm{d}\ln S/\mathrm{d}t = -\beta I/N$ and $\mathrm{d}(S+E+I)/\mathrm{d}t = -\gamma I$, the quantity $s+e+i-\ln(s)/R_0$ is conserved in fractions of N, so $E+I$ peaks where $s = 1/R_0$ and the final size z solves $z = 1-\mathrm{e}^{-R_0 z}$ for a vanishing seed. At $R_0 = 3$ these give a peak $E+I$ of $0.300509N$ and a final size of 0.940490 from the 140 seeded, against 0.940480 for a vanishing seed. Early on, a growth rate r of the infections satisfies $(1+r/\sigma)(1+r/\gamma) = R_0$ for exponential latent and infectious stages, the relation the generation-interval argument gives for this model [11], so the infections double every 4.69 d at $R_0 = 3$ and every 3.43 d at $R_0 = 4$. Section 1.4.3 sets this growth beside the other growth laws of the volume.

Three delays separate the dial from the beds. The dial's effect builds over $\tau_u = 4\,\mathrm{d}$. A person infected today has onset after a mean latent period of 4 days and is reported 7 days after that, so the report describes infections about 11 days old. Admission follows the infectious period, a mean 10 days after infection, and the stay lasts 10 days more, so the census integrates the infections of the last three weeks. A policy that reads the reports acts on the epidemic of about two weeks before, and one that reads the census acts later still. The 7-day average adds about three days more. Section 1.4.5 treats sparse and late samples in general.

🕮 **Pause.** Try Exercise 14.1 before reading on.

14.5 Three Estimators on One Augmented State

All three estimators solve one problem with one model, one prior and one description of the noise, and they differ only in method. Once a day, at $t = k$, they receive day k's report and census and

estimate the augmented state of Equation (14.6).

$$\begin{aligned} x_k &= [S, E, I, H, \beta, \rho, z_1, \ldots, z_L]_k, \qquad L = d_r + 1 \\ x_k &= f_d(x_{k-1}, u, u_e, D) + w_k \\ y_k &= [\rho_k \, w_{\mathrm{dow}(k)} \, z_{L,k}, \; H_k] + v_k \end{aligned} \tag{14.6}$$

where f_d is one day of Equation (14.1) integrated with four Runge–Kutta steps of 0.25 d and the dial's lag in closed form, and z_j holds the onsets of the day j days back, so the report is a delayed sample of a state rather than an output with dead time. The navigation filter of Volume 5, Chapter 13 meets a late measurement the same way, applying each acoustic fix at the instant it describes. The dial and its effect are the controller's own and are known exactly. The process noise is a random walk on β with standard deviation $\sigma_\beta = q_\beta \beta_{ref}$ a day ($\beta_{ref} = 0.5$/d, the prior $R_0 = 3$ times γ), a random walk on ρ with $\sigma_\rho = 0.01$ a day, and a relative error $q_E = 0.03$ on E. The measurement variances are $\max(1, \mu + \mu^2/\kappa)$ for the report and $(\sigma_H H)^2 + 1$ for the census. The prior is $\beta = 0.50 \pm 0.125$/d, $\rho = 0.30 \pm 0.10$, $E = 100 \pm 100$ and $I = 50 \pm 50$ (one standard deviation). The estimators are told N, σ, γ, p_h, L_h, τ_u, d_r, κ and σ_H, and never R_0, ρ, m or h. Each reports $\hat{R}_t = \hat{\beta}(1 - u_e)\hat{S}/(\gamma N)$ with a band.

14.5.1 Extended Kalman Filter

The EKF predicts with one model day, linearized by forward differences of f_d about the estimate, as $P \leftarrow FPF^\top + Q$, then updates with both measurements in the Joseph form. Its only constraint is the common practical one: a component that comes out negative is set to zero. Its band on $\hat{R}_t$ is two standard deviations from P by the delta method.

14.5.2 Particle Filter With Tempered Updates

The PF carries $N_p = 500$ samples of the augmented state through the same model day and draws the same process noise for each [15]. Each is weighted by the negative-binomial likelihood of the report and a Gaussian likelihood of the census, and the cloud is resampled systematically when its effective sample size, $1/\sum w_i^2$, falls below $N_p/2$. A day whose likelihood alone would leave fewer than $N_p/10$ effective particles is taken in stages, each raising the likelihood to the largest power that keeps the effective size at $N_p/2$, with the cloud resampled and jittered by the stage's share of one day's process noise between stages [16]. The jitter makes this a regularized approximation rather than an exact importance-sampling step. The band is the weighted 5 to 95 % range.

14.5.3 Moving Horizon Estimation

The MHE solves, once a day, a window of the last $W = 21$ days, as shown in Equation (14.7).

$$\begin{aligned} \min_{\varphi, w} \; & \|\varphi\|^2 + \sum_j \left(w_{\beta,j}^2 + w_{\rho,j}^2 + w_{E,j}^2 \right) \\ & + \sum_j \left[\frac{(C_j - \mu_j)^2}{R_{C,j}} + \frac{(\tilde{H}_j - H_j)^2}{R_{H,j}} \right] \\ \text{s.t.} \; & x_{k_0} = \bar{x} + L_P \varphi \geq 0, \quad \rho_{lo} \leq \rho_j \leq \rho_{hi}, \\ & \beta_j \geq 0, \quad w_{E,j} \geq -0.9/q_E \end{aligned} \tag{14.7}$$

where the decision variables are the window's starting state, written through the Cholesky factor L_P of the arrival covariance, and the standardized process noises of every day, 77 variables for a full window with 99 linear constraint rows. The arrival cost is the filtering form [5], with $\bar{x}$ the MHE's own estimate made at the window's start and P carried forward by the extended Kalman recursion along its estimates. Each Gauss–Newton step linearizes the residuals by forward differences and solves $\min \frac{1}{2}\|r + J\Delta\|^2 + \frac{1}{2}\lambda\|\Delta\|^2$ with the constraints on the dense dual active-set quadratic program of the shared framework (`core/qp.js`, Section 1.6.3), with $\lambda = 10^{-6}$. A backtracking line search

on the true cost accepts the step, the variances $R_{C,j}$ are re-evaluated at the iterate, and the iteration stops when the cost falls by less than 10^{-6} of itself, the step falls below 10^{-4}, or after 12 iterations. Yesterday's solution, shifted a day, is the warm start. A window whose end state comes out non-finite or with a negative compartment is not used and the previous estimate is carried one model day instead, which happens only with several parameters at their extremes at once. The band is two standard deviations from the Gauss–Newton Hessian, ignoring the active constraints.

14.6 Dial Policies as Implemented

The policy in force sets the dial once a day and holds it until the next day, as a policy decision is held. In manual mode the learner's dial passes through, and the vaccine doses, an advanced second actuator, are given by hand in either mode up to a declared supply.

14.6.1 A Relay and a PI Loop

The threshold policy is a relay with hysteresis on the reported 7-day average $\bar{C}_7$: the dial goes to $u_{on} = 0.8$ when $\bar{C}_7$ passes $C_{on} = 300$ a day and to zero when it falls below $C_{off} = 100$, and it holds between them. The PI policy is the velocity-form controller of Section 1.6.2 with no derivative action, on the census count against a target $H_{sp} = 900$ beds. It is direct acting, $a = -1$, because a higher dial lowers the census. It has $K_c = 0.002$ of the dial per bed and $\tau_I = 20\,\mathrm{d}$, executes once a day, clamps its output to $0 \le u \le 1$ and unwinds its integral by back-calculation with a one-day tracking time T_t. Under any other policy it tracks the dial in force, so a transfer is bumpless.

🕮 **Pause.** Try Exercise 14.3 before reading on.

14.6.2 Economic MPC by Successive Linearization

The MPC plans the dial for 60 days from the selected estimate, never the truth, as shown in Equation (14.8).

$$\begin{aligned} \min_{u,s} \; & \sum_{b} p_b \Big[\sum_{j=0}^{59} \big(c_u u_{b,j}^2 + c_H H_{b,j} \big) + w_\Delta \sum (\Delta u_b)^2 \\ & \qquad + q\, s_b^2 + \ell\, s_b \Big] \\ \text{s.t.} \; & 0 \le u_{b,j} \le 1, \quad s_b \ge 0, \quad H_{b,j} \le H_{\max}(1-\delta) + s_b \\ & u_{b,j} = u_{b',j} \;\; (j < n_a) \end{aligned} \tag{14.8}$$

where b indexes the branches, p_b their weights, $H_{b,j}$ the predicted census on branch b at the end of day j, and s_b one slack per branch priced at $\ell = \$10$ million a bed, far above any day's cost. The dial moves daily for the first $n_a = 7$ days and in blocks of 7 days after, so the nominal controller has 15 moves and one slack. The prediction is the estimators' model day f_d run from $(\hat{S}, \hat{E}, \hat{I}, \hat{H})$ with $\hat{\beta}$ held, the dial's lag in closed form. Each linearization takes the sensitivities of every predicted census to every move by forward differences, solves the quadratic program on `core/qp.js` with at most 600 active-set iterations, replaces the plan and simulates it again. The plan is accepted when no move changes by more than 10^{-3}, or after 4 linearizations, a cap the status reports. Yesterday's plan, shifted a day, starts the next solve, and a failed solve holds yesterday's dial. The plan's last weeks ease the dial, because the census their infections would fill lies past the horizon. Only the first day is applied.

14.6.3 Branches That Share a Week

The scenario MPC solves Equation (14.8) with three branches whose transmission is $\hat{\beta}$ times $m_b = 1.0$, 1.3 and 1.6 from today, each with weight $p_b = \frac{1}{3}$, and the first 7 days of the dial shared by all three. That equality is the non-anticipativity constraint of multi-stage MPC [19]: the dial applied this week must keep even the 1.6 branch under the planning ceiling until the branches may differ. With 31 moves and three slacks the problem stays strictly convex, and the app's test finds 0.0667 as the least eigenvalue of its Hessian. The tree does not know when a variant will arrive. It hedges against $\times 1.3$

Table 14.3 The Ten Cases

#	Case, policy and events	Days	Teaching point
1	Manual, dial by hand	120	the two lags
2	Threshold 300/100 a day at 0.8	240	a relay on a lagged count
3	PI on the census, target 900	150	feedback that comes late
4	Manual, dial 0.66 from day 35	120	three estimators agree
5	As case 4, testing cut on day 70	120	the census anchors
6	MPC on the MHE	180	beds as a constraint
7	As case 6, variant on day 110	180	tree against a margin
8	As case 4, outage days 71–73, dump on day 74	100	where a bound matters
9	Challenge: threshold on the EKF, cut on day 60, variant on day 110, then a 28-day hand-over	180	the whole problem
10	$R_0 = 4$, threshold, no estimator	180	out of reach of a reactive rule

Table 14.4 Measured Outcomes of the Cases. Peak Is the True Census, and Cost Is the Dial's Cost Over the Case's Run

#	Configuration	Peak (beds)	Cost ($M)
1	dial 0	7201, day 73	0
	dial 0.7 on day 39 / 40 / 41	1435 / 1610 / 1799	—
2	threshold as shipped	932, day 55	4768
	reporting delay 3 / 14 d	715 / 2406	—
3	PI, target 900 / 300	1749 / 1051	—
6	MPC / scenario MPC	1421 / 1126	1788 / 2198
7	MPC, margin 5 %	1542, 33 d over	2440
	MPC, margin 10 %	1485	2599
	scenario MPC	1369	2855
10	threshold, no estimator	2098, day 51	—
	MPC on MHE / EKF / PF[a]	1436 / 1435 / 1431	2783 / 2783 / 2680

[a]Peak and cost through the 28-day hand-over with the dial held. Cases 4, 5 and 8 are estimation cases and are reported in the text.

and $\times 1.6$ every day, whether a variant is coming or not, and it branches at day 0 rather than day 7 because a tree branching a week ahead lets its recourse see the branch the moment it begins and keeps no margin. It peaks at 104.8 % of the capacity in case 7 and at 104.0 to 106.5 % on the challenge's seven records.

Three models are kept apart. The plant integrates Equation (14.1) at 0.05 d with the true β, the hidden multipliers and the reports' noise. The estimators carry the model day f_d, the random walks and the declared parameters. The MPC predicts with the same model day from an estimate and, on a branch, a multiplied $\hat{\beta}$. Over 60 days at the shipped parameters the model day reproduces the plant's census to better than 0.001 %, and with a transmission of 15 a day ($R_0 = 6$, a latent day, two infectious days, a variant of 2.5 and a holiday of 2) to within 0.97 %.

14.7 Measurements on Ten Cases

The ten cases appear in Table 14.3. Each runs on its own seed with the shipped settings, with any setting its row names typed in before Run, and the events at the days listed. All values reported below were produced by the app's engine, run without a browser but stepped exactly as the page steps it, and an estimate is compared with the truth of the day whose data it rests on, the alignment the app's own referee uses. Its record is the same whether the case is the first loaded, follows another case from the menu, or is rerun from Reset. The measured outcomes are collected in Table 14.4, and Figure 14.2 draws its peak column as bars.

14.7.1 The Count That Keeps Rising

In case 1 the dial set to 0.7 on day 39 turns the true incidence on that same day, but the reported 7-day average keeps rising for 20 days, to 1,794 a day on day 59, and the census peaks at 1,435 beds on day 61.75. The delays of Section 14.4 add up: the dial's lag, the onsets still to come from everyone

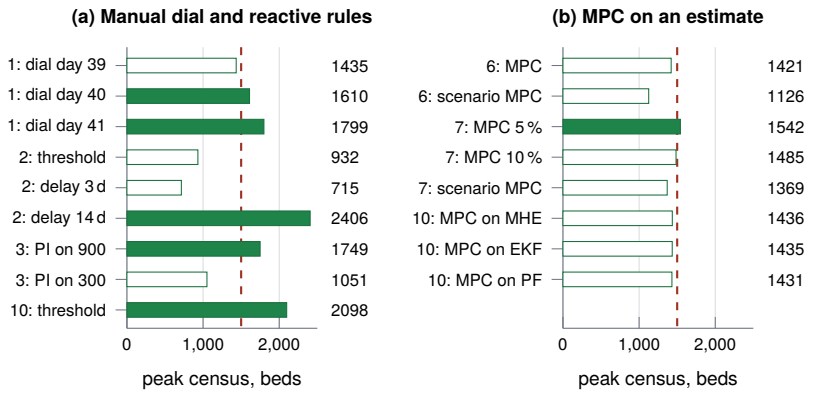


Figure 14.2 The peak census of Table 14.4, case by case (the number before each label), against the capacity of 1500 beds (dashed line). Solid bars pass the capacity, white bars stay under it. The uncontrolled peak of 7201 beds is off the scale. Notice that a day's delay on the dial, a 14-day reporting delay, the PI loop on its target of 900 and the threshold at $R_0 = 4$ carry the reactive peaks past the line, while every MPC on an estimate stays under it except the nominal MPC at its 5 % margin when the variant of case 7 arrives.

already exposed, the 7-day report delay and the 7-day average. The peak moves by about 182 beds for each day the dial waits, to 1,610 beds on day 40 and 1,799 on day 41, which stays over the capacity for 23 days, and the reported average still peaks on day 59 in every one of the three runs.

> **TRY IT IN THE APP** **1 · The Uncontrolled Wave and Its Two Lags**
>
> Load the case and press run with the Intervention dial at 0. Watch the Curves view: the reports trail the dashed true onsets by the 7-day reporting delay, and the census in the right plot climbs past the capacity line of 1,500 beds. Left alone, it peaks at 7,201 beds on day 73.

14.7.2 A Relay and a PI on Late Counts

The threshold of case 2 switches on at day 37, off at 104, on at 129, off at 181 and on at 210, periods of 92 and 81 days, with the census peaking at 932 beds on day 55 and a cost of $4,768 million over 240 days. Lengthening the reporting delay to 14 days stretches the first period to 117 days and lets the census reach 2,406 beds, and shortening it to 3 days gives periods of 80 and 69 days and a peak of 715. The period of a relay on a lagged signal follows the lag. The PI on the census in case 3 first moves the dial on day 37 and reaches the full dial on day 47, when 1,019 beds are already counted, and the census peaks at 1,749 beds on day 57, 1.94 times its target and over the capacity for 22.9 days. A target of 300 beds still peaks at 1,051, 3.5 times the target. Feedback on the census overshoots by the lag, which is built into the count.

14.7.3 Three Estimators Anchored by the Census

On the steady record of case 4, from day 30 on, every daily estimate of R_t is within 0.083 of the truth for the EKF, 0.061 for the PF and 0.082 for the MHE, with a mean error of 0.021, and each estimate of ρ is within 0.065. All three are within 0.2 on all 99 days from day 21, and over seven seeds the three bands contain the true R_t on every day.

Case 5 cuts testing in half on day 70. The reported 7-day average falls from 1,123 a day on day 69 to 490 on day 77 while the census falls only from 996 to 939. The MHE's bound on ρ might be expected to separate the estimators here, and it does not. All three read the cut as testing, with R_t within 0.076 (EKF), 0.097 (PF) and 0.073 (MHE) of the truth from the cut to day 120, and each $\hat{\rho}$ within 0.03 of the truth from day 92 (91 for the MHE). The census pins the number infected, and

in this model only ρ can jump overnight, so the explanation the data leave is a testing change. The bound acts only late and barely. On day 102 the EKF's $\hat{\rho}$ dips to 0.089 and the PF's to 0.086, below the 0.1 floor that stops the MHE's at 0.100, but the drifting true ρ is itself 0.103 that day. Over eight records the EKF crosses the floor on three, and on two of those the true ρ also falls below 0.1 (to 0.074 and 0.092), where the bound holds the MHE away from the truth. A bound is information, and here it is sometimes wrong.

Pause. Try Exercise 14.2 before reading on.

The census decides this result, and the result depends on what the estimators are told about it. With the census withheld from the estimators none finds the cut: the EKF's $\hat{\rho}$ stays above 0.25 and the PF's above 0.22 against a true ρ between 0.10 and 0.17, the MHE's wanders as low as 0.11 and back, the R_t errors grow to 0.22 (EKF), 0.32 (PF) and 0.24 (MHE), and the diverged-estimator alarm, judged on the census they no longer read, lights for all three. The bound cannot help there because the wrong answer lies inside the band. With p_h declared 20 or 30 percent low or high, every R_t stays within 0.089 while $\hat{\rho}$ scales with the error (0.74 to 1.39 times the truth for the EKF and the PF), and a census at 5 % noise with 1 % declared raises the R_t errors to 0.39 (EKF), 0.20 (PF) and 0.33 (MHE). Declared honestly, census noise of 3 to 10 % keeps every R_t error under 0.13.

TRY IT IN THE APP 5 · The Testing Cut: The Count Halves, the Census Does Not

Load the case and press run. Either side of day 70, watch the Reports 7-d avg and Census badges: the reported average falls by more than half within a week while the census hardly moves. Read the three estimates of R_t in the middle plot through the cut. Then set Census to the estimators to Off, press Reset and run again: none of the three finds the cut.

14.7.4 Where the Bound Matters

Case 8 stops the reports for three days and then publishes the held 2,505 with the day's own 1,130 as a dump of 3,635. A day of zero reports lies 4.42 standard deviations below its expected count, and the dump 15.98 above. The estimators' $\hat{\rho}$ through the outage is shown in Figure 14.4a. The EKF's falls to 0.142 through the zeros, jumps to 0.531 on the dump and is not back within 0.05 of the truth until day 91. The PF's falls to 0.152 and is back by day 74. On its own the dump would leave the PF 2.94 effective particles of 500, and the tempered update takes it in up to 6 stages, never with fewer than 119.8. The MHE's stops at its 0.1 floor on day 72, the second zero, and once the dump is in its window it reads the zeros and the dump together and is back by day 77. The same MHE with its band loosened to 0 and 5 falls to zero, where the reports no longer inform the onsets, and is not back until day 90. That loss of 13 of its 14 days of lead over the EKF is the bound's worth, and any floor from 0.05 to 0.12 keeps it, while a floor of 0.01 recovers only on day 87. With 50 or 200 Gauss–Newton iterations instead of 12 the bounded and loosened recoveries are unchanged, so the loosened MHE's slow return comes from the problem and not the solver.

What the bound protects is the estimate of ρ, which no policy reads. Through days 70 to 100 every $\hat{R}_t$ stays within 0.054 (EKF), 0.117 (PF), 0.050 (MHE) and 0.061 (loosened MHE) of the truth, and every estimate of $E + I$ between 0.936 and 1.073 times the truth, because the census carries the infections through the outage. The scene at day 80, with the EKF still at 0.47, is shown in Figure 14.3.

14.7.5 Planning Against the Beds and a Variant

The MPC on the MHE in case 6 lets the census rise to its planning ceiling and rides it while the dial eases with the immunity the epidemic builds. The census peaks at 1,421 beds on day 98, 94.7 % of the capacity, for $1,788 million over 180 days. The scenario MPC on the same case peaks at 1,126 beds for $2,198 million, the price of a hedge against a variant that does not come.

In case 7 a variant multiplies transmission by 1.5 on day 110. The nominal MPC holds the census at 1,392 beds that day, close to its 1,425-bed ceiling, and the variant takes it past the capacity on day 119.6 and to 1,542 beds, over the capacity for 33.05 days, for $2,440 million. The estimators' $\hat{\beta}$ is

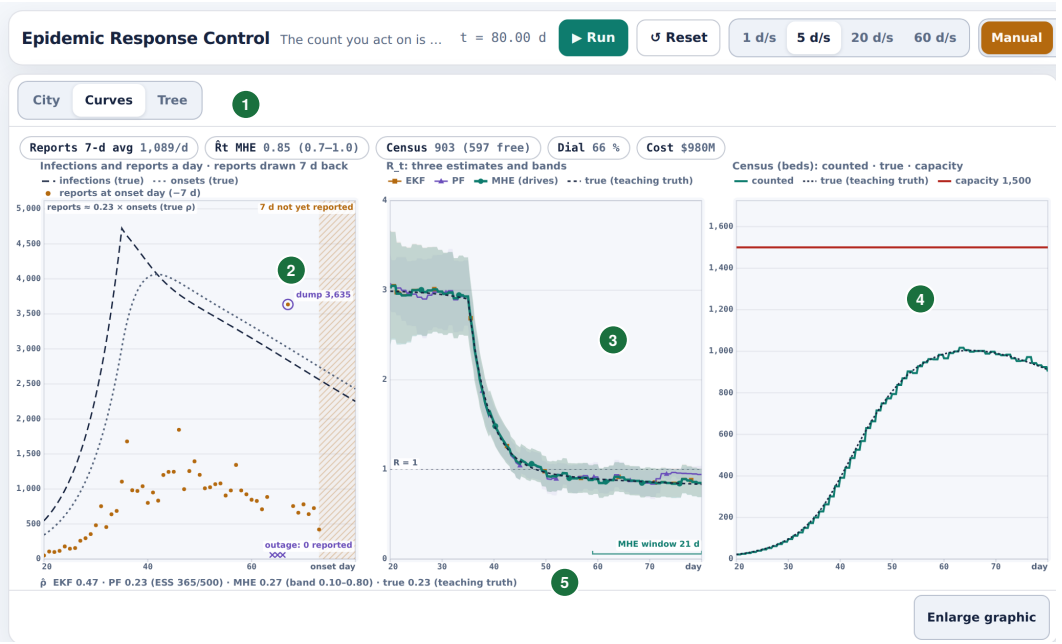


Figure 14.3 The reporting outage of case 8 at $t = 80.00$ d on the Curves view, in Manual with the dial at 66 %. The crop keeps the header and the plant view and cuts the side and lower panels. Callouts: (1) the badges, (2) the dump, (3) the estimates of R_t, (4) the census and the capacity, (5) the line of estimates of ρ. The left plot draws the true infections and onsets (dashed, teaching truth) and each day's report at the onset day it samples, with the three zeros as crosses on the axis, the dump of 3,635 ringed and the last 7 days hatched as not yet reported. The middle plot draws the three estimates of R_t with their bands against the truth, the MHE's 21-day window bracketed, and the right plot the census at 903 counted against the 1,500-bed capacity. The line under the plots shows the EKF's $\hat{\rho}$ still at 0.47 from its jump on the dump, the MHE's at 0.27 after its days on the floor and the PF's at 0.23, against a true 0.23, all made from day 79's data. The badges read a reported 7-day average of 1,089 a day, $\hat{R}_t$ 0.85 from the MHE against a true 0.84 on day 79, and a cost of $980 million.

a quarter higher only 10 days after the variant for the EKF and the MHE and 19 days after for the PF. The scenario MPC holds the census to 1,369 beds for $2,855 million because its shared week kept the census at 1,066 beds on day 110, a margin the variant used before any estimator saw it. A hand-set margin does the same. The nominal MPC with a 10 % planning margin holds at 1,485 beds for $2,599 million. What the tree adds is where its margin comes from, a declared spread of futures and the week before recourse, rather than a number chosen by hand, and at $\times 1.6$ it is the dearer insurance. The tree ends at its cap of 4 linearizations on 178 of 180 days of case 7. Re-solved from the same warm start to 30 linearizations, its first move differs by at most 5.07 points of the dial (median 0.27), and the nominal MPC's by at most 0.33.

14.7.6 The Graded 180 Days

The challenge cuts testing on day 60 and brings a variant on day 110. Its target is a true census never above 1,500 beds, a cost within $3,600 million and the selected estimator's R_t within 0.2 of the truth on at least 90 % of the days from day 21, all judged through a 28-day hand-over in which the city runs on with the dial held where it stands on day 180. The hand-over catches a dial dropped near the end: the suggested run with its dial set to zero on day 170 passes the 180 days alone at 1,375 beds

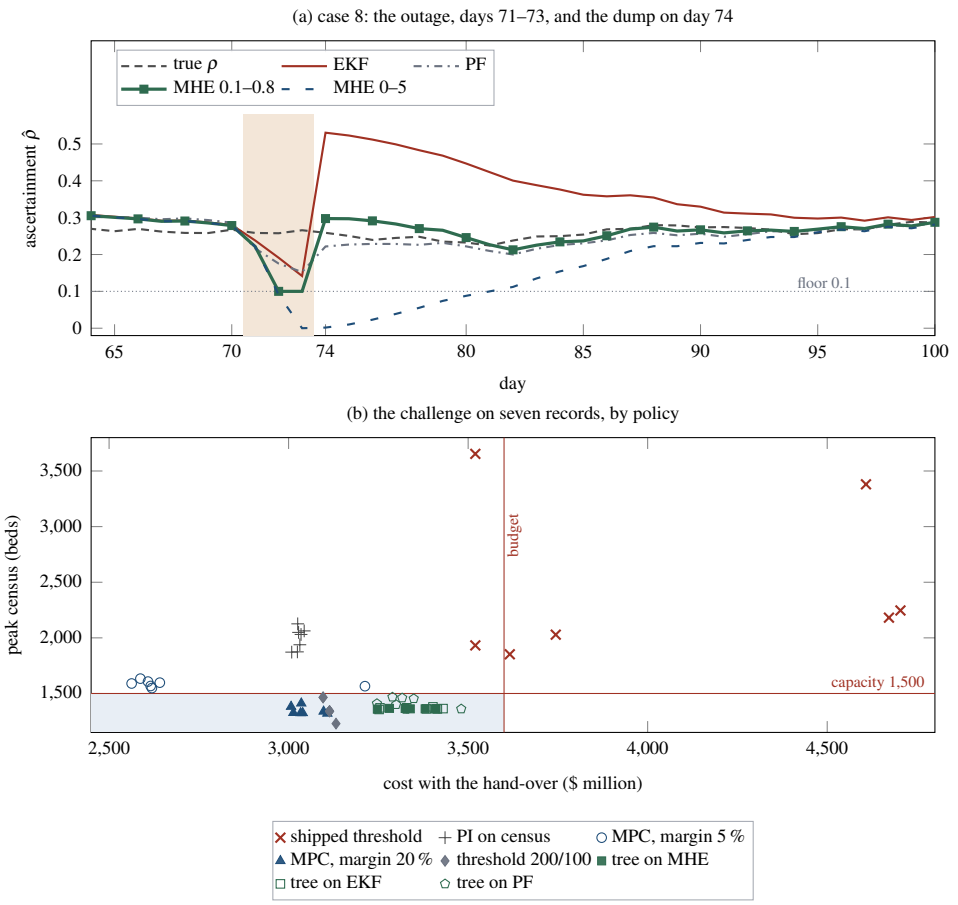


Figure 14.4 The bound and the margin. (a) Each estimator's $\hat{\rho}$ made from a day's data against that day's true ρ (densely dashed) in case 8, with the outage shaded: the EKF solid, the PF dash-dotted, the MHE with its shipped band of 0.1 to 0.8 solid with squares, and the MHE with its band loosened to 0 to 5 loosely dashed. The dotted line is the floor. (b) Peak census against cost on the challenge's seven records for each policy, one marker shape per policy, both through the 28-day hand-over, with the passing region shaded.

Table 14.5 The Challenge on Seven Paired Records (Seeds 109 and 1–6). Peak and Cost Include the Hand-Over

Policy	Pass	Peak (beds)	Cost ($M)
threshold on the EKF (shipped)	0/7	1852–3654	3520–4704
PI on the census	0/7	1872–2126	3009–3043
MPC on the MHE, 5 % margin	0/7	1549–1634	2563–3213
scenario MPC on the MHE (suggested)	7/7	1359–1366	3247–3411
scenario MPC on the EKF	7/7	1356–1380	3252–3430
scenario MPC on the PF	7/7	1357–1468	3246–3481
MPC on the MHE, 20 % margin	7/7	1320–1412	3007–3108
threshold 200/100 a day at 0.6	7/7	1229–1465	3096–3132

and $2,684 million and reaches 4,432 beds in the hand-over. The policies on seven paired records are compared in Table 14.5 and Figure 14.4b.

As shipped, the threshold switches off on day 85, when the halved count falls under 100 a day, and on again only on day 109. By day 110, 29,312 people are infectious, the variant arrives that

day and the census reaches 2,029 beds on day 128, for $3,744 million. The scenario MPC on the MHE holds the census to 1,362 beds and spends $2,868 million plus $471 million in the hand-over, $3,339 million in all, with R_t within 0.2 on 95 % of days. On the shipped EKF the same tree holds 1,364 beds for $3,430 million. The estimator does not decide this challenge. The tree passes on every estimator and record, its runs meet the R_t test on 93 to 96 % of days, and the nominal MPC at its 5 % margin breaches on every record. The margin decides it. The nominal MPC with a 20 % margin passes all seven records for $3,007 to $3,108 million, less than the tree on every record, and the threshold retuned to 200 and 100 reports a day at 0.6 passes all seven too. The R_t test is not idle: with the random walk on β raised to 20 % of its prior a day, the suggested tree holds the beds and the budget and fails on R_t alone, at 84 % of days. The held dial of the hand-over and the policy left running for 28 days give the same verdicts for the shipped and the suggested runs.

> ✦ DID YOU KNOW? **A Plain Threshold, Retuned**
>
> From the simulator: retuned to switch on at 200 reported cases a day and off at 100, with the dial at 0.6, the plain threshold passes all seven records of the challenge and peaks at 1229 to 1465 beds. The nominal MPC at its shipped 5 % margin passes none (Table 14.5).

14.7.7 A Transmissibility No Reactive Rule Holds

At $R_0 = 4$ in case 10 the infections double every 3.4 days, and the shipped threshold without an estimator lets the census reach 2,098 beds on day 51. A sweep in the app's own tests runs 350 threshold settings, 294 PI settings and 1,313 constant dials (101 levels from 13 start days), each judged on the challenge's beds and budget through the hand-over. None holds the beds within $3,600 million, and the cheapest that holds the beds costs $4,077 million, while at $R_0 = 3$ the same sweep finds 126 that pass. A policy that reads only the reports or the census either acts late and overflows or acts early and stays on too long. With an estimate of the present the MPC holds the beds on every estimator (1,436 beds and $2,783 million on the MHE). The claim is scoped. Of 17,732 two-step hand schedules, 6 pass, but each was found by rerunning this same epidemic, and at $R_0 = 4.5$ none of the six holds.

14.7.8 Checks, Published Relations and the Limits of the Evidence

The first question, code verification, is whether the engine solves the equations above. The compartments sum to N within 7.9×10^{-15} of N over 400 days. The uncontrolled run reproduces the conserved quantity's peak $E + I$ to 0.0002 % and the final size to all six digits printed, and the peak census of 7,201.1 beds on day 73.10 agrees with an independent fourth-order integration at 0.002 d to 0.000 04 %. Halving Δt moves the peak census by 0.0000 % in cases 1 and 6, by −0.023 % in case 7 under the tree and by −0.016 % in the suggested challenge (1,362.0 to 1,361.8 beds), with every verdict unchanged. Over 300,000 fresh draws the negative binomial's mean is within 0.08 % and its variance within 0.45 % of the declared values at means of 50, 500 and 2,000. Over seven seeds the bands contain the true R_t on 93.9 % (EKF), 88.7 % (PF, a 5 to 95 % band) and 93.6 % (MHE) of the challenge's days. Averaged over four filter seeds, the 500-particle filter's $\hat{\rho}$ stays within 0.015 of a 20,000-particle filter's from the dump to day 85. On a 6-day window none of 200 random feasible neighbors of the MHE's solution has a lower cost, and on a 21-day one-branch instance the MPC's plan, at $122.6 million, costs less than the best of 2,601 two-block plans on a grid that hold the ceiling, $124.7 million. All 50 of the app's checks pass, 32 written for this city and 18 shared with the other apps. In a desktop browser at 60 days a second, the suggested challenge's 180 days take 3.06 s, the day's three estimators and tree a median of 5.3 ms (the MHE 2.1 ms, the tree 2.5 ms), and frames hold at 16.7 ms.

Agreement with published relations is narrower. The final-size and peak relations are those of the compartmental literature [10, 12, 13], and the early growth rate of the uncontrolled city, fitted over days 8 to 18, is within 0.23 % of the generation-interval relation at $R_0 = 3$ and within 0.47 % at

$R_0 = 4$, the residual being the 0.3 % and 0.7 % of the city already infected by day 18. No parameter was fitted to any outbreak, so the model is verified but not validated. The app's effect on learning has not been evaluated.

14.8 The City on the Estimation Ladder

On a ladder of estimation problems the city stands one rung above fitting a model to a record and one below multirate estimation, such as the kraft digester of Volume 5, Chapter 4, whose analyzer and laboratory report on different clocks. Its estimators and planners build on the options of Section 1.6.3, and its plant is one that a student of public health already knows. A simulation can draw beside every count the truth that no surveillance system has [20], here the infections, the onsets, the ascertainment and the true R_t. Three cases carry most of the lesson. Case 1 puts a dial and a count that keeps rising for 20 days on the same screen. Case 5 turns the census on and off and shows what an anchor is worth, and case 8 shows the one surprise where a bound earns its place. A testing cut that halves the ascertainment overnight, or three days of zeros followed by an exact dump, would not happen so cleanly in a real city, but each isolates one mechanism. At its fastest speed a 180-day case plays in about 3 s, and the export carries every estimate beside the truth, each true value marked as such by its column's name so that no replay mistakes it for a measurement.

As in every app of the series (Section 1.6.4), a case counts as done only when the city responds to something the learner did, and an event the case scripts is never taken for the learner's own action. Case 5, for example, asks for a run past day 90 with the census in use, then the census switched off, a Reset and a second run past day 90, and case 7 asks for a breach under the nominal MPC, then the tree, a Reset and a run to day 180 within the capacity. Reset keeps the learner's settings (Section 1.6.1), and a change made before Run sets the initial state instead of acting as a step. The challenge is graded by the same function the scorecard shows, and changing a model parameter voids the attempt.

14.9 Technology in Practice

The city of Section 14.3 publishes two counts and is moved by one dial. A real public health system has more streams, a set of measures in place of the dial, and people where the policy block sits. Figure 14.5 draws the loop that way.

What Measures It

During the coronavirus disease 2019 (COVID-19) pandemic, surveillance data on the number of cases, hospitalizations and deaths were used to inform response policies [21]. Planning guidelines name further indicators that a state or locality might watch, among them visits to health care providers for influenza-like illness, the geographic spread and school absenteeism [2]. Each stream has its own delay and sees its own share of the infections. The chapter keeps two: the reports stand for the case stream, with a fixed 7-day delay and a hidden ascertainment near 0.30, and the census for the hospital stream.

> **TECHNOLOGY** **Sensor: Case Reports and the Hospital Census**
>
> Surveillance data on cases, hospitalizations and deaths inform response policies, but they give only a partial, time-lagged picture of transmission [21]. A case is counted after an infection, an onset, a test and a report, so the series of new infections has to be rebuilt from observations made long after transmission. The share of infections found moves with test availability, testing criteria and interest in testing. Hospitalizations are less sensitive to testing effort and more sensitive to other factors, such as the beds available [22]. A change in testing therefore shows on a trend as a step in the reports that the census does not share, the signature of case 5. In the chapter the reports sample the onsets of 7 days before at a drifting share, and the census is counted daily within 1 %.

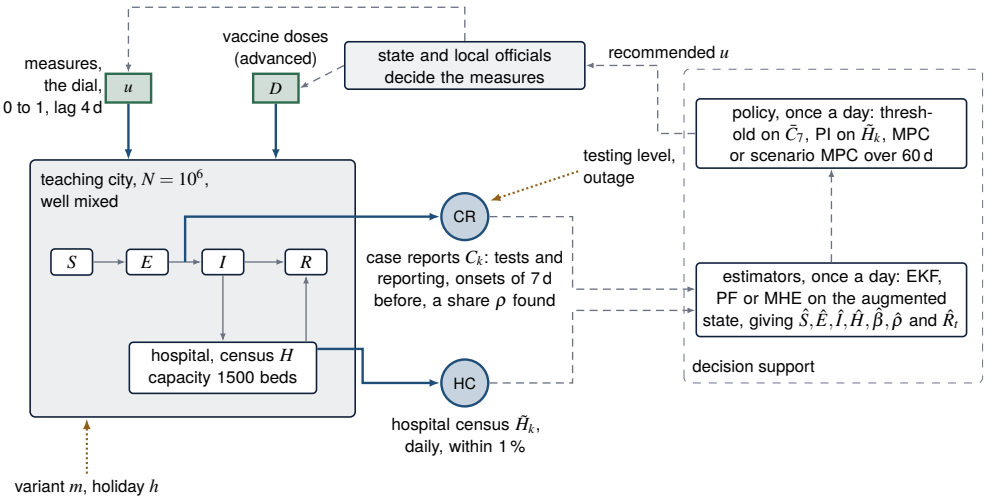


Figure 14.5 The loop of the chapter as a public health system runs it. The city publishes two surveillance streams, case reports (CR) a week late and only a share of them, and the hospital census (HC), counted daily. Estimators and a policy run once a day in a decision-support layer (dashed frame), and people, not the layer, decide the measures. Notice that the testing level and an outage act on the case reports alone, while the census reaches the estimators untouched by either.

What Moves It

The measures are phased in, or layered, on the basis of the pandemic's severity and the local transmission over time, and the more disruptive community measures are decided by state and local officials on the conditions of their jurisdictions. Their timely and effective use depends on community acceptance and active participation [2], which no actuator model carries. The dial of Equation (14.1) compresses all of this into one number between 0 and 1, with the lag τ_u and a quadratic cost.

> **TECHNOLOGY** **Final Element: Community Mitigation Measures**
>
> The handle a city moves is a set of nonpharmaceutical interventions. The CDC guidelines of 2017 group them as personal protective measures, such as home isolation of the ill and hand hygiene, community measures that increase social distancing, such as closing or dismissing schools, distancing in workplaces and postponing or cancelling mass gatherings, and environmental measures, such as routine cleaning of surfaces. Each is only partially effective, so plans call for several at once. Early in a pandemic they may be the only prevention tools at hand, because a pandemic vaccine can take up to six months to produce [2]. The chapter's dial stands for every measure at once, with one lag of four days and a quadratic cost.

Where the Controller Runs

No programmable controller closes this loop. The estimators and planners of this chapter belong to a decision-support layer under the officials who decide. Forecasting models already serve that layer: their predictions informed short-term decisions such as health care staffing, school closures and the allocation of medical supplies [21]. A model-based study of the same problem estimated the hidden states every day but revised its policy only every 25 days, because planned suppression strategies likely cannot be altered quickly [6]. The chapter's policy executes once a day, the period at which its counts arrive. Whether, and how far, any such rule should steer real measures is a policy question that the simulator does not answer.

What Is Modernizing It

The first development is wastewater surveillance. It measures community infection trends independently of health care seeking, access to care and testing capacity, and it has shown trends before case reports and hospital admissions. One statewide network sampled 65 treatment plants twice a week, analyzed the samples within 3 to 4 days and posted results daily. In another state, case rates fell while the number of people tested also fell, and steadily falling levels in wastewater showed that the decline was real [23], the question of case 5 answered by a third stream. The second is collaborative forecasting, in which a hub combined the predictions of many groups into an ensemble [21]. The third is model-based planning with state estimation and periodic revision of the policy [6], the structure this chapter teaches. None of them removes the delay between an infection and its count. Wastewater shortens it, and the other two plan around it.

> **IN PRACTICE** **Triggers Written Before the Pandemic**
>
> Community measures are planned before a pandemic arrives. The guidelines ask planners to identify activation triggers so that measures start before explosive growth, to plan for long school dismissals, and to weigh the social and economic costs of the measures against their public health benefits. Possible indicators include more visits to health care providers for influenza-like illness and a wider geographic spread, and the choice differs among states and localities with their resources. Rising influenza-associated hospitalizations are among the indicators that might confirm that measures should continue [2]. The threshold policy of case 2 is a trigger of this kind, written as a rule on the reported count.

14.10 Limitations

The city is a declared teaching model, well mixed, without age structure, waning immunity or deaths, with exponential stays and a fixed reporting delay, so it says nothing about where an epidemic spreads or who is admitted. The estimation results rest on the estimators knowing p_h, the census noise and the weekday pattern exactly, and Section 14.7 shows how a wrong p_h scales the ascertainment estimates and how an undeclared noisy census spoils R_t. The particle filter's tempering is a regularized approximation, checked against a 20,000-particle filter on one case only. The tree's protection is a margin. A nominal MPC with a hand-set margin passes the challenge for less, and the tree is preferred here only because its margin follows from a declared spread of futures. The MPC stops at 4 linearizations on most days of the tree, with a first move within about 5 points of a converged plan, its plan eases at the end of its horizon, and a policy switched on in mid run starts from a cold plan that can be far off for its first day. One recorded epidemic per case is graded, and seven records are checked for the challenge. The engine logs each whole-day row before the controller's execution at that instant, so an exported estimate must be read against its `est_day`. A simulated pass is not a public-health recommendation.

> **OPEN CHALLENGE** **Estimating Transmission in Real Time**
>
> Policy makers and public health officials use the effective reproduction number to judge whether interventions work. Two challenges are common to every method of estimating it: specifying the generation interval, and rebuilding the series of new infections from observations made long after transmission. Naive corrections for the delay, such as subtracting delays drawn from a distribution, can bias the estimate [22]. Methods that assume cases, hospitalizations and deaths all arise from one infection process might infer the infections more accurately while estimating the delays and the changes in ascertainment of each source [22]. Progress would be such joint estimates, tested where the truth is known, as it is in this chapter's city.

14.11 Conclusion

A city that acts on week-old, partial counts is an estimation problem before it is a control problem, and this chapter puts a declared compartmental city, its reports and its census, three estimators and four policies on one screen with the truth beside every count. A dial turned the day after the reports

pass 300 a day still sees them rise for 20 days, and feedback on the census peaks at 1.94 times its target. With the census to anchor them the EKF, the PF and the MHE all read a testing cut as testing, within 0.10 in R_t, and the MHE's bound earns its place only on a reporting outage, where it brings the ascertainment estimate back on day 77 against the EKF's 91 while every R_t stays within 0.12. A scenario tree holds a variant that breaches the nominal MPC, and so does a 10 or 20 percent margin, for less. At $R_0 = 4$ no threshold, PI or constant dial in a sweep of 1,957 holds the beds within the budget, while the MPC on any estimate does. The model is well mixed and every value is a teaching value. A fuller model would add age groups and waning immunity, delays as distributions, a census whose noise the estimators must learn, and a tree whose branches are revealed when the data can tell them apart. The simulator shows that what a constrained estimator or a scenario tree is worth depends on what else the measurements already pin down.

KEY POINTS

- A city that acts on its case reports acts on the epidemic of about two weeks before, and a dial turned the day after the reports pass 300 a day still sees them rise for 20 days.
- Feedback on the census peaks at 1.94 times its target, because the lag is built into the count.
- With the census to anchor them, all three estimators read a testing cut as testing. The MHE's bound earns its place only on a reporting outage.
- A scenario tree holds a variant that breaches the nominal MPC, and so does a 10 or 20 percent margin, for less.

Where else this applies. Volume 5, Chapter 4 estimates a plant from a late laboratory result, and Volume 9, Chapter 7 plans on a scenario tree of inflow forecasts.

References

[1] R. J. Hatchett, C. E. Mecher, and M. Lipsitch, "Public health interventions and epidemic intensity during the 1918 influenza pandemic," *Proceedings of the National Academy of Sciences*, vol. 104, no. 18, pp. 7582–7587, 2007.

[2] N. Qualls, A. Levitt, N. Kanade, N. Wright-Jegede, S. Dopson, M. Biggerstaff, C. Reed, and A. Uzicanin, "Community mitigation guidelines to prevent pandemic influenza—United States, 2017," *MMWR Recommendations and Reports*, vol. 66, no. 1, pp. 1–34, 2017.

[3] J. F. Lawless, "Adjustments for reporting delays and the prediction of occurred but not reported events," *Canadian Journal of Statistics*, vol. 22, no. 1, pp. 15–31, 1994.

[4] E. L. Haseltine and J. B. Rawlings, "Critical evaluation of extended Kalman filtering and moving-horizon estimation," *Industrial & Engineering Chemistry Research*, vol. 44, no. 8, pp. 2451–2460, 2005.

[5] C. V. Rao, J. B. Rawlings, and D. Q. Mayne, "Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations," *IEEE Transactions on Automatic Control*, vol. 48, no. 2, pp. 246–258, 2003.

[6] C. Tsay, F. Lejarza, M. A. Stadtherr, and M. Baldea, "Modeling, state estimation, and optimal control for the US COVID-19 outbreak," *Scientific Reports*, vol. 10, no. 1, p. 10711, 2020.

[7] J. Köhler, L. Schwenkel, A. Koch, J. Berberich, P. Pauli, and F. Allgöwer, "Robust and optimal predictive control of the COVID-19 outbreak," *Annual Reviews in Control*, vol. 51, pp. 525–539, 2021.

[8] M. M. Morato, S. B. Bastos, D. O. Cajueiro, and J. E. Normey-Rico, "An optimal predictive control strategy for COVID-19 (SARS-CoV-2) social distancing policies in Brazil," *Annual Reviews in Control*, vol. 50, pp. 417–431, 2020.

[9] A. Cori, N. M. Ferguson, C. Fraser, and S. Cauchemez, "A new framework and software to estimate time-varying reproduction numbers during epidemics," *American Journal of Epidemiology*, vol. 178, no. 9, pp. 1505–1512, 2013.

[10] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM Review*, vol. 42, no. 4, pp. 599–653, 2000.

[11] J. Wallinga and M. Lipsitch, "How generation intervals shape the relationship between growth rates and reproductive numbers," *Proceedings of the Royal Society B: Biological Sciences*, vol. 274, no. 1609, pp. 599–604, 2007.

[12] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 115, no. 772, pp. 700–721, 1927.

[13] J. Ma and D. J. D. Earn, "Generality of the final size formula for an epidemic of a newly invading infectious disease," *Bulletin of Mathematical Biology*, vol. 68, no. 3, pp. 679–702, 2006.

[14] K. M. Gamado, G. Streftaris, and S. Zachary, "Modelling under-reporting in epidemics," *Journal of Mathematical Biology*, vol. 69, no. 3, pp. 737–765, 2014.

[15] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings F Radar and Signal Processing*, vol. 140, no. 2, p. 107, 1993.

[16] N. Oudjane and C. Musso, "Progressive correction for regularized particle filters," in *Proceedings of the Third International Conference on Information Fusion*, vol. 2. IEEE, 2000, pp. THB2/10–THB2/17.

[17] E. L. Ionides, C. Bretó, and A. A. King, "Inference for nonlinear dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 103, no. 49, pp. 18 438–18 443, 2006.

[18] M. Ellis, H. Durand, and P. D. Christofides, "A tutorial review of economic model predictive control methods," *Journal of Process Control*, vol. 24, no. 8, pp. 1156–1178, 2014.

[19] S. Lucia, T. Finkler, and S. Engell, "Multi-stage nonlinear model predictive control applied to a semi-batch polymerization reactor under uncertainty," *Journal of Process Control*, vol. 23, no. 9, pp. 1306–1319, 2013.

[20] J. Park, R. A. Martin, J. D. Kelly, and J. D. Hedengren, "Benchmark temperature microcontroller for process dynamics and control," *Computers & Chemical Engineering*, vol. 135, p. 106736, 2020.

[21] E. Y. Cramer, E. L. Ray, V. K. Lopez, J. Bracher, A. Brennen *et al.*, "Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States," *Proceedings of the National Academy of Sciences*, vol. 119, no. 15, p. e2113561119, 2022.

[22] K. M. Gostic, L. McGough, E. B. Baskerville, S. Abbott, K. Joshi, C. Tedijanto, R. Kahn, R. Niehus, J. A. Hay, P. M. De Salazar *et al.*, "Practical considerations for measuring the effective reproductive number, R_t," *PLOS Computational Biology*, vol. 16, no. 12, p. e1008409, 2020.

[23] A. E. Kirby, M. S. Walters, W. C. Jennings, R. Fugitt, N. LaCross, M. Mattioli, Z. A. Marsh, V. A. Roberts, J. W. Mercante, J. Yoder, and V. R. Hill, "Using wastewater surveillance data to support the COVID-19 response—United States, 2020–2021," *MMWR Morbidity and Mortality Weekly Report*, vol. 70, no. 36, pp. 1242–1244, 2021.

14.12 Exercises

14.1 ○ Concept check · 2 min

In case 1 the Intervention dial is set to 70 % on day 39, the day after the reported 7-day average passes 300 a day, and the true infections turn that same day. Why does the reported 7-day average keep rising until day 59?

(a) The dial's effect builds over four days, and the infections keep rising until it is complete
(b) The reports sample onsets a week old, from people infected before the dial moved
(c) The ascertainment drifts upward as testing catches up with a larger epidemic
(d) The reports count admissions, which follow the infections by about ten days

14.2 ○ Concept check · 2 min

In case 5 testing is cut in half on day 70, and the reported 7-day average falls from 1,123 to 490 a day within a week. With the census in use, all three estimators read the change as testing and not as a turn in the epidemic. What lets them do so?

(a) The MHE's floor of 0.1 on ρ, which the other two estimators share through the prior
(b) The weekday pattern of the reports, which the estimators know and the cut disturbs
(c) The census, which pins the number infected, so only ρ is left to explain the drop
(d) The 21-day window, long enough to hold the reports from before and after the cut

14.3 ○ Concept check · 1 min

The PI policy of case 3 moves the dial on the census count against a target of 900 beds. In the convention of Section 1.6.2, what is its action, and why?

(a) Reverse acting, $a = +1$, because the census rises when the dial is raised
(b) Reverse acting, $a = +1$, because the error is the target minus the census
(c) Direct acting, $a = -1$, because the dial is bounded below at zero
(d) Direct acting, $a = -1$, because a higher dial lowers the census

14.4 Theory and calculation · 5 min

The city has a latent period $1/\sigma = 4\,\mathrm{d}$ and an infectious period $1/\gamma = 6\,\mathrm{d}$. Early in the epidemic the infections grow at a rate r that satisfies $(1 + r/\sigma)(1 + r/\gamma) = R_0$.

(a) For $R_0 = 3$, find the transmission β_0, the growth rate and the doubling time. *Checkpoint:* $0.5\,\mathrm{d}^{-1}$, $0.1477\,\mathrm{d}^{-1}$ and 4.69 d.

(b) Repeat for $R_0 = 4$, the transmissibility of case 10. *Checkpoint:* $0.667\,\mathrm{d}^{-1}$, $0.2020\,\mathrm{d}^{-1}$ and 3.43 d.

(c) A student takes $r = \beta_0 - \gamma$, the rate of a model with no latent class. Find the doubling time it predicts at $R_0 = 3$, and say why it is wrong for this city. *Checkpoint:* 2.08 d.

14.5 Theory and calculation · 6 min

With the dial at zero, the quantity $s + e + i - \ln(s)/R_0$ is conserved, where s, e and i are the fractions of $N = 10^6$ in S, E and I. Take $R_0 = 3$ and a vanishing seed, so that $s = 1$ and $e = i = 0$ at the start.

(a) Find the largest number of people in E and I together. *Checkpoint:* About 300,500 people, $0.3005N$.

(b) Find the final size z from $z = 1 - \mathrm{e}^{-R_0 z}$ by fixed-point iteration from $z = 0.9$, to four digits. *Checkpoint:* 0.9405.

(c) Find the share still susceptible at the end, compare it with $1/R_0$, and explain the difference. *Checkpoint:* 0.0595 against 0.333.

14.6 Theory and calculation · 5 min

The daily report is negative binomial with mean μ and variance $\mu + \mu^2/\kappa$, with $\kappa = 20$.

(a) For $\mu = 1000$, find the standard deviation and the number of standard deviations by which a report of zero lies below the mean. *Checkpoint:* 226 and 4.43.

(b) Show that a report of zero can never lie more than $\sqrt{\kappa}$ standard deviations below its mean, and evaluate the bound. *Checkpoint:* 4.47.

(c) Case 8 finds its three zeros 4.42 standard deviations low and the dump 15.98 high. What does the bound mean for an outage alarm that flags any report more than 5 standard deviations from its mean?

14.7 App challenge · 8 min

Load scenario 1, The uncontrolled wave and its two lags. Use the default seed and set the speed to 1 d/s.

(a) Run until the Reports 7-d avg badge first shows 300 a day or more, and pause. Read the day. *Checkpoint:* Day 38.

(b) Before day 39 is reached, set the Intervention dial to 70 % and run to day 120. Read the Peak census tile, and find in the exported file the largest value of `reported_7d` and its day. *Checkpoint:* 1,435 beds on day 62; 1,794 a day on day 59.

(c) Press the reload button, run past day 39, pause before day 40, set the dial to 70 % and run to day 120. Read the Peak census tile. *Checkpoint:* 1,610 beds on day 62, 15 days over the capacity.

Accept a peak within 2 beds. State what one day of waiting costs and why the reports peak on the same day in both runs.

14.8 App challenge · 8 min

Load scenario 3, PI on the census: too late by construction. Use the default seed and run the case to its end at day 150, then export the file.

(a) Find the first row with `dial_pct` above zero, and the first row in which it reaches 100. Read `census_beds` in that second row. *Checkpoint:* Day 37; day 47 with 1,019 beds.

(b) Read the Peak census tile, and divide the peak by the target. *Checkpoint:* 1,749 beds on day 57, 23 days over; 1.94.

(c) Set the Census target (PI) to 300 beds, press Reset and run to day 150. Read the Peak census tile and divide by the new target. *Checkpoint:* 1,051 beds; 3.50.

Accept a peak within 2 beds. Explain why the lower target does not cure the overshoot.

14.9 App challenge · 9 min

Load scenario 7, The variant: nominal MPC breaches, the tree pre-empts. Use the default seed and run each case for 180 days. Read the Peak census and Intervention cost tiles each time.

(a) Run the case as shipped, with the nominal MPC and a 5 % planning margin. *Checkpoint:* 1,542 beds, 33 days over; $2,440 million.

(b) Set the Policy to Scenario MPC, press Reset and run again. *Checkpoint:* 1,369 beds; $2,855 million.

(c) Set the Policy back to MPC and the Planning margin to 10 %, press Reset and run again. *Checkpoint:* 1,485 beds; $2,599 million.

Accept a peak within 2 beds and a cost within $2 million. Which protection is cheaper, and what does the tree give that the margin does not?

14.10 Application area · 3 min

The estimators are told the hospitalized share $p_h = 0.03$ exactly. Suppose the true share is 0.03 but the estimators are told 0.021, 30 % low.

(a) By what factor does the number infected that the census implies change, and what does that do to the estimate $\hat{\rho}$? *Checkpoint:* About 1.43 times; $\hat{\rho}$ about 0.7 times the truth.

(b) Which of the chapter's estimates can a health department still publish, and which needs an independent check of p_h first?

14.11 Application area · 4 min

The dial costs $c_u u^2$ a day with $c_u = \$50$ million.

(a) The shipped threshold of case 2 sets the dial to 0.8 while its measures are on. Find the daily cost while on, and the number of days on that make up the $4,768 million of its 240-day run. *Checkpoint:* $32 million a day; 149 days.

(b) Find the constant dial that, held for the 180 days of the challenge and its 28-day hand-over, spends exactly the $3,600 million budget. *Checkpoint:* 0.588.

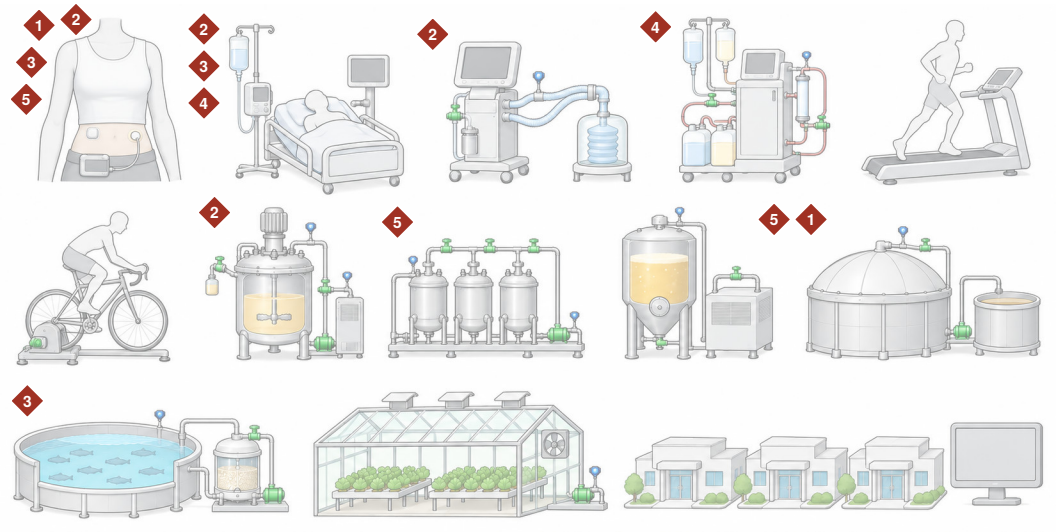
(c) Compare the cost of a dial of 0.5 for 60 days with a dial of 1.0 for 30 days. *Checkpoint:* $750 million against $1,500 million.

14.12 Review the agent · 5 min

An AI assistant is asked how to pass the challenge and replies as follows. "Testing is cut in half on day 60. The EKF carries ρ with no bound, so it reads the halved count as a collapse in transmission, its $\hat{R}_t$ drops and any planner fed by it relaxes too early. The MHE holds ρ in a band of 0.1 to 0.8 and cannot be fooled this way. Use the scenario MPC, and it is essential to switch the estimator to the MHE, because on the EKF the tree will fail." Find the flaw and say how you would check it.

CHAPTER 15

Summary and Outlook



Open gaps 1 unmeasured variability 2 virtual populations 3 one-way actuators 4 measurement checks 5 objectives

Notice that the insulin pump carries four gaps and the anesthesia bed three, and that no gap marks the athletes, the fermenter, the glasshouse or the city. Illustration, not to scale.

IN PLAIN TERMS **What Thirteen Living Plants Taught**

The main idea. Each chapter of this book put an automatic loop on something alive: a patient, an athlete, a culture of cells, a farm or a city. The loop sets a pump, a valve or a pace by itself. No two were alike, and what mattered was often measured late or not at all.

The challenge. A pump that adds a drug cannot take it back. A reading can look right while the body drifts away, and a score that treats a low and a high alike can pick the wrong loop.

Why it matters. The loops that worked measured the effect, checked their readings, kept a margin and were judged on what the plant is for.

Thirteen chapters put a loop on a patient, an athlete, a culture, a farm and a city. Section 1.4.7 read each plant in three parts: what varies between individuals or during the run, the limit the plant may not cross, and the structure that keeps it inside. This chapter adds the measured results to that reading and compares how each loop handled variability and limits it could not exceed. It then gives a short guide for choosing a structure, looks at where practice is heading, reports how the volume was made with AI agents and what they got wrong, and lists the problems that remain open.

15.1 What the Chapters Showed

Tables 15.1 and 15.2 extend Table 1.1. Each row names what varied and the limit that bound, a structure that held the plant, one that did not, and the measured result with the table of the chapter that reports it. Every number is the chapter's own, and Figure 15.1 draws one pair from each row.

The columns of the two tables make five points.

Table 15.1 How Each Loop Handled Variability and Its Limits, Chapters 2 to 8

Chapter	What varied, the limit	What held it	What failed	Measured result
2, insulin pump	Exercise raises insulin sensitivity unseen. The pump stops at zero, and a low costs more than a high.	Direct-acting PI with back-calculation. A predictive controller told of the meal.	A symmetric score, which ranks the exercise day above the large meal. The predictive controller not told.	In band 78.9 to 100 %. Exercise low 47.8 mg/dL at an IAE of 118.5 against 210.3 (Table 2.5). Meal IAE 50.94 told, 98.98 not told, 83.67 PID (Table 2.7).
3, anesthetic depth	Sensitivity 0.70 to 2.0 times the model's. A depth band, a pressure floor, and a way down only by elimination.	A depth loop that moves the target of a model-inverting pump. A signal-quality hold. A vasopressor loop.	Model inversion alone. A controller that trusts the monitor through cautery.	Index spread 18.3 open, 1.1 closed (Table 3.7). Cautery low 31.8 against 48.9 and 333 mg against 178 (Table 3.6). Challenge in band 17 to 91 % (Table 3.8).
4, ventilator	Compliance, resistance, effort and leak. A plateau limit of 30 cmH_2O.	Identification on every breath under a protective cap. Leak compensation with adaptive support.	Volume control on a stiff lung. Pressure control on the same lung. Volume control with a leak.	Plateau up to 31.5 at 420 mL, or 179 mL at a held pressure. With the cap 393 to 396 mL at 29.9 to 30.0 (Table 4.5). Hour in band 27.34 against 99.74 % (Table 4.6).
5, kidney circuit	Pump tubing, wear, scale drift and fouling. Filter and access pressures, and the patient's true balance.	A weighed balance with each scale checked against its pump. Overrides. Effluent headroom.	Pump counts alone. A balance loop on a drifting scale. A drift under the check's tolerance.	466.3 mL behind in a day on counts. Error held within 28.4 while the patient falls 304.3 mL behind (Table 5.6). 35 g/h never flagged (Table 5.4). Challenge −893.0 to −51.2 mL (Table 5.7).
6, runner	Runner, heat and fuel. Lactate, heart rate and core temperature, and an hour to run.	Two PI loops on a low selector with pace feedforward. A predictive controller with a rollout terminal value.	The selector without feedforward. A terminal value written as a frozen product.	9.584 km selector, 9.947 predictive, both compliant, 8.676 without feedforward (Table 6.8). 210 of the 363 m by 3 min (Table 6.9). Under-fueled 8.35 against 9.09 km (Table 6.10).
7, cyclist	Rider, and a work capacity that drains. Heart rate and lactate limits.	A power loop with two overrides on margin-scaled reset. A predictive controller on the grade underfoot.	The treadmill's fixed reset band. A target above critical power. The whole route given to the predictive controller.	Circuit 28.90 km PID, 29.05 predictive, 28.97 with the route, 28.80 and not compliant at 600 W (Table 7.8). Fixed band 18.50 km (Table 7.11).
8, antibody culture	An appetite that doubles, a cooled culture, probe drift. Glucose 1 to 6 g/L, osmolality.	A continuous feed with an hourly Raman trim. An osmolality cap.	A daily bolus. A capacitance feed without the Raman. A glucose target of 4.5 g/L.	Titer 4.14 by bolus, 34.8 h outside, 3.43 without the Raman, 5.03 with it (Table 8.5). Challenge 4.25 to 5.21 g/L (Table 8.6).

Table 15.2 How Each Loop Handled Variability and Its Limits, Chapters 9 to 14

Chapter	What varied, the limit	What held it	What failed	Measured result
9, insulin refolding	Folding constants and the harvest load. An unfolded ceiling, hold clocks and a 72 h campaign.	A charge scheduled on an assay-corrected estimate at a loose ceiling.	A tight ceiling. A late harvest. A plant that folds 20 % slower than the controller assumes.	5.56 against 5.47 g/h for a bolus. At 0.35 g/L a lot folds 79.6 %, 665 g expire and the campaign fails (Table 9.7). Slow folding 3.68 against 5.01 (Table 9.6).
10, lager fermenter	Yeast viability, mixing that stops with the gas. A recipe within 1 °C, diacetyl, shared glycol.	Time-proportioned PI on three probes. A rest ended on the corrected estimate. A glycol coordinator.	One mid-height probe. A rest ended on the day-old laboratory result. The estimate without its correction.	Batch done on day 12.57 with the cone 2.73 °C off, then day 10.47 within 0.52 (Table 10.6). Uncorrected, the beer ships at 0.123 mg/L (Table 10.5).
11, digester	Feed, temperature and a hidden inhibitor. A titration ratio of 0.4 and a fold at 287.6 m^3/d.	A sampled PI on the daily titration under a supervisor.	A constant feed. A ratio target of 0.28. A predictive feed under hidden inhibition.	Constant feed sours at \$166,751, the supervisor passes at \$428,198 (Table 11.6). Predictive feed 6 high reports against 1 and less revenue (Table 11.5).
12, fish farm	A biofilter that must grow, and the fish biomass. Nitrite, ammonia and oxygen.	A feed governor, an oxygen soft sensor and an economic ration.	The feeding table. Fingerlings stocked on new media.	Hours over 57.0 to 20.7 (Table 12.4). Challenge 69.8 h over, 1,359 killed and 0 of 11 seeds, then 6.6 h, none killed and 11 of 11 (Table 12.5).
13, glasshouse	Weather and its forecast. A daily mean temperature and an hourly humidity limit.	A planner over the forecast together with the humidity override.	Either one alone. Fixed setpoints.	Shipped \$4292, three targets failed. Planner alone \$3780 with 715 min humid. Override alone \$6312. Both \$5594 and pass (Table 13.6).
14, city	Reported share, testing and a variant. Hospital beds.	Estimators anchored by the census. A scenario tree, or a nominal plan with a 20 % margin.	PI on the census. A plan with a 5 % margin. The shipped threshold.	PI peaks at 1,749 beds for a target of 900 (Table 14.4). Shipped and 5 % margin pass 0 of 7 records, tree and 20 % margin 7 of 7 (Table 14.5).

Variability was handled by feedback on the effect, when the effect was measured. Model inversion in Chapter 3 delivered the modeled concentration to every patient and left the depth index spread over 18.3 units. Closing a loop on the depth monitor held the same patients within 1.1 units, because the loop moved the commanded concentration from 1.90 to 5.03 µg/mL. That is the P-1 example of Section 1.4.4 at full scale. The athlete chapters handled the individual with a schedule on one known number, threshold pace or critical power, and Chapter 4 identified the lung on every breath. Where the variation reached no measurement, nothing caught it in time. Exercise in Chapter 2 took glucose from 74.9 to 47.8 mg/dL before the controller next executed (Section 2.7), and a hidden inhibitor in Chapter 11 turned a predictive feed into the loser of the comparison.

A loop that holds its measurement need not hold the plant. The balance loop of Chapter 5 kept its measured error within 28.4 mL while the patient fell 304.3 mL behind, because it followed a drifting scale exactly. The ventilator of Chapter 4 read 1521 to 1830 mL behind a mask leak while the lung received 103 to 414 mL (Section 4.6). The anesthesia controller that trusted the monitor through cautery let the true index fall to 31.8. The remedy in each case was a second view of the same

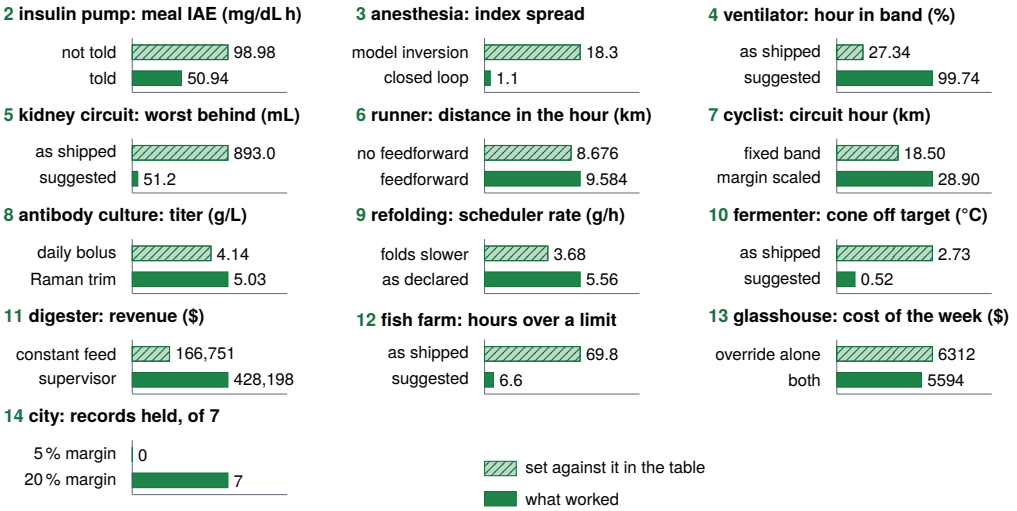


Figure 15.1 One pair of results from each chapter, with the numbers as the "Measured result" column of Tables 15.1 and 15.2 prints them. Each panel has its own scale from zero. A larger value is better in the panels of Chapters 4, 6, 7, 8, 9, 11 and 14, and a smaller one in the others. The kidney circuit's worst errors of −893.0 and −51.2 mL are drawn as amounts behind, and the city's 0 of 7 records draws no bar. Notice that in most rows the better result came from a change of structure or measurement, not of tuning alone.

quantity: the pump counts, the leak estimate, the signal-quality hold, three probes in the brewery tank and the census beside the case counts of Chapter 14.

One-way actuators were held by margin and by acting early. A pump that only adds leaves the way down to the plant. In Chapter 2 the pump sat at zero for 0.84 h while the insulin already given went on acting. In Chapter 3 the index fell to 35.5 a quarter of a minute after a closing, and only lightening six minutes ahead stopped the dip at 43.3 (Section 3.6). The governor of Chapter 12 and the humidity override of Chapter 13 act before their limits, not at them. The margin must cover the spread of the plant. On the epidemic's variant a 5 % margin breached every record and a 20 % margin held every one.

The objective decided what counted as held. A symmetric IAE in Chapter 2 ranked a run that fell to 47.8 mg/dL above one that never fell below 64 mg/dL, and the ranking reversed only at a low-side weight of 7.9. In Chapter 9 the best fold of a lot cost the campaign. In Chapter 11 the mean margin to the fold fell from 54.7 to 32.7 m^3/d as the revenue rose from $406,506 to $440,312 (Section 11.6). The glasshouse planner gave the cheapest week and failed the humidity limit. A score has to say which side of a limit matters and what the plant is paid for.

Prediction paid where it had a deadline and an anchored estimate. The predictive controller of Chapter 6 won the hour by 363 m, and that of Chapter 7 covered 29.05 against 28.90 km. More than half of the runner's lead was won in the first three minutes, a share the regulator recovers by shortening its reference ramp. In Chapter 14 no threshold, PI or constant dial in a sweep of 1,957 held the beds at $R_0 = 4$ within the budget, while the predictive controller did on every estimator (Section 14.7). Prediction lost where the model missed the biology. Without the meal announced, the predictive insulin controller scored 98.98 against the PID's 83.67, and when the Raman was lost in

Chapter 8 the model-based feed on the predictive controller's own estimate made 5.06 against its 5.01 (Table 8.7). The estimate carried most of the value.

15.2 Choosing a Method

The chapters suggest an order of questions for a loop on a living plant. Each question names the chapter where its answer decided the result, and Figure 15.2 draws the same questions as a flowchart.

1. *What varies between individuals, and does any measurement reveal it?* If the effect is measured, close a slow outer loop on it and let it correct the model's target (Chapter 3). If one number is known per individual, schedule on it (Chapter 6, Chapter 7). If nothing reveals it, carry margin, because no loop will learn it in time (Chapter 2).
2. *Is the measurement the plant?* Check each measurement the loop holds against a second view of the same flow or state, with a stated detection limit. In Chapter 5 a drift of 35 g/h passed the check and left the patient 379.3 mL behind.
3. *How old is the sample, and how fast does the plant grow?* The uptake between samples must fit inside the band, Equation (1.7). A culture held warm in Chapter 8 eats more in a day than its band is wide, so no daily bolus holds it (Section 1.4.5). Correct the model at the time of the draw.
4. *Can the actuator take back what it gave?* If not, size the margin for the spread of individuals plus the slow way back, Equation (1.9), and act on the event that will arrive, not on the limit it will reach (Chapter 3, Chapter 12).
5. *Which limit binds, and on which loop?* Give each limit an override on a selector, and test the selector's rules one by one. Removing the margin scaling of Chapter 7 cost 10.4 km of the hour.
6. *What is the plant paid for?* Score the two sides of a limit separately, Equation (1.10), judge the campaign rather than the vessel (Chapter 9) and the season rather than the day (Chapter 11).
7. *Is there a deadline, a forecast or an announcement?* Then a plan can earn its place (Chapter 6, Chapter 13, Chapter 14), provided its estimate is anchored by a measurement and its margin is sized for the futures it cannot see.

15.3 Where Practice Is Heading

Automated dosing at the bedside. Automated insulin delivery has reached clinical care. A joint consensus report of the European and American diabetes associations records that regulators have approved the first systems, that they have been adopted into care and that a community of people with type 1 diabetes builds its own, and it reviews the systems with a focus on their safety [1]. A meta-analysis of twelve randomized trials with 1284 patients found that index-guided closed-loop delivery of intravenous anesthesia used less propofol at induction, held the target depth more often and shortened recovery [2]. A survey of the field finds that anesthesia control has moved from single controllers to strategies of several components without yet reaching an integrated system [3]. Physiological closed-loop ventilation has moved from gas exchange toward preventing lung injury, early commercial systems are appearing, and many clinicians do not yet know them, which limits their adoption [4]. In the editor's view the lessons of Chapters 2 to 4 mark the edge of this work: an unannounced meal, a one-way pump, and a monitor that can be wrong while it looks right.

Evidence from simulation. Regulators have started to accept evidence obtained by modeling and simulation, and a framework for the credibility of such models asks for a stated context of use, a risk analysis that sets the acceptance thresholds, and verification, validation and uncertainty quantification [5]. The apps of this volume are verified and not validated, and each chapter says so. In the editor's view that distinction, made routinely and stated with every result, is the first thing a simulator used for clinical or manufacturing decisions must carry.

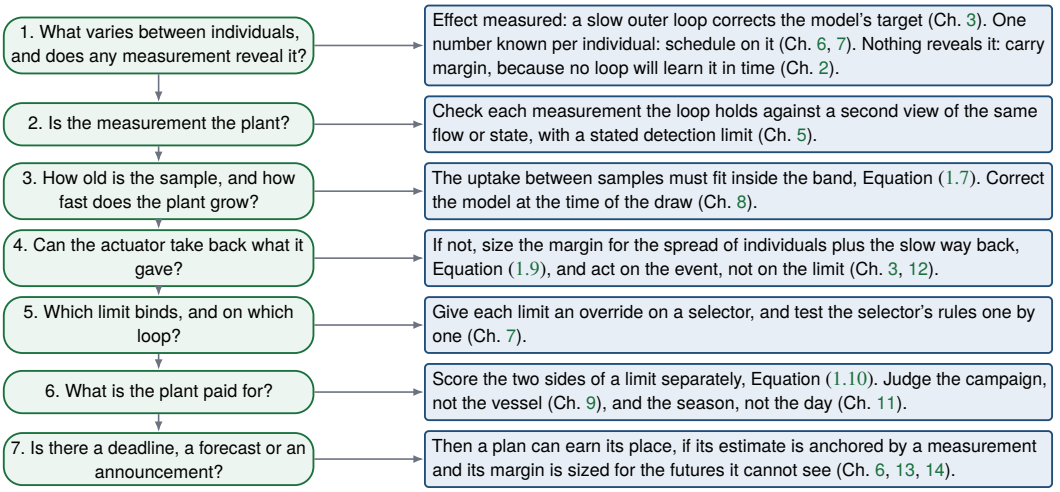


Figure 15.2 The seven questions of this section in order, with the list's answer and the chapters where it decided the result. Notice that the questions about the individual and the measurement come first, and that only the last question asks for a plan.

> ⧖ HISTORY **1978 to 2017: From Separate Pumps to One Machine**
>
> The first history box of Chapter 1 left the kidney circuit in 1978, balanced by gravity with scales and electronic clamps. A review written forty years after the first continuous treatment records what followed. Peristaltic pumps soon replaced those simple mechanisms, and veno-venous circuits took over with machines that carry a blood pump. The first such circuits joined separate pumps that could not communicate and operate together, with the possibility of amplifying errors. Integrated machines with four pumps and a pre-assembled circuit then made the therapy possible in almost every intensive care unit. The delivered dose still often differed from the prescription [6]. In the editor's reading the arc ends where Chapter 5 begins: one machine, one balance, and a check on every pump.

Models and spectra in biomanufacturing. A review of process models in biopharmaceutical manufacturing finds that the sector lags other industries in digitalization and names models that turn process data into decisions as key enablers under quality by design and process analytical technology [7]. Raman spectroscopy has been integrated into real-time release testing, continuous manufacturing and statistical process control, with a growing number of in situ applications in bioprocesses [8]. The hourly Raman trim of Chapter 8, worth 5.03 against 3.43 g/L of titer, is a small instance of that direction.

Autonomous growing and estimated epidemics. In the second autonomous greenhouse challenge, five teams grew cherry tomatoes for six months with climate and crop strategies set by their own algorithms, and every team outperformed the greenhouse operated by people [9]. For epidemics, guidance on estimating the effective reproductive number names two challenges common to all methods, the generation interval and the reconstruction of infections from observations made long after transmission, and warns that naive treatment of observation delays introduces bias [10]. That is the problem Chapter 14 solves first, before any policy acts. In the editor's view, control of living plants will move toward loops whose estimators carry most of the design, with the controller downstream of them.

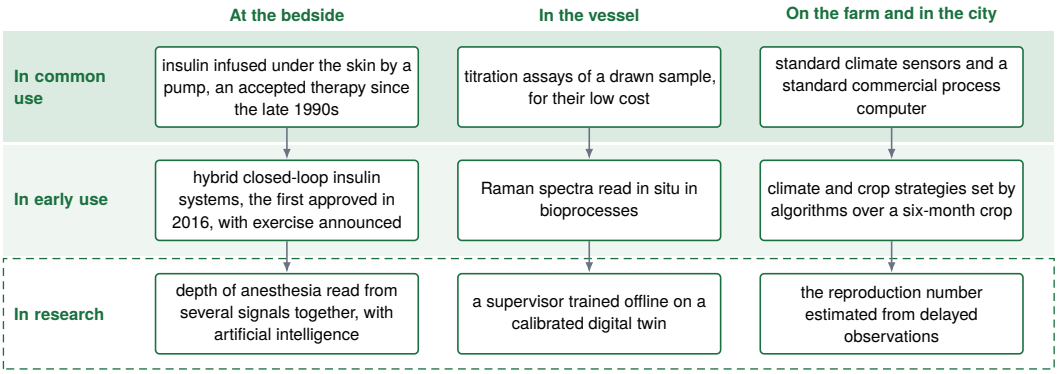


Figure 15.3 Developments in three threads. The top lane holds what the cited source calls standard, accepted or common, the middle lane what it reports in service or grown on a real crop, and the bottom lane what it reports from simulation or as a direction still to be shown, in the editor's reading. Sources by thread: at the bedside [11, 14, 15], in the vessel [8, 12, 13] and on the farm and in the city [9, 10, 16]. Notice that each thread moves from a reading taken by hand or by a standard instrument toward a decision made by a model.

Depth read from more than one signal. A review of depth-of-anesthesia monitoring notes that the indices in use reflect only the statistical features of cortical electrical activity. It expects multimodal monitoring, brain network analysis and artificial intelligence to improve the accuracy and the individualization of anesthesia management [11]. The depth loop of Chapter 3 closes on one index, and its signal-quality hold exists because that index can be wrong while it looks right. In the editor's view a loop on several signals will need the same check on each of them.

Online acids for the digester. A review of sensors for anaerobic digestion finds that simple titration assays remain common in research and in industry, largely for their low cost, while offline sampling introduces delays and limits the temporal resolution. Microbial electrochemical sensors have reached an accuracy comparable to gas chromatography for the volatile acids, and electrochemical sensors remain constrained by their longevity, reproducibility and biofouling [12]. The supervisor of Chapter 11 acts on one titration a day. In the editor's view an online acid reading would shorten that sample period only if it carries its own check against the titration.

Supervisors trained on a digital twin. A study of a full-scale digester notes that such plants are often operated conservatively because of delayed process feedback, nonlinear stability constraints and limited observability. It trained a supervisory policy offline on a calibrated mechanistic digital twin, without experiments on the live plant, and found a modest potential gain of 2.7 percent in biogas production, subject to substantial uncertainty [13]. The comparison of Chapter 11 gives the same caution, since its predictive feed lost to the PI loop under a hidden inhibitor. In the editor's view a supervisor trained in simulation inherits the credibility of its model, which is gap 2 of Section 15.5. Figure 15.3 sets the developments of this section in three threads.

15.4 Engineering With AI Agents

An AI agent is a language model that works toward a goal in a loop. It plans, calls tools such as a compiler, a simulator or a browser, reads the results and continues [17]. In the editor's view, agents can now carry out much of the work that used to fill the hours of a control project. They derive a model from balances or fit one to plant data. They write the simulator and its tests, code the controller,

tune it against stated criteria, script commissioning checks, write the documentation and watch running loops for degradation. The apps and chapters of this series were produced that way under the editor's direction, and the account that follows this section reports what the agents got wrong in this volume and how it was caught.

In the editor's view, this changes where an engineer's value lies. When a model, a controller and a report can each be produced in minutes, producing them is no longer the scarce skill. The work moves to five things.

Quality control. An agent's output arrives as a claim: this model fits, this tuning settles in four minutes, this code implements the equation. The engineer checks claims against the plant or the simulator. A reported number is measured again on a fresh run at the stated seed. The code is read for what it clamps, what it initializes and what it reads, because a display that shows the controller's model in place of the plant looks correct until the two differ. A tuning is tried on a disturbance the agent did not test. The most useful checks are independent of the author, made by a second agent or a person who has not seen the first answer, and they are checks that can fail.

Supervision. The engineer sets the scope of each task and decides what an agent may touch. Reading a historian and proposing a tuning is one level of access. Writing a setpoint, downloading a configuration or changing an interlock is another, and it passes through the plant's management of change with a named approver. Supervision also fixes the order of work: a baseline before a change, one change at a time and a comparison that shows what moved.

Creativity. An agent answers the question it is given. Choosing the question remains the engineer's task: which loop is worth improving, whether a disturbance calls for feedforward, a cascade or a change to the process, which step test the operation can tolerate, and which comparison would show that a predictive controller earns its place over a retuned PID loop. The experiment that separates two explanations is designed by someone who can imagine both.

Process insight. A model contains what its author put in. The equipment does more. A valve sticks, the dead time of an analyzer grows when its sample flow drops, a thermowell lags, a heat exchanger fouls and an operator intervenes at the same hour each night. The engineer who knows the unit recognizes a result that is too good, asks what the model left out and knows which measurement to distrust first.

Accountability. Responsibility does not transfer to a tool. Functional safety standards assign the activities of the safety lifecycle to persons and organizations of stated competence, and they require an assessment that includes someone independent of the design team [18]. An agent may draft the hazard analysis or the test procedure. A person owns the decision to accept it, to start up and to keep running.

These five are the professional form of the habit introduced in Section 1.7: check before trusting, and know enough to check.

15.4.1 How This Volume Was Made

Each of the thirteen apps was built and tested by agents. An agent then wrote the paper on it, keeping verification notes that trace each number to a run. Seven of the thirteen referee reports call themselves self-reviews or an internal review by the drafting agent, and the other six write in the voice of a reviewer. None names a reviewer independent of the writing. Four carry later referee passes on predictive-control or estimator material, three of them in the insulin pump's paper. The defect logs hold 157 entries with a severity. Of these, 33 are wrong output or behavior, 86 documentation or

content, 24 cosmetic or display and 14 something else, such as a completion rule or an entry checked and found not to be a defect. Some round notes list fixes without a severity, so the counts are a floor. The editor set the scope of each round and what each agent could change. Four cases show what went wrong and what caught it. In each, a loop met variability or a limit in a way its text did not say.

Controllers that knew the individual. Four entries in four logs record a controller that read the run's hidden parameters or the plant's true state. The runner's predictive controller started every plan from the runner's hidden state and stepped the run's own parameters. With the readings held fixed, 24 of 50 parameters scaled by 1.25 moved its plan, and a hidden core temperature raised by one percent moved its predicted heart rate. In the rider's predictor 35 of 61 parameters and an unmeasured headwind moved the plan, and under the fingerstick it read the true lactate in place of the reading. The insulin pump's controller took its starting bias from the run's balances, and the antibody culture's estimators started from the broth itself. In a volume about variability between individuals, a controller that reads the individual's true parameters has no variability to handle. Each fix gave the controller a declared individual and its own estimate. The runner's predictive hour fell from 9.991 to 9.947 km, and two of its hours now break the lactate ceiling, one of them because the controller is no longer told that the room is warm. The probe that found them is the one the shared section describes: move a hidden value with the measurements held, and ask whether the controller moves.

A lead measured where it was not earned. The runner app's texts said that almost the whole lead of the predictive controller was earned in the last five minutes. The paper agent, measuring every number the app asserts, sampled both runs once a second and found 215 of a final 406 m already won by three minutes, because the regulator walks a five-minute reference ramp up from rest. The same texts said the terminal value was what made the predictive controller work. Deleting it changed the flat hour by one meter. Its product form carried a kink from the fuel store, so on the under-fueled hour the belt alternated between 6.4 and 11.9 km/h and covered 8.33 km, behind the regulator. A smooth rollout fixed the alternation, and the chapter now reports the terminal value as measured, within 31 m of none at all. In the same app a claim that the sustainable pace falls through the hour was true in direction and 0.5 % in size on the hour it was attached to.

Protective layers that leaked. The anesthesia app's signal-quality hold froze the reading through cautery but let the monitor's filter go on integrating the artifact. On release the controller received a reading 24.7 index units above the patient, and its propofol target stepped from 1.81 to 2.83 µg/mL just after the cautery it was meant to ignore. The predictive-control round found it because one comparison was dominated by the minute after the hold released. The antibody culture's predictive controller, as first built, ran its pump on the declared stock and not the label, and its plan was not offset-free, so a weaker lot took the glucose under 1 g/L for 22.4 h. A technical review found these and a third fault, and Table 8.8 reports the controller before and after. The cyclist's predictive controller never started the bicycle on level ground, since its predictor turned zero cadence into zero power. The circuit hid the defect, because its start rolls downhill.

Causes the runs did not support. The book round tested each chapter's causal sentences by run and corrected six in three chapters without adding a number. The anesthesia paper said the end of a case is shortened by anticipation rather than by tuning. Twice the depth gain cut the time below 45 from 12.95 to 4.35 min, about what the opioid cut gives, while the lowest index stayed at 35.5. Only anticipation raises the bottom of the dip, and the chapter now says so. The kidney circuit's paper said no tuning helps a loop that follows a drifting scale. A slower loop left the patient 244.8 mL behind at six hours instead of 304.3, and caught up with the ramp by twelve. The brewery paper blamed an undershoot partly on the glycol left in the jacket. Runs over the probe lag moved the minimum by 0.18 °C and runs over the glycol holdup by 0.02 °C.

The book round. The thirteen chapter records list 177 rows of numeric verification, each a stated value or group of values set beside its reproduction, and all thirteen report no drift. The writer of Chapter 1 read the apps' action flags and found five chapters that named a loop's action against the convention of Section 1.3. Its own first outer loop for P-1 compared the measured effect site with the setpoint, so it integrated the effect-site lag and made the population patient overshoot. Comparing with the model's prediction made the loop act on model error only, which is the lesson of Chapter 3. Chapter 1 was later corrected from three chapter passes, on the order of an override, the meaning of a lag and the notation. Seven of the thirteen chapter passes logged an app defect, most of them stale numbers in briefings. One was behavior: at the smallest glycol holdup its slider allows, the brewery's integration step is too long and the run breaks down once a valve opens. One chapter was finished in two runs after its first agent was cut off, and the second agent checked the first one's editorial pass before going on.

Every error above was caught by a check that could fail: a hidden value moved while the measurements were held, a lead decomposed over time, a term deleted to see what it was worth, a protective layer released on purpose, a causal sentence tested against the variable it names. None was caught by rereading the text that contained it. Four working rules follow. Give a controller only what the plant measures and the design declares. Decompose every claimed advantage before explaining it. Test each protective layer on the event it guards and on its release. Run every "no tuning helps" with two tunings.

15.5 Gaps in the Control of Living Systems

Five gaps belong to this volume's topic, and the gaps common to agent-era practice follow them in Section 15.5.1. In the editor's view none is closed. Each item gives the chapters' evidence, then the state of the literature with its sources. Why each gap is hard and what progress would look like are the editor's view. Figure 15.4 places the gaps by the chapters that met them and by the maturity of the method available, and the same numbers mark the plants in the picture that opens this chapter.

1. *Variability that no measurement reveals.* Exercise multiplies insulin sensitivity by 1.6 in Chapter 2 with no change in any measured quantity, and the hidden inhibitor of Chapter 11 misleads a forecast that has no state for it. A review of closed-loop insulin systems finds that excursions after meals remain a challenge because insulin under the skin is absorbed with a delay, that users usually announce exercise to the algorithm in advance, and that heart rate or accelerometry may detect activity sooner than the glucose monitor alone [15]. Full-scale digesters are often run conservatively, in part because of limited observability [13]. The gap is hard because an unmeasured parameter shows itself only in the harm it does. Practice needs estimators of an individual's unmeasured parameters that stay safe while they learn, and a way to bring signals such as activity or a second analyzer into them.

2. *Virtual populations with stated credibility.* Every chapter runs one declared individual: a patient with one covariate set and a slider in Chapter 3, one virtual subject in Chapter 2, a scripted lung in Chapter 4 and a culture with teaching constants in Chapter 8. A margin sized for the spread of individuals needs that spread, drawn from data, with a statement of the context of use in which the model may be trusted. The Open challenge box after this list takes the gap further.

3. *Margins for one-way actuators.* No standard method sizes the margin a one-way pump needs in front of a hard limit from the spread of individuals and the slow way back, and the chapters set it by run. In Chapter 2 the pump sat at zero for 0.84 h while the insulin already given went on acting. The anesthesia dip at the end of a case is raised only by feedforward from an event nobody instruments, and the fish farm's governor cannot stop the nitrite of meals already eaten. Insulin given under the skin acts late, because it must first be absorbed, with peaks up to 120 min after a bolus of regular insulin [14]. The gap is hard because the margin must be set before the

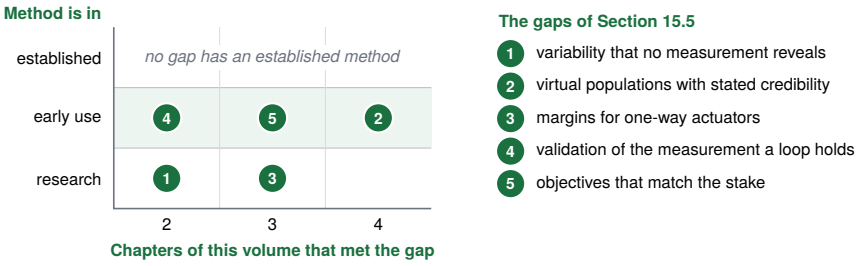


Figure 15.4 The five gaps of this section. Across is the number of chapters that each item names as evidence: two for gap 1 (Chapters 2 and 11), four for gap 2 (2, 3, 4 and 8), three for gap 3 (2, 3 and 12), two for gap 4 (3 and 5) and three for gap 5 (2, 9 and 11). Up is the maturity of the method available, the editor's reading of the sources cited in each item and in the Open challenge box. Asymmetric targets are agreed for glucose alone, so gap 5 sits in early use. Notice that the two gaps about what a loop cannot see or cannot take back, gaps 1 and 3, are still research.

individual is known, and it is paid for on every ordinary day. A method that sizes the margin from the spread of individuals and the slow way back, and states the risk it leaves, would replace setting it by run.

4. *Validation of the measurement a loop holds.* The kidney circuit's check missed every drift under its tolerance, and the monitor that is wrong during cautery reads like a light patient. Weighing is the most common way to balance fluid in continuous kidney therapy and the most reliable over long treatments, and its disadvantages include limited scale capacity, user errors and disturbances of the operating environment [19]. Muscle activity and electrical interference can pass into a depth index without being displayed as artifact, and raise its value spuriously [20]. The gap is hard because a fault that drifts slowly looks like the plant moving, and a tighter check raises more false alarms. Plausibility checks with stated detection limits, and redundant measurements designed into the equipment, are needed wherever a loop can hold a measurement while the plant moves.

5. *Objectives that match the stake.* A symmetric error, a vessel's yield and a day's revenue each chose the wrong controller in this volume. An international consensus on glucose monitoring already scores the two sides of the range apart, with targets of more than 70 % of readings in range, fewer than 25 % above 180 mg/dL and fewer than 4 % below 70 mg/dL, and it names raising the time in range while reducing the time below it as the primary goal [21]. The gap is hard because the price of each side is set outside control engineering, and most of these plants have no agreed price. Agreed asymmetric measures, campaign-level scores and a monitored margin to a fold, reported beside every result, would make a comparison mean what a clinician, a plant manager or a public-health officer needs it to mean.

⚠ OPEN CHALLENGE A Population, Not One Individual

Four chapters name gap 2, and every chapter runs one declared individual. A simulator of 300 virtual subjects with type 1 diabetes, built from real individuals' data to span the observed variability, was accepted by the Food and Drug Administration as a substitute for animal trials in the preclinical testing of closed-loop control [22]. The commercial models used to dose anesthetic drugs were built on selected populations that often left out children, obese and elderly patients [23]. In the editor's view the gap is hard because the spread that sizes a margin lies in the individuals that studies leave out. Progress would be a population drawn from pooled data and stated with the context in which it may be trusted.

15.5.1 Gaps Common to Agent-Era Practice

The final chapter of each volume lists gaps specific to its topic. Seven more are common to every topic in the series. In the editor's view none of them is closed.

1. *Verifying agent-written control code and models before they reach a plant.* Tests show that a controller behaves on the cases tested. For controllers learned from data, guarantees of stability and constraint satisfaction rest on assumptions about the model and its uncertainty, and open problems remain [24]. In the editor's view, code written by an agent has no comparable body of method yet. Practice needs a routine path from an agent's draft to code a plant can trust: independent derivation, tests of what the code reads and clamps, and trials with hardware in the loop.

2. *Trustworthy models from sparse or poor plant data.* Plant records are short, taken in closed loop, noisy and marked by unmeasured disturbances. A fitting routine returns a model from any record. What is missing is a dependable statement, delivered with the model, of where it can be trusted, and a method that says so when the data cannot identify a parameter.

3. *Safety cases and standards for agents in commissioning and operation.* Functional safety standards define a lifecycle of specified requirements, verification and assigned human roles [18]. In the editor's view, how an agent that writes configuration, runs commissioning checks or advises an operator is qualified, limited and audited under them is not yet settled.

4. *Security of agent access to control systems.* An agent with credentials to a control system is a new path into it. Industrial security standards state the requirements a control system must meet at each security level [25]. An agent that reads documents, messages or web pages can also be steered by text placed there. In the editor's view, least privilege, read-only access by default and human approval of every write are a starting point and not a solution.

5. *Benchmarks and reproducibility.* An agent asked twice may deliver two different controllers, and a result reported without its plant, seed and scoring cannot be checked. Control needs shared benchmark plants with fixed disturbances and published measures, on which agent-built and conventional designs are compared like for like. The apps of this series, with a default seed and an exported record of every run, are a start.

6. *Transfer from simulation to hardware.* A controller developed on a simulator inherits what the simulator omits. The robotics literature documents the gap between simulated and real performance and the methods used to narrow it, such as randomizing the simulated plant and adapting on the real one [26]. Process plants add slow dynamics, rare upsets and little freedom to experiment, so each transfer still needs a staged plan from the simulator to a bench rig to the unit.

7. *Keeping engineers' judgment sharp when agents do the routine work.* Automation removes the routine work through which skill is built and leaves the person the hard cases, an irony described for process operators in 1983 [27]. The same now applies to engineers. An engineer who has never tuned a loop by hand or closed a balance on paper is poorly placed to review an agent that

has. Deliberate practice, of the kind the exercises in this book provide, has to be designed into education and into work.

KEY POINTS

- A loop coped with the differences between individuals when it could measure the effect it was meant to hold.
- A loop that holds its measurement need not hold the plant.
- One-way actuators were held by margin and by acting early.
- The objective decided what counted as held.
- Prediction paid where it had a deadline and an anchored estimate.

References

[1] J. L. Sherr, L. Heinemann, G. A. Fleming, R. M. Bergenstal, D. Bruttomesso, H. Hanaire, R. W. Holl, J. R. Petrie, A. L. Peters, and M. Evans, "Automated insulin delivery: Benefits, challenges, and recommendations. A consensus report of the joint diabetes technology working group of the European association for the study of diabetes and the American diabetes association," *Diabetes Care*, vol. 45, no. 12, pp. 3058–3074, 2022.

[2] L. Pasin, P. Nardelli, M. Pintaudi, M. Greco, M. Zambon, L. Cabrini, and A. Zangrillo, "Closed-loop delivery systems versus manually controlled administration of total IV anesthesia: A meta-analysis of randomized clinical trials," *Anesthesia & Analgesia*, vol. 124, no. 2, pp. 456–464, 2017.

[3] M. Ghita, M. Neckebroek, C. Muresan, and D. Copot, "Closed-loop control of anesthesia: Survey on actual trends, challenges and perspectives," *IEEE Access*, vol. 8, pp. 206 264–206 279, 2020.

[4] P. von Platen, A. Pomprapa, B. Lachmann, and S. Leonhardt, "The dawn of physiological closed-loop ventilation—a review," *Critical Care*, vol. 24, no. 1, p. 121, 2020.

[5] M. Viceconti, F. Pappalardo, B. Rodriguez, M. Horner, J. Bischoff, and F. Musuamba Tshinanu, "In silico trials: Verification, validation and uncertainty quantification of predictive models used in the regulatory evaluation of biomedical products," *Methods*, vol. 185, pp. 120–127, 2021.

[6] C. Ronco, "Continuous renal replacement therapy: Forty-year anniversary," *The International Journal of Artificial Organs*, vol. 40, no. 6, pp. 257–264, 2017.

[7] H. Narayanan, M. F. Luna, M. von Stosch, M. N. Cruz Bournazou, G. Polotti, M. Morbidelli, A. Butté, and M. Sokolov, "Bioprocessing in the digital age: The role of process models," *Biotechnology Journal*, vol. 15, no. 1, p. 1900172, 2020.

[8] K. A. Esmonde-White, M. Cuellar, C. Uerpmann, B. Lenain, and I. R. Lewis, "Raman spectroscopy as a process analytical technology for pharmaceutical manufacturing and bioprocessing," *Analytical and Bioanalytical Chemistry*, vol. 409, no. 3, pp. 637–649, 2017.

[9] S. Hemming, F. de Zwart, A. Elings, A. Petropoulou, and I. Righini, "Cherry tomato production in intelligent greenhouses—sensors and AI for control of climate, irrigation, crop yield, and quality," *Sensors*, vol. 20, no. 22, p. 6430, 2020.

[10] K. M. Gostic, L. McGough, E. B. Baskerville, S. Abbott, K. Joshi, C. Tedijanto, R. Kahn, R. Niehus, J. A. Hay, P. M. De Salazar, J. Hellewell, S. Meakin, J. D. Munday, N. I. Bosse, K. Sherrat, R. N. Thompson, L. F. White, J. S. Huisman, J. Scire, S. Bonhoeffer, T. Stadler, J. Wallinga, S. Funk, M. Lipsitch, and S. Cobey, "Practical considerations for measuring the effective reproductive number, R_t," *PLOS Computational Biology*, vol. 16, no. 12, p. e1008409, 2020.

[11] H. Yu, B. Gao, and Y. Li, "Depth of anesthesia monitoring: Current evidence, clinical impact, and future directions," *Journal of International Medical Research*, vol. 54, no. 6, p. 03000605261457781, 2026.

[12] M. Marycz, I. Turowska, S. Glazik, and P. Jasiński, "Artificial intelligence in anaerobic digestion: A review of sensors, modeling approaches, and optimization strategies," *Sensors*, vol. 25, no. 22, p. 6961, 2025.

[13] A. I. Yunus, S. Jalla, J. F. Bozeman, and Y. Chen, "Reinforcement-learning-enabled supervisory control of a full-scale anaerobic digestion facility using a calibrated digital twin," *Environmental Science & Technology*, vol. 60, no. 39, pp. 28 064–28 078, 2026.

[14] C. Cobelli, E. Renard, and B. Kovatchev, "Artificial pancreas: Past, present, future," *Diabetes*, vol. 60, no. 11, pp. 2672–2682, 2011.

[15] C. K. Boughton and R. Hovorka, "New closed-loop insulin systems," *Diabetologia*, vol. 64, no. 5, pp. 1007–1015, 2021.

[16] S. Hemming, F. de Zwart, A. Elings, I. Righini, and A. Petropoulou, "Remote control of greenhouse vegetable production with artificial intelligence—greenhouse climate, irrigation, and crop production," *Sensors*, vol. 19, no. 8, p. 1807, 2019.

[17] L. Wang, C. Ma, X. Feng, Z. Zhang, H. Yang, J. Zhang, Z. Chen, J. Tang, X. Chen, Y. Lin, W. X. Zhao, Z. Wei, and J. Wen, "A survey on large language model based autonomous agents," *Frontiers of Computer Science*, vol. 18, no. 6, p. 186345, 2024.

[18] International Electrotechnical Commission, "Functional safety – Safety instrumented systems for the process industry sector – Part 1: Framework, definitions, system, hardware and application programming requirements," IEC 61511-1:2016, Edition 2.0, 2016.

[19] G. Villa, M. Neri, R. Bellomo, J. Cerda, A. R. De Gaudio, S. De Rosa, F. Garzotto, P. M. Honore, J. Kellum, A. Lorenzin, D. Payen, Z. Ricci, S. Samoni, J.-L. Vincent, J. Wendon, M. Zaccaria, and C. Ronco, "Nomenclature for renal replacement therapy and blood purification techniques in critically ill patients: practical applications," *Critical Care*, vol. 20, p. 283, 2016.

[20] A. A. Dahaba, "Different conditions that could result in the bispectral index indicating an incorrect hypnotic state," *Anesthesia & Analgesia*, vol. 101, no. 3, pp. 765–773, 2005.

[21] T. Battelino, T. Danne, R. M. Bergenstal, S. A. Amiel, R. Beck, T. Biester, E. Bosi, B. A. Buckingham, W. T. Cefalu, K. L. Close, C. Cobelli, E. Dassau, J. H. DeVries, K. C. Donaghue, K. Dovc, F. J. Doyle, III, S. Garg, G. Grunberger, S. Heller, L. Heinemann, I. B. Hirsch, R. Hovorka, W. Jia, O. Kordonouri, B. Kovatchev, A. Kowalski, L. Laffel, B. Levine, A. Mayorov, C. Mathieu, H. R. Murphy, R. Nimri, K. Nørgaard, C. G. Parkin, E. Renard, D. Rodbard, B. Saboo, D. Schatz, K. Stoner, T. Urakami, S. A. Weinzimer, and M. Phillip, "Clinical targets for continuous glucose monitoring data interpretation: Recommendations from the international consensus on time in range," *Diabetes Care*, vol. 42, no. 8, pp. 1593–1603, 2019.

[22] B. P. Kovatchev, M. Breton, C. Dalla Man, and C. Cobelli, "In silico preclinical trials: A proof of concept in closed-loop control of type 1 diabetes," *Journal of Diabetes Science and Technology*, vol. 3, no. 1, pp. 44–55, 2009.

[23] O. Vandemoortele, L. N. Hannivoort, F. Vanhoorebeeck, M. M. R. F. Struys, and H. E. M. Vereecke, "General purpose pharmacokinetic-pharmacodynamic models for target-controlled infusion of anaesthetic drugs: A narrative review," *Journal of Clinical Medicine*, vol. 11, no. 9, p. 2487, 2022.

[24] L. Brunke, M. Greeff, A. W. Hall, Z. Yuan, S. Zhou, J. Panerati, and A. P. Schoellig, "Safe learning in robotics: From learning-based control to safe reinforcement learning," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5, pp. 411–444, 2022.

[25] International Electrotechnical Commission, "Industrial communication networks – Network and system security – Part 3-3: System security requirements and security levels," IEC 62443-3-3:2013, Edition 1.0, 2013.

[26] W. Zhao, J. P. Queralta, and T. Westerlund, "Sim-to-real transfer in deep reinforcement learning for robotics: A survey," in *Proc. 2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2020, pp. 737–744.

[27] L. Bainbridge, "Ironies of automation," *Automatica*, vol. 19, no. 6, pp. 775–779, 1983.

15.6 Exercises

15.1 Discussion · 8 min

An agent proposes to remove the depth loop of Chapter 3 and give the target-controlled pump a richer population model with more covariates, "so that the monitor, which fails during cautery, is no longer needed". Using Table 3.7 and scenario 9 of the chapter, say what the proposal would gain and what it would lose.

15.2 Discussion · 8 min

The balance loop of Chapter 5 keeps its measured error within 28.4 mL while the patient falls 304.3 mL behind in six hours. Explain why retuning the loop cannot fix this, and discuss how you would choose the tolerance of the scale check of Table 5.4.

15.3 Discussion · 7 min

In Chapter 2 exercise takes glucose from 74.9 to 47.8 mg/dL before the controller next executes, and the pump then sits at zero for 0.84 h. A colleague proposes a higher gain "so the loop reacts to lows faster". Discuss the proposal, and say what does protect a one-way loop from a fast fall.

15.4 Discussion · 9 min

An agent reports that its predictive controller covers 9.947 km in the runner's hour of Chapter 6 against 9.584 km for the selector, "because it plans the whole hour". Using Tables 6.8 to 6.10 and Section 15.4.1, list the checks you would run before accepting the claim and its explanation.

15.5 Discussion · 8 min

Three chapters chose a different controller depending on the score: Chapter 2 with a symmetric IAE, Chapter 9 with the fold of one lot, and Chapter 13 with the cost of the week. Pick one of them, state what the score hid, and propose a score that a person responsible for that plant would accept.

15.6 Discussion · 8 min

On the epidemic challenge of Chapter 14, the nominal predictive controller with a 20 % margin passes all seven records for less than the scenario tree, and the chapter still prefers the tree. Using Table 14.5 and the gaps of Section 15.5, argue for one of the two as the policy to recommend, and say what evidence would change your choice.

Answers to Selected Exercises

Short answers are given for the odd-numbered exercises. Numeric answers from an app are for its default seed and settings unless the exercise says otherwise.

Chapter 1. Introduction

1.1 (a) 20 min, 3 mg/min and 60 mg. (b) About 60 min, three time constants.

1.3 (c). The steady concentration is the target divided by the clearance ratio, $2/1.6 = 1.25\,\mathrm{mg/L}$.

1.5 (a) 2.30, 0.52 g/L and 1.19 g/L. (b) Day 3, the fourth day, with 6.29 g/L.

1.7 (a) 16.5 % low. (b) 0.784 of the error.

1.9 (a) 6.7 min and 27.5 min. (b) 44.4 s, 700 W and 367 W.

Chapter 2. Artificial Pancreas With One-Sided Insulin

2.1 (c). More insulin lowers glucose, so the process gain is negative and the output must rise as the measurement rises.

2.3 (a). The symmetric score barely moves while the low side moves a great deal.

2.5 (a) $\tau_p = 1.18\,\mathrm{h}$, $\theta_p = 0.16\,\mathrm{h}$. (b) 9.6 against 11.1 min and 70.8 against 81.3 min. The glucose balance takes 1.0 min. (c) 0.048 and 0.20, both below the shipped 0.3.

2.7 (a) 47.77 mg/dL, a secant gain of −16.1 mg/dL per µU/min, 60 % of the tangent gain. (b) About 1.17 h, shorter than the 1.34 h of the one-unit step.

2.9 (a) 50.9, 98.6, 75.0, 0 h. (b) 99.0, 130.9, 64.9, 0.84 h. (c) 83.7, 116.8, 66.2, 0.67 h. Told of the meal the MPC wins, and not told it loses to the PID.

2.11 (a) 17.0 h against 14.2 h, 2.9 h a day more. (b) 1.34 h. Every one of those samples lies between 64 and 70 mg/dL.

Chapter 3. Closed-Loop Anesthesia by Model Inversion

3.1 (b). The inner loop reads the pump's integration of the population model, which never sees the patient's sensitivity ρ.

3.3 (d). The effect site lags the plasma by 2.19 min, so the bottom of the dip is set before the closing.

3.5 (a) $U = 0.900$ and 3.74 µg/mL. (b) $U_{50} = 0.65$, 35 % less. (c) 66.7.

3.7 (a) 67.4 and 1.708 µg/mL. (b) 59.6. (c) 4.07 µg/mL, still short of the 4.27 the patient needs.

3.9 (a) Fail, 10.1, 99.7 min, 0.0 min. (b) Pass, 3.3, 4.4 min, 1.2 min. (c) Fail on the two missing switches, 4.4, 4.4 min, 11.3 min.

3.11 (a) The pump's population model, the depth monitor and the patient's sensitivity ρ. (b) It drives the composition the wrong way for ten minutes. Hold the last good value while the signal is flagged.

Chapter 4. Mechanical Ventilator Breath Cycle Control

4.1 (b). Volume control imposes the flow, so the pressure pays: PEEP $+ V_T/C$ goes from 16.4 to 31.3 $\mathrm{cmH_2O}$.

4.3 (d). The effort is an unmodeled pressure in the same phase of every breath it triggers, so the error is systematic: about −22 % on R and +11 % on C.

4.5 (a) 1.00 s, 1.50 s, 2.41 $\mathrm{cmH_2O}$. (b) 0.90 $\mathrm{cmH_2O}$. (c) About 7 and 6 % above, because each expiration starts from the peak pressure and the circuit takes a moment to come down to PEEP.

4.7 (a) 31.3 $\mathrm{cmH_2O}$, highest 31.5 $\mathrm{cmH_2O}$. (b) 395 to 396 mL at 29.9 to 30.0 $\mathrm{cmH_2O}$, highest still 31.5 $\mathrm{cmH_2O}$, because the first breaths after the step pass before the layer has learned the stiffer lung.

4.9 (a) 540 of 1975 breaths in band (27.3 %), auto-PEEP 6.93 cmH_2O, 1211 auto-triggers, CO_2 111.0 mmHg: the volume, auto-PEEP and trigger targets fail. (b) 1157 of 1160 (99.7 %), 1.95 cmH_2O, one missed effort and two auto-triggers, CO_2 29.3 mmHg: the target is met.

4.11 (a) 4.32 L/min and 39.95 mmHg. (b) 4.44 mmHg a minute. (c) A real patient would breathe harder as the CO_2 rises, and would tire. The scripted effort ignores both.

Chapter 5. Kidney Replacement Fluid Balance Control

5.1 (a). The loop follows the drifting reading by moving the patient, so its own error stays small and the true error grows.

5.3 (c). At the limit the gain of the dose loop changes sign, and only a structural cap on the dialysate removes the sign change.

5.5 (a) $-1 \pm \mathrm{j}$ per hour, $\zeta = 0.707$, $\omega_n = 1.41$ rad/h. (b) 48.4 mL at 0.785 h. (c) A double pole at -2 per hour and a peak of 27.6 mL at 0.5 h.

5.7 (a) About −304 mL true, with the measured error inside about ±30 mL. (b) About −13 mL, with the effluent scale flagged. (c) About −37 mL: the substitute pump count cannot follow the tubing wear.

5.9 (a) −893 mL, 81 %, 284 mmHg and 21.1 %: the band, the dose and the TMP limit fail. (b) −51 mL, 90 %, 206 mmHg and 22.5 %, a pass. (c) Yes, with −37 mL and 88 %.

5.11 About 93 mL, close to the 88.1 mL reported. Catch-up repays the backlog, and changing bags together halves the pauses.

Chapter 6. Treadmill Pacing by Constraint Selection

6.1 (d). The tracked loop reproduces u exactly and wins the next comparison by an arbitrarily small margin.

6.3 (b). Ψ_G is flat above 54.1 g of store and sloped below it, and $T - t_N$ multiplies the slope.

6.5 (a) 1059.8 W, 101.0 W, 586 s. (b) 0.0118, 579 s, within 0.4 % of the measured 577 s. (c) 88 m, against the 90 m the chapter reports at 141 s.

6.7 (a) 29 mmol s/L, peak 4.52 mmol/L. (b) Zero excess, peak 3.50 mmol/L. (c) $2.787 \times 4.15/6.49 = 1.78$ m/s, which the file shows.

6.9 (a) 9.584 km, 189 m. (b) 9.947 km, 36 m. (c) 210 m of 363 m, 58 %. (d) The startup share is the regulator's reference ramp, and the finish is the deadline a regulator cannot represent.

6.11 (a) 620 W and 909 W. (b) Only with the fan restored. Restoring it bought 1.5 °C of peak core temperature and 320 m in the chapter's runs.

Chapter 7. Cycling Exertion With a Depleting Actuator

7.1 (c). Above E_c the exertion buys a share of $P_{top} - CP$, and P_{top} falls with the W' left.

7.3 (a). A watt buys 0.0187 m/s on the level and 0.0167 m/s on a 6 % climb, and lactate averages the effort over about four minutes.

7.5 (a) 112.2 W, 30.8 W, 146.7 W, the 146.67 W that 50 % exertion buys. (b) 53.4 and 59.9 W per m/s. (c) 0.0187 and 0.0167 m/s per watt, 11 % less on the climb.

7.7 (a) 0.0, 12.3, 14.9 kJ; 87 W, 218 W. (b) $16000/(0.5 \times 131) = 244$ s. (c) $16(1 - \mathrm{e}^{-360/244}) = 12.3$ kJ and $16(1 - \mathrm{e}^{-660/244}) = 14.9$ kJ, as read.

7.9 (a) 27.93 km, compliant, short of 28.8 km. (b) 28.90 km, compliant, passes. (c) 28.94 km, not compliant, 14 mmol s/L of lactate excess. (d) 29.05 km, compliant, passes.

7.11 (a) Tailwind: $a = \sqrt{0.167/8} = 0.14$, factor 0.53. Headwind: $a = \sqrt{6.83/8} = 0.92$, factor 0.96, 1.8 times as much. (b) 39.2 against 38.2 °C. A trainer has $v = 0$, so only a fan gives airflow.

Chapter 8. Feeding an Antibody Culture on Daily Samples

8.1 (c). Each bolus can at best start the interval at the top of the band, and the cells then take a day's uptake before the next bolus arrives.

8.3 (a). With the bubbles nearly saturated, only more gas raises the bound $60\,000\,\dot{n}_g(y_{eq} - y_C)$.

8.5 (a) $2 \times 11.1 \approx 22$ mOsm/kg. (b) 27.6 mol. (c) About 76.7 mol. The sodium stays, because nothing in the model consumes it.

8.7 (a) 34.8 h outside the band, 4.14 g/L, 435 mOsm/kg. (b) 0.7 h, 2.73 g/L, 511 mOsm/kg. (c) The high target keeps the glucose above the switch threshold, so the cells keep making lactate and the base drives the osmolality up.

8.9 (a) 5.03 g/L, 80.9 %, 27.6 mol. (b) Day 3: 4.33 g/L, 87.4 %, 3.1 mol. Day 8: 3.57 g/L, 68.2 %, 45.6 mol, harvested before day 14. (c) Early cooling leaves too few cells, late cooling leaves a warm culture making lactate that the base turns into osmolality.

8.11 More gas strips more carbon dioxide, so less bicarbonate has to be made with base to hold the pH, and less sodium enters the broth. A higher gas flow brings foam, the antifoam it needs and shear at the bursting bubbles. Oxygen first costs pure oxygen and leaves the carbon dioxide high.

Chapter 9. Insulin Refolding and Recovery Scheduling

9.1 (b). The closed form of Equation (9.4) depends on the charge only through α, and a smaller α gives a larger folded share.

9.3 (d). The refolding and conversion vessels set the lot rate, and the campaign horizon does not move when the harvest does.

9.5 (a) $k_f = 4.61 \times 10^{-5} \mathrm{e}^{0.693} = 9.22 \times 10^{-5}$ and $k_a = 2.21 \times 10^{-5} \mathrm{e}^{1.351} = 8.54 \times 10^{-5}$ in the units above. (b) $\alpha = 1.39$ and $\ln(2.39)/1.39 = 62.7\,\%$, against 75.3 % at 15 °C. (c) Both constants rise, and the one that destroys protein rises faster, so the hours saved are paid for in aggregate.

9.7 (a) 67.7 % folded, 30.0 % aggregated, 18.8 h. (b) 84.7 % folded, 13.1 % aggregated, 39.0 h. (c) The bolus, by 112.7 g to nothing: the fed charge keeps 17 points more of its own lot and holds the vessel twice as long, so 665 g of harvest behind it passes its hold limit at 52 h and the campaign isolates nothing.

9.9 (a) 0.00 g/h: both lots are isolated but rejected on host-cell protein, at 1654 and 1685 ppm against a limit of 1000 ppm, because the shipped recipe washes only 1.5 volumes. (b) 5.56 g/h, 401 g in three lots, target met. (c) 5.47 g/h: the scheduler is ahead by 0.09 g/h at this seed, a margin smaller than the 0.12 g/h that moving the ceiling from 0.90 to 0.75 g/L is worth, while both configurations hold the vessel far inside the temperature and pH tolerances.

9.11 (a) Those three cases all end with recipe lots about 0.76 of a point inside the purity limit, so their verdicts turn on a declared purification parameter with 19 % of margin; a plant would measure the selectivity of its own capture and polish steps, that is, how much related substance and host-cell protein each step carries into the bulk, before believing the ranking. (b) From the product's specification, the pharmacopeia and regulatory guidance such as ICH Q6B; a simulated pass is not a batch release.

Chapter 10. Lager Fermenter With On/Off Glycol Valves

10.1 (a). The holdup is refilled with cold glycol at every pulse and keeps removing heat while the valve is shut, so the characteristic is concave.

10.3 (c). The textbook yeast form less VDK and reduce it faster than this strain, so the uncorrected estimate reaches the limit too soon.

10.5 (a) 1.23 GJ/K and 18.45 GJ. (b) 102.5 kW. (c) About 5.9 °C.

10.7 (a) 10.66 to 11.27 °C and 10.97 to 11.04 °C. (b) 29 switches (6 to 35) and 802 (104 to 906). (c) $0.61/0.07 = 8.7$, nearly nine times. The probe reports the cooling ten minutes late.

10.9 (a) Day 10.5, 0.087 mg/L, 0.52 °C, never over capacity, met. (b) Day 10.5, 0.083 mg/L, 0.52 °C, 39.4 kW over capacity for 1.2 h, not met. Full openings meet the neighbor's crash.

10.11 The VDK is one lumped state, so diacetyl formed after the crash from remaining precursor is not modeled. The estimator differs from the plant only in its constants, where a real one would differ in structure.

Chapter 11. Feeding a Digester on a Daily Titration

11.1 (d). The acids displace bicarbonate, and the carbon dioxide released leaves with the biogas, which reaches 2.97 times its design flow of carbon dioxide on day 3.08.

11.3 (b). No methanogen population can grow as fast as 300 m^3/d washes it out, so the acids climb whatever the pH.

11.5 (a) 302 kW. (b) $G = 14.09\,\mathrm{kW/K}$, $\tau = 13.73\,\mathrm{d}$, 13.58 °C, 27.4 °C. (c) About 140 m^3/d, below the 250 m^3/d the tank is fed.

11.7 (a) 9504 Nm^3/d on day 2.63. (b) Day 3.58, 0.55, 0.95 days after the peak. (c) Row 2.04, after the trip on day 2.01; soured on day 4.5. The carbon dioxide driven out by the acids took the holder's gas below the upgrader's 45 % methane floor.

11.9 (a) 27.4 °C on day 8.42; report over 0.4 on day 7.58; soured on day 10.6. (b) Not soured, four reports over 0.4. (c) The fold is below 160 m^3/d only while the tank is cold, and the acids need days to climb through the basin.

11.11 (a) \$1.81 per Nm^3, 94 % from the credits. (b) 103 t and about \$41,200, close to 12 times the supervisor's 40-day gain.

Chapter 12. Recirculating Aquaculture Control

12.1 (d). Both populations were grown for a third of the load, and the slower nitrite oxidizers, with less reserve, must grow on the extra nitrite before they can remove it.

12.3 (b). The model's respiration coefficient is 165 against the cohort's hidden 150, so the measured oxygen divides into fewer kilograms.

12.5 (a) Growth $R_{sat} e_{max} (\rho - c\rho^2 - k)$ over feed ρR_{sat}. (b) $\rho^* = \sqrt{k/c} = 0.569$. (c) $e_{max} = 1.508$, lowest conversion 1.0068, 8.47 % better than 1.1.

12.7 (a) 57.0 h, 3.07 mg N/L, 873 fish. (b) 20.7 h, 1.34 mg N/L, 1 fish; FCR 3.29 against 1.05. (c) The governor sees the nitrite only in the four-hourly analysis, after the meals that make it.

12.9 (a) 2.58 t, 69.8 h, 1.28, 1,359 killed, \$1,326, not met. (b) 2.54 t, 6.6 h, 1.08, none killed, \$2,148, met. (c) The shipped full table breaks the nitrite limit after the upset and kills fish, leaving fewer, larger ones on more feed. The governor and the economic ration cut the feed while the film regrows.

12.11 (a) $0.448 - 0.154 = 0.294\,\mathrm{d^{-1}}$, a factor of 1.34 a day. (b) $\ln 3 / 0.294 = 3.73\,\mathrm{d}$ at the very least. A ramp of 25 % a day stays below the bound, and the economic ration in case 3 rose only 4.8 to 12.6 % a day.

Chapter 13. Greenhouse Climate Under a Humidity Limit

13.1 (c). The growth rate G of Equation (13.10) depends on $\bar{T}_{24}$, so the heating setpoint is a budget for the day.

13.3 (a). Equation (13.12) replaces the integral by a lag of the clipped measured pipe temperature: external reset feedback.

13.5 (a) 2.5 $m^3/(m^2\,h)$, 120 min. (b) 28.3 $m^3/(m^2\,h)$, 10.6 min. (c) 4.5 min and 800 min.

13.7 (a) Fixed: 19.02 °C, 1040 m^3. Planner: 18.91 °C, 865 m^3. (b) Tuesday 1547 m^3 against 1501 m^3, about 3 %; Monday 17 %. (c) The planner saves gas by banking free sun, and the dull day has little.

13.9 (a) Six days off the recipe (worst 1.00 K), 775 humid dark minutes against 150 and 578 ppm against 600, at \$4292. (b) Yes: \$5594, no day off (worst 0.33 K), 78 minutes, 611 ppm. (c) \$196 less (\$5398), but 192 humid dark minutes fail the humidity target.

13.11 The price is the same in every hour of the day, so there is little to move between hours, and the daily mean must still be paid for. A heat buffer and a combined heat and power unit.

Chapter 14. Epidemic Response on Week-Old Counts

14.1 (b). A report counts an onset seven days after it happens, and the onsets still to come from everyone already exposed arrive over the latent period, with the 7-day average on top.

14.3 (d). The dial removes transmission, so the census falls as the dial rises, and a census above the target must raise the dial.

14.5 (a) About $0.3005N$, some 300,500 people. (b) $z = 0.9405$. (c) $s = 0.0595$, far below $1/R_0 = 0.333$, because the epidemic is at its largest when s passes $1/R_0$.

14.7 (a) Day 38. (b) 1,435 beds on day 62, and 1,794 reports a day on day 59. (c) 1,610 beds, over the 1,500-bed capacity for 15 days. One day of waiting costs about 175 beds, and the reports peak on day 59 either way.

14.9 (a) 1,542 beds, 33 days over the capacity, \$2,440 million. (b) 1,369 beds, \$2,855 million. (c) 1,485 beds, \$2,599 million. The 10 % margin holds the beds for \$256 million less than the tree. The tree's margin comes from a declared spread of futures, not from a number set by hand.

14.11 (a) \$32 million a day and 149 days on. (b) 0.588. (c) \$750 million against \$1,500 million: the same dial-days cost half as much spread over twice the time.

Chapter 15. Summary and Outlook

15.1 A good answer reads the sweep: under model inversion the settled index spans 52.1 to 70.4 as sensitivity moves from 0.70 to 2.0, and with the depth loop it spans 50.6 to 51.7 because the loop moves the commanded concentration from 1.90 to 5.03 µg/mL. A better model narrows the first range only for the variation its covariates explain. Scenario 9 shows what no model sees: a fifth of the central volume lost, the pump's model at its target and the patient's index at 43.2 with nothing displayed. The answer keeps the loop, keeps the signal-quality hold for cautery, and accepts a better model as a better starting target for the loop.

15.3 A good answer notes that the pump cannot remove insulin, so the loop's only answer to a low is to stop, which it already does at the next execution. A higher gain moves the pump harder upward after meals and so raises the insulin on board that causes the next low. The fall is faster than the control period and is caused by an unmeasured change in sensitivity. What protects the loop is margin and anticipation: a setpoint and tuning sized for the spread of sensitivity and the slow way down, suspension on a predicted low, a bound on insulin on board, and an announcement of exercise, as Section 1.4.6 describes.

15.5 A good answer for the pump says that a symmetric IAE ranks the exercise day (118.5) above the large meal (210.3) although the first fell to 47.8 mg/dL, and proposes time in range with separate low thresholds or Equation (1.10) with a weight above 7.9. For refolding, the 0.35 g/L ceiling folds the first lot 79.6 % and the campaign fails at 4.34 g/h, so the score is accepted insulin per campaign hour. For the glasshouse, the planner alone gives the cheapest week, \$3780, and fails the humidity limit, so cost counts only among weeks that meet every limit. Each proposal names a limit that is scored apart from the average.

Glossary

Each term is given as this volume uses it, in one plain sentence, followed by the chapter where it is first used in earnest. The two indexes give every page.

acute respiratory distress syndrome (ARDS) A lung injury in which the lungs stiffen and fill with fluid, so a ventilator must breathe with small volumes at limited pressure. *Chapter 4.*

alkalinity The capacity of a liquid to neutralize acid, and in a digester it is the buffer that the microbes' acids use up. *Chapter 11.*

anaerobic digester A closed tank in which microbes without oxygen turn manure and food waste into biogas. *Chapter 11.*

anti-windup Any method that stops a controller's integral from growing while its output is held at a limit. *Chapter 1.*

artificial pancreas A system that measures glucose and sets an insulin pump automatically, closing the loop a person with diabetes otherwise closes by hand. *Chapter 2.*

ascertainment The share of all infections that are found by testing and reported. *Chapter 14.*

asymmetric cost A score that weights errors on one side of a target more heavily than on the other, such as low glucose over high. *Chapter 1.*

back-calculation A way to stop windup: when the output is limited, the integral is pulled toward the output actually applied. *Chapter 2.*

bias update A correction that shifts a model's estimate by part of the difference between a laboratory result and the model at the time of sampling. *Chapter 1.*

biofilter A bed of bacteria on a support that turns the ammonia excreted by fish into nitrite and then nitrate. *Chapter 12.*

biogas The mixture of methane and carbon dioxide that a digester releases. *Chapter 11.*

bispectral index A number from 0 to 100 computed from the brain's electrical signal and read as the depth of anesthesia. *Chapter 3.*

blood lactate A product of muscle metabolism in the blood that rises steeply once exercise passes a sustainable intensity. *Chapter 6.*

bolus A dose given all at once, as opposed to a steady infusion. *Chapter 1.*

cascade control Two nested loops in which the outer controller sets the target of the inner one. *Chapter 3.*

Chinese hamster ovary cells The cultured mammalian cells that secrete the antibody in the fed-batch culture of the antibody chapter. *Chapter 8.*

clearance The volume of blood or water cleared of a drug or solute per unit time. *Chapter 1.*

compartment model A model that treats each part of a body or plant as one well-mixed volume exchanging material with the others. *Chapter 1.*

compliance The change in lung volume for each unit change in pressure, so a stiff lung has a low compliance. *Chapter 4.*

continuous glucose monitor (CGM) A sensor under the skin that reads the glucose of the fluid there every few minutes. *Chapter 2.*

continuous kidney replacement therapy (CRRT) Treatment that cleans the blood and removes water around the clock for a patient whose kidneys have failed. *Chapter 5.*

critical dilution rate The highest rate at which a continuous vessel can be fed before its microbes wash out. *Chapter 1.*

critical power The highest power a person can sustain for a long time without drawing down a limited reserve. *Chapter 1.*

depth of anesthesia How deeply a patient is unconscious and unresponsive under anesthetic drugs. *Chapter 3.*

diacetyl A buttery-tasting compound made during fermentation that the yeast must take back up before beer is finished. *Chapter 10.*

diacetyl rest A warm hold near the end of fermentation that lets the yeast remove diacetyl. *Chapter 10.*

dissolved oxygen Oxygen held in solution in water or broth, which cells and fish need to breathe. *Chapter 8.*

driving pressure The pressure that stretches the lung in each breath, the plateau pressure minus the pressure left at the end of exhalation. *Chapter 4.*

effect site A notional compartment where a drug acts, joined to the blood by a delay. *Chapter 1.*

electroencephalogram (EEG) A recording of the brain's electrical activity through electrodes on the head. *Chapter 3.*

external reset Feeding the output actually applied

back into a controller's integral, so a controller that is not selected does not wind up. *Chapter 6.*

fed-batch culture A culture that is fed during its run but not emptied until harvest. *Chapter 1.*

feed conversion ratio The mass of feed given for each unit of mass the fish gain. *Chapter 12.*

feedforward Acting on a measured disturbance before it shows up in the controlled variable. *Chapter 3.*

final size The share of a population infected by the end of an epidemic. *Chapter 14.*

fold The point at which two steady states meet and disappear, beyond which a process cannot hold a steady state. *Chapter 1.*

gain scheduling Changing a controller's settings with a known quantity, such as a runner's threshold pace, so one tuning suits many cases. *Chapter 6.*

glycol plant A refrigeration plant that chills a glycol solution and sends it to the jackets of the fermenting tanks. *Chapter 10.*

gravimetric control Balancing fluids by weighing them, using the weight in place of the flow. *Chapter 5.*

Haldane kinetics A growth law in which a substrate feeds the organism at low levels and inhibits it at high levels. *Chapter 1.*

hospital census The number of patients in hospital beds on a given day. *Chapter 14.*

hybrid closed loop An automated insulin system that adjusts the background delivery by itself and leaves the meal doses to the user. *Chapter 2.*

in silico trial A test of a treatment or controller on simulated patients instead of real ones. *Chapter 2.*

inclusion bodies Dense clumps of unfolded protein that bacteria form when they make a foreign protein, here the insulin precursor. *Chapter 9.*

infusion pump A device that pumps fluid or drug into a patient at a controlled rate. *Chapter 3.*

insulin on board The insulin already delivered that has not yet acted. *Chapter 1.*

insulin sensitivity How much glucose falls for a given amount of insulin, and exercise raises it. *Chapter 2.*

Kalman filter An estimator that blends a model's prediction with each new measurement according to their uncertainties. *Chapter 2.*

lactate threshold The exercise intensity above which blood lactate begins to accumulate. *Chapter 6.*

load cell A metal element fitted with strain gauges that turns the force on it into an electrical signal, and scales rest on load cells. *Chapter 1.*

low selector A block that passes the lowest of several controller outputs, so the most limiting one takes the actuator. *Chapter 1.*

maximal lactate steady state The highest intensity at which blood lactate stays constant during a long, steady effort. *Chapter 6.*

minimal model A small model of glucose and insulin used to describe how a person's glucose responds to insulin. *Chapter 2.*

model inversion Computing the input that makes a model reach a target, then applying it to the real plant. *Chapter 1.*

model predictive control (MPC) A controller that plans moves over a horizon with a model, applies the first one and plans again. *Chapter 2.*

monoclonal antibody An antibody made by cells that all descend from one cell, so every molecule is the same. *Chapter 8.*

Monod kinetics A growth law in which growth rate rises with substrate and levels off when substrate is plentiful. *Chapter 1.*

moving horizon estimation (MHE) An estimator that fits the model to a window of recent measurements to find the present state. *Chapter 2.*

net ultrafiltration The water removed from a patient by the kidney machine beyond the fluid it gives back. *Chapter 5.*

nitrification The conversion of ammonia to nitrite and then nitrate by bacteria. *Chapter 12.*

nitrite toxicity The harm that nitrite does to fish, which chloride in the water greatly reduces. *Chapter 12.*

nonpharmaceutical interventions Measures other than drugs and vaccines that slow an epidemic, such as closing schools or limiting gatherings. *Chapter 14.*

one-sided actuator An actuator that can push only one way, like a pump that can add a drug but not remove it. *Chapter 1.*

osmolality The concentration of dissolved particles in a liquid, and added base and salts raise it in a culture. *Chapter 8.*

override control A structure in which a limit controller takes the actuator from the normal controller when a limit is near. *Chapter 1.*

oxygen cone A pressurized cone in which water and pure oxygen are mixed to supersaturate part of a fish farm's flow. *Chapter 12.*

patient variability The differences between individuals in how they take up, clear and respond to a drug. *Chapter 1.*

patient-ventilator asynchrony A mismatch between the breaths a patient tries to take and the breaths the ventilator gives. *Chapter 4.*

pharmacokinetics The study of how a drug spreads through and leaves the body over time. *Chapter 1*.

plateau pressure The pressure in the lungs when flow stops at the end of a breath, a measure of lung stretch. *Chapter 4*.

population model A model of the typical individual of a group, with the spread of individuals around it. *Chapter 1*.

positive end-expiratory pressure (PEEP) The pressure a ventilator keeps in the lungs at the end of each breath out. *Chapter 4*.

power meter A strain-gauge sensor in a bicycle's crank or pedal that reports the rider's power. *Chapter 7*.

pressure support A ventilator mode that adds a set pressure to each breath the patient starts. *Chapter 4*.

propofol An intravenous anesthetic drug given by infusion to keep a patient unconscious. *Chapter 3*.

protein refolding Dissolving unfolded protein and letting it fold into its working shape under controlled conditions. *Chapter 9*.

Raman spectroscopy A measurement of light scattered by molecules that can read several concentrations in a broth at once. *Chapter 8*.

reproduction number The average number of people one infected person infects. *Chapter 14*.

resistance temperature detector (RTD) A temperature sensor whose metal wire changes its resistance with temperature. *Chapter 10*.

sampled measurement A value known only at the moments a sample is drawn and analyzed, often reported later. *Chapter 1*.

scale drift A slow error that builds up in a scale's reading while the true weight is unchanged. *Chapter 5*.

SEIR model An epidemic model that divides people into susceptible, exposed, infectious and removed groups. *Chapter 14*.

soft sensor A model that estimates an unmeasured variable from variables that are measured. *Chapter 3*.

solenoid valve An on/off valve opened or closed by current through an electric coil. *Chapter 10*.

split-range control One controller output shared between two actuators, each working over part of its range. *Chapter 8*.

target-controlled infusion (TCI) A pump that computes from a drug model the infusion that reaches a chosen concentration in the blood or effect site. *Chapter 1*.

temperature integration Holding a crop's average temperature over a day or more while letting the hourly temperature vary to save energy. *Chapter 13*.

thermal screen A movable sheet drawn over a glasshouse crop at night to hold the heat in. *Chapter 13*.

time in range The share of time that glucose stays inside a target band. *Chapter 1*.

time-proportioned control Turning an on/off valve open for a share of each cycle in proportion to a controller's output. *Chapter 10*.

transmembrane pressure The pressure difference across a filter membrane that drives water through it. *Chapter 5*.

volatile acids to alkalinity ratio A titration result that compares a digester's acids with its buffer, where a rising ratio warns of an upset. *Chapter 11*.

washout The loss of a microbe population from a continuous vessel fed faster than the microbes can grow. *Chapter 1*.

work capacity above critical power A limited store of work a person can spend above critical power, which refills slowly at lower efforts. *Chapter 1*.

yeast Single-celled fungi that ferment sugar into alcohol and carbon dioxide. *Chapter 10*.

Index of Control Topics

activation trigger, 368
active constraint
  moving with the conditions, 149
actuator saturation, 33, 121, 169
  on a store, 181
adaptive control, 97
AI agents
  causal claims, 381
  engineering practice, 379
  open problems, 384
  reproduction, 382
  truth audit, 381
  verification record, 380
  while studying, 17
anti-windup
  back-calculation, 243
Arrhenius law, 261
artificial intelligence
  greenhouse, 346
artificial pancreas
  open problems, 48
asymmetric cost, 10, 23, 376

back-calculation, 33, 150, 184, 213, 289, 313
batch control, 256
batch scheduling, 234
  bottleneck vessel, 237
bias update, 8, 214
biological actuator, 304
bistability, 280
bumpless transfer, 16, 96, 122, 150

cascade control, 65, 244, 313, 336
charge balance, 286
closed-loop control
  history, 3, 57
  physiological, 105
compartment model, 5, 57
compartmental model, 353
conditional integration, 186
continuous glucose monitor, 11
control layers, 12
control structure
  choice before tuning, 72, 377
controller action, 23, 66, 96, 121, 213, 244, 265, 289, 335, 359
convex hull, 337
critical dilution rate, 288

daily mean temperature, 327
data export, 17
data reconciliation, 113
delayed assay, 234
depth of anesthesia
  open problems, 80
digital twin, 379
dissolved oxygen sensor, 319
drug synergy, 60

economic optimization, 314
effect site, 5, 57
external reset, 151, 185
external reset feedback, 336

fault detection, 113
feedforward, 122, 185, 294
  from an uninstrumented event, 72
  meal, 313
  measured load, 151
feedforward control
  feedback trim, 214, 244
  model-based, 96, 214
fermentation monitoring, 275
final control element, 11
final size
  history, 353
first-order plus dead-time model, 31
flow scheme, 5
fold bifurcation, 281

gain scheduling, 150
golden-section search, 314
gravimetric control, 111
growth kinetics, 7

Haldane kinetics, 7, 285
high select, 336
humidity control, 326
humidity sensor, 344
hybrid systems, 86

in silico trial, 377
  virtual population, 384

in-process control, 250
infeasible setpoint, 218
inferential control, 54, 266
infusion pump, 12
integrated absolute error, 36
integrating process, 119

Kalman filter, 16
    extended, 358
    input disturbance, 34, 216

laboratory measurement, 8, 290
least squares, 96
limit cycle, 268
linear programming, 337
linearization, 215
low select, 314

measurement artifact, 61
measurement delay, 313, 357
measurement lag, 213
measurement redundancy, 111
medical device regulation, 47
model inversion, 7
model predictive control
    adaptive, 67
    against PID, 376
    economic, 137, 186, 215, 354
    in the apps, 16
    move blocking, 151, 215, 359
    multi-stage, 359
    nominal model, 290
    preview of a disturbance, 35
    successive linearization, 359
model-based measurement, 66
Monod kinetics, 7, 207, 310
moving horizon estimation, 16, 35, 358
    arrival cost, 358

negative binomial distribution, 356
non-anticipativity, 359
nonpharmaceutical interventions, 367

offset-free control, 216
on-off control, 256
    differential gap, 275
    hysteresis, 265, 359
one-sided actuator, 10, 23, 64, 94, 376
online analyzer
    aquaculture, 321
oscillation
    aggressive tuning, 155

outlier, 362
override control, 11, 122, 137, 169, 314, 336
    equipment protection, 125

particle filter, 353
    tempered update, 358
patient variability, 7, 63, 375
patient-ventilator asynchrony
    automatic detection, 105
    detection, 105
pharmacokinetics, 5
PID control
    anti-windup, 15, 65
    direct acting, 32
    velocity form, 15, 96, 121, 213, 243, 289, 313
PID tuning
    internal model control, 32
plant-model mismatch, 54
population model, 7
positive feedback, 121
process analytical technology, 227, 251
process gain
    sign change, 121
    variable, 63
    varying, 148, 170

receding horizon, 151, 186
receding-horizon control, 327
recursive least squares, 67
relay autotuning
    switching band, 36
reporting delay, 351
    correction, 353
reproduction number
    real-time estimation, 368
resistance temperature detector, 273
response surface, 60

safety limits, 10, 11, 97
sample interval
    choice of, 195
sampled analyzer, 213, 241
    at-line, 248
sampled measurement, 8, 184, 311
    band condition, 211
    correction at the draw, 244
    daily titration, 280
    gain limit, 294
scale, 12
scenario tree, 359

scenarios, 17
  challenge, 101
  completion rules, 17, 37
  graded challenge, 127, 222, 295
scorecard, 17, 152, 187
secant gain, 32
selector
  min/max, 122, 137, 169
sensor bias, 120, 221, 375
sensor fault
  dropout, 126
sensor placement, 270, 346
sequential control, 315
shared utility, 265
signal validation, 66, 120, 151, 186, 375, 381
simulation engine, 14
simulator framework, 14
soft constraint, 216, 359
soft sensor, 56, 202, 315
  bias update, 214, 266
solenoid valve, 274
split-range control, 214
state estimation, 244, 290, 315, 351
  anchoring measurement, 361
  augmented state, 358
steady-state map, 30
step test
  on a nonlinear plant, 31
stratification, 262
supervisory control, 124, 214, 266, 290, 337
  digester feed, 299
surveillance
  case reports, 366
system identification
  bias from an unmeasured input, 95
  breath by breath, 86
system type, 120

target-controlled infusion, 55
  practice, 79
temperature integration, 326, 329
  history, 329
terminal cost, 151
time-proportioned control, 265
time-scale separation, 120
titration
  volatile acids and alkalinity, 297
two-step digestion model, 285

variable frequency drive, 297
volatile acids to alkalinity ratio, 298

washout, 288
weather forecast, 333

Index of Applications

acid precipitation, 240
acute respiratory distress syndrome, 87
adaptive support, 97
ADM1, 283
aerodynamic drag, 175
aggregation kinetics, 235
alkalinity, 282
anaerobic digester, 280
    farm, 282
    history, 282
    sensors, 379
anaerobic work reserve, 142, 171
anesthesia, 54
    closed-loop, 54
apparent wind, 178
arterial pressure, 60
artificial pancreas, 23
    history, 25
ascertainment, 352
auto-PEEP, 89
auto-trigger, 100
automated insulin delivery, 377

bag change, 117
bag ventilation, 98
Balling relations, 263
base addition, 212
beer fermentation, 256
Bergman minimal model, 26
bicarbonate buffer, 209
bicarbonate dosing, 283
bicycle
    gearing, 175
biofilter, 306
    startup, 312, 315
biogas, 280
biogas upgrader, 283
biomass estimation, 315
biomedical control, 1
    milestones, 2
    open problems, 382
bioprocess control, 1
bioreactor, 202
bispectral index, 57
blood glucose control, 23
blood lactate, 143, 177
blood loss, 71
blower, 89
boiler, 330
bronchospasm, 100

cadence, 174
capacitance probe, 209
    size and count, 228
carbohydrate feeding, 145, 179
carbon dioxide enrichment, 330
    vent cap, 337
carbon dioxide stripping, 208
    gas-flow bound, 211
charge pump, 249
Chinese hamster ovary cells, 203
climate computer, 328
    history, 329
co-digestion, 281
continuous glucose monitor, 28
    enzyme electrode, 46
continuous kidney replacement therapy, 111
    history, 113, 378
core temperature, 144, 177
crash cooling, 271
critical power, 142, 169
    history, 172
    work capacity, estimation of, 196
cycle-off, 94
cylindroconical tank, 258

dairy manure, 281
degasser, 306
density meter, 265
depth of anesthesia monitor, 61, 379
    processed electroencephalogram, 77
diacetyl, 259
    flavor threshold, 267
diacetyl rest, 270
dialysis dose, 117
    measurement, 131
digester recovery, 293
dissolved oxygen, 208
distancing measures, 354
double trigger, 100
downtime, 130

driving pressure, 87

electrocautery, 71
end of a case, 71
enzymatic conversion, 240
epidemic, 351
epidemic growth rate, 357
equation of motion (lung), 90
ergometer, 171
Escherichia coli, 236
exercise physiology, 137, 169
expiratory hold, 100

fed-batch charging, 239
fed-batch culture, 202
    dilution, 206
feed conversion ratio, 307
feed governor, 314
fermentation
    heat of, 257
film observer, 315
filter fouling, 117
filtration fraction, 116
final size, 357
fish
    appetite, 309
fish farm, 304
flare, 283
flow sensor
    ventilator, 104
flow trigger, 94
fluid balance, 111
fluid ledger, 118
fogging, 330
food waste, 294

gas holder, 283
glasshouse, 327
glucose feeding
    daily bolus, 214
    dosing pump, 227
glycogen, 144, 178
glycol plant, 257
gravity, apparent extract, 263
greenhouse, 326
gross efficiency, 176

harvest decision, 214
    viability limit, 221
heart rate, 143, 176
    feedback control in practice, 161
heart rate monitor
    chest strap, 160
heat stress, 144
heater fault, 294
heating pipe rail, 344
hemodiafiltration, 111
hold time, 239
hospital census, 351
host-cell protein, 239
humidity control
    practice, 345
hybrid closed loop, 26

in silico trial, 26
inclusion bodies, 234, 236
inclusion-body wash, 239
infusion pump, 61
    target-controlled, 79
instrument, 11
insulin, 234
    history, 4, 236
insulin on board, 25
insulin pump, 24
    subcutaneous infusion, 47
insulin sensitivity
    exercise, 28

jacket, cooling, 262

lactate
    noninvasive measurement, 162
lactate switch, 207
    glucose threshold, 212
lactate threshold, 143
lager, 256
LC50, 306
leak
    estimate, 96
    past the flow sensor, 91
lung-protective ventilation, 97

maximal lactate steady state, 143
meal disturbance, 28
mechanical ventilation, 86
Minto model, 57
missed effort, 100
monoclonal antibody, 202
    origin, 204

net ultrafiltration rate, 131
new-tank syndrome, 312
nitrification, 310

nitrifying bacteria, 310
nitrite toxicity, 306
  chloride, 307
nonpharmaceutical interventions
  history, 352

orifice law, 91
osmolality, 209
  cost of base, 212
overflow metabolism, 207
oxygen cone, 306, 320
  history, 306

pacing, 137, 172
  cue on a display, 195
patient-ventilator asynchrony, 89
  history, 99
Penman–Monteith equation, 332
perceived exertion, 170
pH probe drift, 221
pharmacokinetics, 5
photosynthesis, 333
plasma refill, 117
plateau pressure, 93
power failure, 321
power meter
  strain gauge, 194
pressure control, 93
pressure support, 88
process analytical technology, 378
propofol, 54
proportional valve, 89
protein refolding, 234

rainbow trout, 305
Raman spectroscopy, 213
  in-line probe, 227
recirculating aquaculture, 304
refolding tank, 240
refrigeration, 258
remifentanil, 54
renewable natural gas, 283
reporting outage, 356
reproduction number, 355
respiratory mechanics, 90
road cycling, 169
roller pump, 114
  tubing segment, 129
  under-filling, 114
rolling resistance, 175
running economy, 141

scale
  gravimetric balancing, 129
scale drift, 117
Schnider model, 57
SEIR model, 353, 354
souring, 293
sparger, 208
stuck fermentation, 272
substrate oxidation, 144
surgical stimulation, 60

tank respirator, 88
target-controlled infusion
  history, 56
temperature shift, 214
testing cut, 356
thermal screen, 328
thermal-growth coefficient, 307
three-minute all-out test, 171
time in range, 36
titration ratio, 282
  history, 282
tomato crop, 326
transmembrane pressure, 116
transpiration, 332
treadmill, 137
  belt drive, 161
  belt motor, 145
  heart-rate control, 140
type 1 diabetes, 24

un-ionized ammonia, 309
under-reporting, 352
urea clearance, 117

vaccination, 355
valve
  ventilator, 104
variant, 355
vasopressor, 60
Venlo glasshouse, 329
ventilation, natural, 332
ventilator
  alarms, 97
  history, 89
  intensive care, 86
volatile fatty acids, 282
volume control, 93

water-quality analyzer, 311

yeast
  growth and uptake, 261
  viability, 272

About the Editor

John D. Hedengren is a professor in the Department of Chemical and Biological Engineering at Brigham Young University, where he leads the Process Research and Intelligent Systems Modeling (PRISM) group. He received the B.S. and M.S. degrees in chemical engineering from Brigham Young University and the Ph.D. degree from the University of Texas at Austin in 2005. He developed advanced process control applications for polymer production at ExxonMobil Chemical Company and commissioned nonlinear model predictive controllers with PAS, Inc. before joining the faculty at Brigham Young University in 2011.

His teaching covers process dynamics and control, engineering computing, data science, machine learning, and dynamic optimization, with course material published online for use inside and outside the classroom. He developed the Temperature Control Lab, a portable heat transfer device now used by many universities for process control and data science instruction, and he maintains the APMonitor and GEKKO optimization software used in those courses. He serves as Chair of Control Education for the IEEE Control Systems Society and as a trustee of the CACHE Corporation. He received the AIChE David Himmelblau Award in 2014, the AIChE Computing Practice Award in 2018, and the John R. Ragazzini Award in 2024, and he was inducted into the Process Automation Hall of Fame in 2025.

Volumes in This Series

Interactive Control Engineering: Principles and Applications has 20 volumes. Nine are organized by control technology and eleven by industry or application area. Every chapter in every volume has a simulator at https://apmonitor.com/apps/.

1 **Process Dynamics, Modeling, and Identification.** How plants respond, and how to find the model.
2 **Practical PID Control.** Tuning, anti-windup and implementation that works on real loops.
3 **Feedforward, Cascade, and Ratio Control.** Choosing a structure from where the disturbance enters.
4 **Override, Split-Range, and Constraint Control.** When limits decide what the controller may do.
5 **Delays, Slow Measurements, and State Estimation.** Controlling what arrives late or cannot be measured.
6 **Multivariable and Plantwide Control.** Pairing loops, managing inventories, closing the recycle.
7 **Nonlinear Dynamics, Oscillations, and Stabilization.** Plants that will not stay put without control.
8 **Batch, Run-to-Run, and Hybrid Control.** Recipes, endpoints and learning from the last run.
9 **Model Predictive Control and Economic Optimization.** Planning moves ahead, and choosing where to operate.
10 **Control in Upstream and Midstream Oil and Gas.** From the wellbore to the terminal.
11 **Control in Petroleum Refining.** Throughput, quality and limits from crude unit to blending.
12 **Control in Petrochemicals and Polymers.** Severity, selectivity and grade from cracker to pellet.
13 **Control in Chemical Plants.** Recycle, equilibrium and energy in continuous chemical processes.
14 **Control in Semiconductor Manufacturing.** Run-to-run control where little can be measured in the chamber.
15 **Control in Mining and Mineral Processing.** Variable ore, late assays and equipment at its limit.
16 **Control in Iron, Steel, and Metals Production.** Slow furnaces, fast strip, and measurements that arrive once.
17 **Control of Energy Conversion and Storage.** Fast regulation and slow dispatch on a changing grid.
18 **Control of Nuclear and Fusion Systems.** Reactor physics, protection limits and coordinated control.
19 **Control of Biological Systems and Human Performance.** Variable subjects, sparse samples and limits that cannot be crossed.
20 **Motion Control, Robotics, and Autonomous Systems.** Tracking, navigation, contact and allocation.