INTERACTIVE CONTROL ENGINEERING

Process Dynamics, Modeling, and Identification

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 1

Process Dynamics, Modeling, and Identification

How plants respond, and how to find the model

John D. Hedengren

Editor

Advanced Process Solutions, LLC

Interactive Control Engineering: Principles and Applications
Volume 1: Process Dynamics, Modeling, and Identification
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/v01-dynamics-modeling/

How to cite this volume. John D. Hedengren, Ed., *Process Dynamics, Modeling, and Identification*, vol. 1 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 how plants respond and how to find the model that describes the response. Chapters 2 to 10 explain why a plant behaves as it does: a tank whose time constant moves with its level, two tanks whose order depends on a valve, a heater pair, an unstable column of air, a flash vessel whose level first moves the wrong way, water inertia in a penstock, the swing of a power grid, a pool that only integrates, and a market with a slow supply response. Chapters 11 to 13 show how to obtain a model from data, check that its structure can be identified, and watch a running loop for the signs that the model no longer fits.

The volume is the starting point of the series. Volume 2 takes the models found here and tunes controllers for them.

John D. Hedengren
2026

Contents

Preface v

Notation x

1 Introduction: Process Models From Balances and Data 1
- 1.1 The Question This Volume Answers . . . 1
- 1.2 A Short History . . . 2
- 1.3 Notation and Conventions . . . 3
- 1.4 From Balances to Models That Fit Data . . . 3
- 1.5 Instruments and Final Elements in This Volume . . . 12
- 1.6 The Simulator Framework . . . 14
- 1.7 Working With AI Agents While Studying . . . 18
- 1.8 Chapter Roadmap . . . 18
- 1.9 Exercises . . . 21

2 Gravity-Drained Tank Laboratory for Level Loops With a Moving Time Constant 23
- 2.1 Introduction . . . 24
- 2.2 Background . . . 25
- 2.3 Process Model . . . 26
- 2.4 Dynamic Analysis . . . 30
- 2.5 Controller Design . . . 32
- 2.6 Results . . . 36
- 2.7 Technology in Practice . . . 45
- 2.8 Limitations . . . 47
- 2.9 Discussion . . . 47
- 2.10 Conclusion . . . 48
- 2.11 Exercises . . . 50

3 Two-Tank Level Control With a Splitter Valve That Sets the Plant Order 53
- 3.1 Introduction . . . 54
- 3.2 Background . . . 55
- 3.3 Process Model . . . 56
- 3.4 Effect of the Splitter Fraction on Plant Order . . . 61
- 3.5 Controller Design . . . 63
- 3.6 Results . . . 67
- 3.7 Load Disturbance Rejection . . . 70
- 3.8 Predictive Control on Both Plant Structures . . . 72
- 3.9 Technology in Practice . . . 76
- 3.10 Limitations . . . 78
- 3.11 Discussion . . . 78
- 3.12 Conclusion . . . 79
- 3.13 Exercises . . . 81

4 A Two-Heater Digital Twin That Runs Beside the Device It Models 84
- 4.1 Introduction . . . 85
- 4.2 Background . . . 86
- 4.3 Process Model . . . 87
- 4.4 Steady-State Gains and Open-Loop Dynamics . . . 90

4.5 Controller Design . . . 91
4.6 Hardware Interface . . . 96
4.7 Results . . . 97
4.8 Technology in Practice . . . 104
4.9 Limitations . . . 106
4.10 Discussion . . . 106
4.11 Conclusion . . . 107
4.12 Exercises . . . 109

5 Air-Levitated Ball in a Tube Whose Height No Valve Angle Holds 112
5.1 Introduction . . . 113
5.2 Floating Balls in Teaching Laboratories . . . 114
5.3 The Rig as Modeled . . . 115
5.4 An Integrator Behind a Valve . . . 120
5.5 The Controller as Coded . . . 121
5.6 Ten Cases and a Challenge, Measured . . . 123
5.7 The Rig Beside Its Twin . . . 127
5.8 Technology in Practice . . . 128
5.9 Limitations . . . 130
5.10 Conclusion . . . 130
5.11 Exercises . . . 132

6 A Flash Steam Vessel Whose Level Rises While It Loses Water 135
6.1 Introduction . . . 136
6.2 Flash Steam, Recovered Heat and the Level Problem . . . 137
6.3 The Unit as Modeled . . . 138
6.4 Dynamic Character . . . 144
6.5 Implemented Control System . . . 145
6.6 Fifteen Cases, Measured . . . 148
6.7 Technology in Practice . . . 151
6.8 Limitations . . . 154
6.9 Conclusion . . . 154
6.10 Exercises . . . 156

7 Water Inertia, Transient Droop and Closure Laws in a Hydroelectric Governor Simulator 160
7.1 Introduction . . . 161
7.2 From Dashpot Governors to Closure Laws . . . 162
7.3 The Unit as Modeled . . . 162
7.4 Why the Water Limits the Governor . . . 166
7.5 The Governors as Implemented . . . 169
7.6 Nine Cases, Measured . . . 170
7.7 Technology in Practice . . . 174
7.8 Limitations . . . 176
7.9 Conclusion . . . 176
7.10 Exercises . . . 178

8 Grid Frequency Simulator for Inertia, Droop Offset and Integral Restoration 181
8.1 Introduction . . . 182
8.2 Background . . . 183
8.3 Process Model . . . 184
8.4 Dynamic Analysis . . . 188
8.5 Controller Design . . . 190

8.6 Results . . . 192
8.7 Technology in Practice . . . 196
8.8 Limitations . . . 198
8.9 Discussion . . . 198
8.10 Conclusion . . . 199
8.11 Exercises . . . 200

9 Spent-Fuel Pool Simulator for Integrating Inventories and Time-to-Limit Prediction 204
9.1 Introduction . . . 205
9.2 Background . . . 206
9.3 Pool Model . . . 207
9.4 Dynamic Analysis . . . 214
9.5 Controller Design . . . 216
9.6 Results . . . 219
9.7 Technology in Practice . . . 225
9.8 Limitations . . . 227
9.9 Discussion . . . 228
9.10 Conclusion . . . 228
9.11 Exercises . . . 230

10 Oil Price Versus Market Share in a Cartel's Quota Loop 234
10.1 Introduction . . . 235
10.2 From Exhaustible Resources to the Shale Decade . . . 236
10.3 One Tank for the World Market . . . 237
10.4 Two Speeds of the Market . . . 241
10.5 Deciding at Meetings . . . 242
10.6 Ten Years at a Time . . . 246
10.7 Technology in Practice . . . 250
10.8 Limitations . . . 252
10.9 Conclusion . . . 252
10.10 Exercises . . . 254

11 Model Structure Selection and Identifiability on a Temperature Control Lab 258
11.1 Introduction . . . 259
11.2 Background . . . 260
11.3 Process Models and Data Generator . . . 261
11.4 Results . . . 267
11.5 Technology in Practice . . . 275
11.6 Limitations . . . 277
11.7 Discussion . . . 278
11.8 Conclusion . . . 279
11.9 Exercises . . . 280

12 Multivariable Plant Identification Workbench with Exact Structural Zeros 284
12.1 Introduction . . . 285
12.2 Plant Testing and Identification for Multivariable Control . . . 286
12.3 Identification Problem and Data Model . . . 287
12.4 Identification Algorithms as Implemented . . . 292
12.5 Predictive Control from the Fitted Model . . . 297
12.6 Measured Results . . . 299
12.7 Technology in Practice . . . 306
12.8 Limitations . . . 308

12.9 Discussion . . . 309
12.10 Conclusion . . . 309
12.11 Exercises . . . 311

13 Process Monitoring Workbench That Diagnoses Records From Declared Simulated Plants 315
13.1 Introduction . . . 316
13.2 Balances, Charts and Principal Components on Plant Records . . . 317
13.3 The Records and Their Declarations . . . 318
13.4 Five Tools as Implemented . . . 320
13.5 Ten Cases, Measured . . . 323
13.6 MPC Health From an Exported Run . . . 328
13.7 Verification . . . 328
13.8 Technology in Practice . . . 330
13.9 Limitations . . . 332
13.10 Conclusion . . . 332
13.11 Exercises . . . 334

14 Summary and Outlook 338
14.1 What the Chapters Showed . . . 338
14.2 Choosing a Method . . . 341
14.3 Where Practice Is Heading . . . 342
14.4 Engineering With AI Agents . . . 344
14.5 Gaps in Modeling and Identification . . . 347
14.6 Exercises . . . 351

Answers to Selected Exercises 352

Glossary 357

Index of Control Topics 360

Index of Applications 364

About the Editor 366

Volumes in This Series 367

Notation

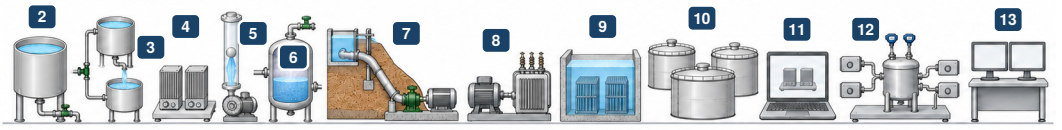
The chapters share the symbols below. A symbol used by one chapter only is defined where it first appears, and each chapter states its own units.

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: Process Models From Balances and Data



The chapters 2 draining tank 3 two tanks 4 two-heater board 5 levitated ball 6 flash steam vessel 7 hydro governor 8 grid frequency 9 spent-fuel pool 10 oil market stocks 11 heater model choice 12 identification workbench 13 monitoring workbench

Notice that the first four units are laboratory plants, the next five carry the same balances to a steam vessel, a hydroelectric unit, a grid, a fuel pool and a market, and the last three build and judge models from records. Illustration, not to scale.

IN PLAIN TERMS **How a Plant Answers a Push**

The main idea. A controller needs a model, a short account of how its plant answers a push. The model comes from counting what flows in, what flows out and what is stored, or from pushing the real plant and fitting a curve to the record.

The challenge. The answer changes as the plant moves, so a model that is right at one level can be wrong at another. Some plants answer late or start the wrong way, and a curve that fits one record can fail on the next.

Why it matters. The same few shapes describe tanks, heaters, turbines, power grids and the oil market, so reading them well pays almost everywhere.

THREE WAYS TO READ THIS CHAPTER

Explore	7 min	This page, Section 1.2 with its timeline and two History boxes, the Did you know? boxes in Sections 1.4.6, 1.4.8 and 1.4.9, Figure 1.11, and Key points in Section 1.8. No equations.
Study	40 min	The whole chapter in order, with each number of the tank example worked, then Exercises 1 to 10.
Practice	10 min	This page, Sections 1.4.8 and 1.4.9 with Table 1.1, Section 1.5 with Figures 1.6 and 1.9, and Section 1.8.

1.1 The Question This Volume Answers

Every controller in this series acts on a model of its plant. Sometimes the model is only the sign of a gain and a rough time scale, and sometimes it is a set of differential equations solved at every decision. A controller tuned on the wrong model is mistuned however good its algorithm. This volume answers the question that comes before control: how is a model of a dynamic process built, how is it reduced to the few numbers a controller needs, and how is it checked against data?

Two routes lead to a model, and most plants need both. The first writes the balances of mass, energy or momentum that the plant must obey and solves them. The second runs an experiment on the plant and fits a model to the record. A balance says which form the model should take and how its constants depend on the operating point. An experiment supplies the numbers that the balance leaves unknown and shows where the balance is wrong.

Twelve chapters follow the two routes, each on a plant with an app that runs in a browser. Chapter 2, Chapter 3, Chapter 4 and Chapter 5 are laboratory plants: a tank that drains under its own head, two tanks whose order changes with a valve, a board with two heaters and a ball floating in a tube of air. Chapter 6, Chapter 7, Chapter 8, Chapter 9 and Chapter 10 are larger plants in which the model carries the lesson: a vessel whose level rises while it loses water, a turbine whose power first falls when its gate opens, the frequency of a power grid, a fuel pool that holds its temperature only while

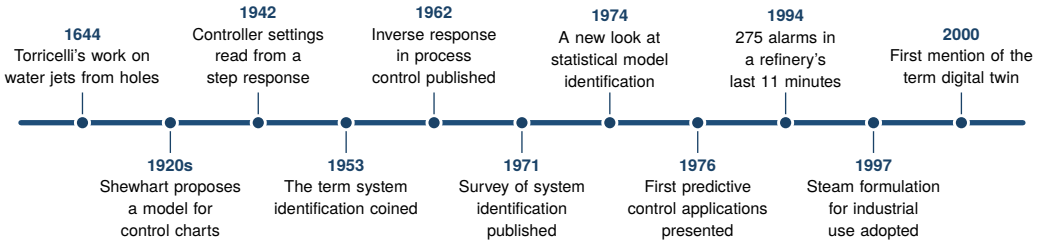


Figure 1.1 Milestones of dynamic modeling, identification and monitoring. The dates are from [4] (1644), [6] (1920s), [7] (1942), [8] (1953), [9] (1962), [10] (1971), [11] (1974), [12] (1976), [13] (1994), [5] (1997) and [14] (2000). Notice that the balance is centuries older than the fitted model, and that the checks on a model in service came last. Events are evenly spaced, not to a time scale.

its pumps run, and an oil market treated as one tank of stocks. Chapter 11, Chapter 12 and Chapter 13 build and judge models from records: choosing between energy balances, identifying a multivariable plant and diagnosing faults in plant data.

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

- Write the mass and energy balances of a process and find its steady state.
- Linearize the balances about an operating point in deviation variables, and say how the gain and the time constant move with that point.
- Recognize first-order, second-order, integrating, dead-time and inverse responses in a step test, and name the physical feature that produces each.
- Fit a model to a step test with a graphical rule and by least squares, in continuous and in discrete time.
- Judge a fitted model by its residuals and on data it was not fitted to, and say which of its parameters the data can determine.

This chapter develops each skill on one small example and names the chapters that use it. A reader new to process models should read Section 1.4 in order. It assumes calculus and the Laplace transform at the level of a first course. Standard texts give fuller derivations [1, 2], and a survey places the methods of identification in context [3]. Later volumes build on this one. Volume 2 tunes controllers from the models fitted here, and Volume 5 and Volume 7 return to dead time and to nonlinear models.

1.2 A Short History

The milestones of Figure 1.1 follow the two routes to a model, the balance and the test, and the monitoring that checks a model once it is in service. The balance is the older route. Torricelli's work of 1644 on water jets leaving holes in a vessel is one of the founding results of fluid mechanics [4], and the orifice of TK-1 drains by the same physics. A balance also needs the properties of what it stores. International conferences on steam tables met in 1929, 1930 and 1934, equations replaced the tables as computers spread through industry, and a formulation for industrial use was adopted in 1997 [5]. Chapter 6 closes its balances with that formulation.

The test is younger. Tuning rules published in 1942 read two numbers from a step response [2, 7], and the term system identification dates from 1953 [8]. A paper on inverse response in process control appeared in 1962 [9], the two search methods of Section 1.4.8 in 1963 and 1965 [15, 16], a survey of system identification in 1971 [10] and the information criterion of Equation (1.23) in 1974 [11]. The first industrial applications of predictive control, presented in 1976, identified an impulse response model from plant test data [12], the model family of Chapter 12.

> ⧖ HISTORY **1942: Two Numbers From a Bump Test**
>
> The first tuning rules, from Ziegler and Nichols, rested on one simple experiment. The process was brought to steady state, its input was changed by a suitable amount, and the output was recorded and scaled to a unit step. Two numbers were read from the record, the intercepts of its steepest tangent with the axes: an apparent time delay and, with it, the steepest slope. The test could stop once the response had passed its inflection point. The rules had a huge impact and makers of controllers adopted them for routine use, but they used too little process information [2, 7]. The model of Equation (1.15) adds the gain, which needs the final value, and a time constant read at 63 % of the change, as Equation (1.18) does.

Monitoring asks whether a model still holds. During the 1920s Shewhart proposed the control chart, with limits a fixed number of standard deviations from the mean [6]. Before a refinery explosion of 1994, two operators had to recognize, acknowledge and act on 275 alarms in the last 11 minutes [13], the problem behind the alarm tools of Chapter 13. The term digital twin is traced back only to 2000 [14], and Chapter 4 runs such a twin beside its board.

> ⧖ HISTORY **1940s to 1970s: Two Ways to Write Down a Plant**
>
> When control became a discipline in the 1940s, it saw dynamics through the electrical engineer's input and output view, in which a system is described by its recorded response to a step or to sinusoids. A second wave from the late 1950s took the state space of mechanics, with space flight as a typical example, and the two views merged into the state-space model of an input-output system [2]. The term system identification was coined in 1953. The field then followed two routes, a statistical one built on parameter estimation such as maximum likelihood, and a realization one that builds a linear system from input-output descriptions such as impulse responses [8]. Equation (1.7) is the state-space form, and Section 1.4.8 takes the statistical route.

1.3 Notation and Conventions

The chapters share the symbols of the notation page. The setpoint is SP, the process variable is PV and the manipulated variable is MV. A model of one input and one output has a gain K_p, a time constant τ_p and a dead time θ_p, in PV units per MV unit and in the time unit of the plant. A prime marks a deviation from the operating point, so $L' = L - \bar{L}$, and a bar marks the operating point itself. A hat marks a value fitted or predicted by a model, so $\hat{y}_k$ is the model's output at sample k. Samples are taken every T_s, and a dead time of d samples is dT_s. The difference $\varepsilon_k = y_k - \hat{y}_k$ between a measurement and a model is a residual. Time in the running example is in minutes.

Several letters need care across the series. The shared framework uses h for the integration step, a for the controller action and e for the control error, so this chapter writes the tank level as L, the coefficients of a discrete model as ϕ and γ and a residual as ε. The chapters keep their own notation. Chapter 2 and Chapter 3 write the level as h, and Chapter 2 writes the process gain as K_u and the disturbance gain as K_d. Chapter 5 writes the integrating gain as K_i, Chapter 9 writes it as K_I and the pool level as z, and Chapter 8 uses K_I for the gain of its integral restoring controller. Chapter 12 writes the coefficients of its discrete models as a and b and its dead times in samples as d_{ij}.

1.4 From Balances to Models That Fit Data

1.4.1 The Running Example

The running example is a heated water tank TK-1 that drains under its own head, shown in Figure 1.2. Water enters at the flow F, in $\mathrm{m^3/min}$, and at $T_{in} = 20\,^{\circ}\mathrm{C}$. It leaves through an orifice at the bottom at the rate $\beta\sqrt{L}$, where L is the level in meters and $\beta = 0.2\,\mathrm{m}^{2.5}/\mathrm{min}$ lumps the orifice area, its discharge coefficient and gravity. The tank has a cross-section of $A_1 = 1\,\mathrm{m}^2$. An electric heater supplies W percent of 280 kW. The outlet water passes through a pipe that holds $0.2\,\mathrm{m}^3$ to a thermometer TT-1 in a thermowell. The level transmitter LT-1 reads L directly. At the operating point the feed is $\bar{F} = 0.2\,\mathrm{m}^3/\mathrm{min}$, the level $\bar{L} = 1\,\mathrm{m}$, the heater $\bar{W} = 50\,\%$ and the temperature $\bar{T} = 30\,^{\circ}\mathrm{C}$. Two variants

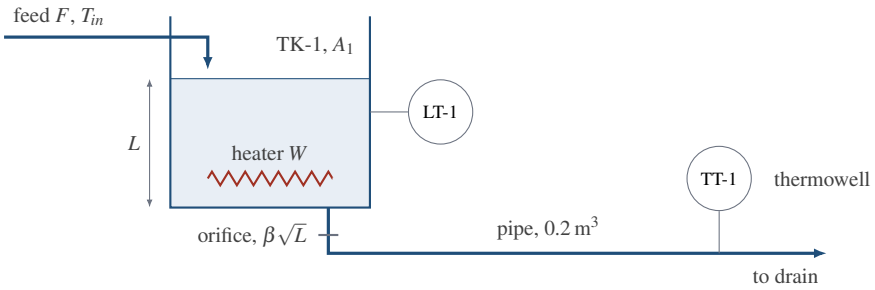


Figure 1.2 The Running Example. The Tank TK-1 Is Fed at F and Drains Through an Orifice. A Heater Warms the Water, and the Thermometer TT-1 Reads the Outlet at the End of a Pipe, Inside a Thermowell.

appear later: a pump in place of the orifice in Section 1.4.4, and a second tank under the first in Section 1.4.5.

1.4.2 Balances Become Differential Equations

Every model in this volume starts from one statement, written for a chosen quantity and a chosen boundary, as shown in Equation (1.1).

$$\text{rate of accumulation} = \text{rate in} - \text{rate out} + \text{rate generated} \tag{1.1}$$

The quantity may be mass, energy, momentum or a number of barrels, and the boundary may be a tank, a heater or a market. When the stored amount is written in terms of a measurable variable, the statement becomes a differential equation in that variable.

For TK-1 the stored mass is $\rho A_1 L$ at constant density ρ, so the mass balance is Equation (1.2).

$$A_1 \frac{\mathrm{d}L}{\mathrm{d}t} = F - \beta\sqrt{L} \tag{1.2}$$

The outflow term is the square-root law of an orifice under a head L. The stored energy is $\rho c_p A_1 L T$ relative to a reference temperature, where c_p is the heat capacity. The water brings in $\rho c_p F T_{in}$, carries out $\rho c_p \beta\sqrt{L} T$ and gains the heater power q_W in kJ/min. Expanding the derivative of the product LT and subtracting T times Equation (1.2) gives the energy balance of Equation (1.3).

$$A_1 L \frac{\mathrm{d}T}{\mathrm{d}t} = F\,(T_{in} - T) + \frac{q_W}{\rho c_p} \tag{1.3}$$

The model rests on four assumptions: the tank is well mixed, the density and heat capacity are constant, the walls lose no heat and the outlet carries the tank's temperature. With $\rho c_p = 4200\,\mathrm{kJ/(m^3\,K)}$ the flow $\bar{F}$ carries 14 kW for each kelvin of warming, so each percent of heater power, 2.8 kW, raises the outlet by 0.2 °C.

A steady state sets every derivative to zero. Equation (1.2) then gives $\bar{L} = (\bar{F}/\beta)^2 = 1\,\mathrm{m}$, and Equation (1.3) gives $\bar{T} = T_{in} + 0.2\bar{W} = 30\,^\circ\mathrm{C}$. The steady state answers where the plant settles. The derivatives answer how fast it gets there, which is the subject of the rest of this section.

Each plant of the volume is an inventory written this way. Chapter 2 and Chapter 3 write the mass balance of Equation (1.2) for one and two tanks. Chapter 4 writes an energy balance for each heater and each sensor, and Chapter 11 writes three energy balances of increasing detail for one heater. Chapter 6 and Chapter 9 write mass and energy balances together, closed by steam tables in one case and by decay heat and evaporation in the other. Chapter 8 writes the swing equation, an energy balance on the rotating machines whose stored energy is read as frequency. Chapter 7 writes a momentum balance on the water in a penstock, and Chapter 5 one on a floating ball. Chapter 10 writes a balance on commercial oil stocks, filled by supply and drained by demand.

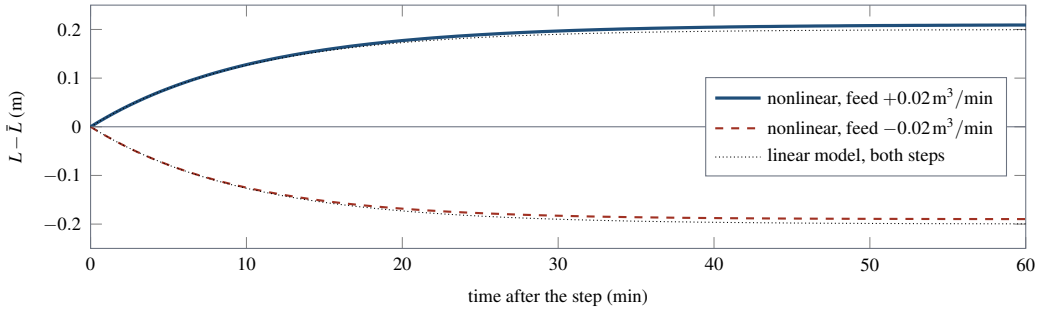


Figure 1.3 Level of TK-1 After Feed Steps of $\pm 0.02\,\mathrm{m}^3/\mathrm{min}$. Solid: Nonlinear Balance, Step Up. Dashed: Nonlinear Balance, Step Down. Dotted: the Linear Model of Equation (1.5), Which Predicts $\pm 0.20\,\mathrm{m}$.

1.4.3 Linearization and Deviation Variables

Equation (1.2) is nonlinear because of the square root, and Equation (1.3) because of the products of L with $\mathrm{d}T/\mathrm{d}t$ and of F with T. A model valid near the operating point follows from a first-order Taylor expansion. For a state equation $\dot{x} = f(x,u)$ that is at rest at $(\bar{x},\bar{u})$, the expansion is Equation (1.4).

$$\frac{\mathrm{d}x'}{\mathrm{d}t} \approx \left.\frac{\partial f}{\partial x}\right|_{\bar{x},\bar{u}} x' + \left.\frac{\partial f}{\partial u}\right|_{\bar{x},\bar{u}} u', \qquad x' = x - \bar{x}, \quad u' = u - \bar{u} \tag{1.4}$$

The constant term $f(\bar{x},\bar{u})$ is zero at a steady state, which is why deviation variables remove it. Applied to Equation (1.2), the slope of $\beta\sqrt{L}$ is $\beta/(2\sqrt{\bar{L}}) = \bar{F}/(2\bar{L})$, and the linear model is Equation (1.5).

$$\tau_L \frac{\mathrm{d}L'}{\mathrm{d}t} = -L' + K_L F', \qquad \tau_L = \frac{2A_1\bar{L}}{\bar{F}}, \qquad K_L = \frac{2\bar{L}}{\bar{F}} \tag{1.5}$$

At the operating point $\tau_L = 10\,\mathrm{min}$ and $K_L = 10\,\mathrm{m}$ per $\mathrm{m}^3/\mathrm{min}$. Both constants depend on where the tank operates. At $\bar{L} = 0.25\,\mathrm{m}$ they are 5 and 5, and at 4 m they are 20 and 20. Their ratio $K_L/\tau_L = 1/A_1$ does not change, because the first slope of a level response is the inflow change divided by the area, whatever the head. Chapter 2 derives the same model for its tank and finds that the loop gain of a fixed controller grows 3.46 times from the bottom of the tank to the top.

How far does the linear model reach? Figure 1.3 steps the feed by $\pm 0.02\,\mathrm{m}^3/\mathrm{min}$, a tenth of $\bar{F}$. The linear model predicts $\pm 0.20\,\mathrm{m}$. The nonlinear balance settles 0.21 m higher after the rise and 0.19 m lower after the fall, and it takes 10.69 min and 9.32 min to cover 63.2 % of the change against the 10 min of τ_L. The tank slows as it fills. A step test therefore returns the gain and time constant of the range it crosses, not of its starting point. Chapter 3 finds a fitted gain 14 % above the linearized one for this reason.

The energy balance has two inputs. At a steady state $\mathrm{d}T/\mathrm{d}t = 0$, so the factor L on the left of Equation (1.3) enters its linearization only as $\bar{L}$, and the result is Equation (1.6).

$$\tau_T \frac{\mathrm{d}T'}{\mathrm{d}t} = -T' + K_W W' + K_F F', \qquad \tau_T = \frac{A_1\bar{L}}{\bar{F}}, \quad K_F = -\frac{\bar{T} - T_{in}}{\bar{F}} \tag{1.6}$$

with $\tau_T = 5\,\mathrm{min}$, $K_W = 0.2\,^\circ\mathrm{C}/\%$ and $K_F = -50\,^\circ\mathrm{C}$ per $\mathrm{m}^3/\mathrm{min}$. The time constant is the residence time of the water, and more feed cools the tank. A plant with several states is linearized the same way, one partial derivative per pair, which gives the state-space model and its transfer function of Equation (1.7).

$$\frac{\mathrm{d}x'}{\mathrm{d}t} = \mathbf{A}x' + \mathbf{B}u', \qquad y' = \mathbf{C}x', \qquad G(s) = \frac{Y'(s)}{U'(s)} = \mathbf{C}\,(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} \tag{1.7}$$

where $\mathbf{A} = \partial f/\partial x$ and $\mathbf{B} = \partial f/\partial u$ at the operating point. The eigenvalues of $\mathbf{A}$ are the poles of G, and each real pole $-1/\tau$ contributes one time constant. For the level, $G(s) = K_L/(\tau_L s + 1)$.

Chapter 3 writes Equation (1.7) for two tanks and reads both poles off the diagonal. Chapter 5 linearizes its ball about the hover angle, and Chapter 7 linearizes the water column to find a zero that limits the governor. Chapter 10 computes the four eigenvalues of its market and finds that they differ on each side of the balance, so one linearization holds on one branch at a time. A gain that moves with the operating point is the reason for gain scheduling in Volume 2 and for the nonlinear methods of Volume 7.

1.4.4 First-Order and Integrating Responses

After a step Δu at $t = 0$, the model $\tau_p \dot{y}' = -y' + K_p u'$ responds as Equation (1.8).

$$y'(t) = K_p \Delta u \left(1 - \mathrm{e}^{-t/\tau_p}\right) \tag{1.8}$$

The response covers 63.2 % of its change at $t = \tau_p$ and 95 % at $3\tau_p$. It starts at its steepest, with the slope $K_p \Delta u/\tau_p$, and has no inflection. TK-1 after a feed step of $0.02\,\mathrm{m}^3$/min starts to rise at 0.02 m/min. A plant of this kind is self-regulating: its own outflow grows with the level until it matches the new feed.

Replace the orifice with a pump that draws a fixed F_p, and the outflow no longer depends on the level. Equation (1.2) becomes Equation (1.9).

$$A_1 \frac{\mathrm{d}L}{\mathrm{d}t} = F - F_p, \qquad G(s) = \frac{K'}{s}, \qquad K' = \frac{1}{A_1} \tag{1.9}$$

The tank is now an integrating process. The same feed step makes the level rise at 0.02 m/min until it overflows or an operator acts, and no steady state exists unless $F = F_p$. The gain of an integrating process is a slope, in PV units per unit time per MV unit, and its step test has no final value to read. The integrating gain $K' = 1/A_1$ is the ratio K_L/τ_L of the self-regulating tank, so the integrator is the limit of a first-order lag whose time constant grows without bound.

Over a short test the two are hard to tell apart. One minute after the step TK-1 has risen 0.0190 m against the integrator's 0.0200 m, 4.8 % less. After five minutes the gap is 21 %. Chapter 9 makes the point on a full plant. Its pool temperature has a time constant of 3.3 h while the cooling trains run and 339 h when they stop, and a test read over its first hour returns an integrating gain that the chapter calls a correct measurement of the wrong model.

Integrating inventories run through the volume. Chapter 2 closes the outlet valve of its tank and obtains Equation (1.9) exactly. The pool level of Chapter 9, the water inventory of Chapter 6, the height of the ball in Chapter 5 and the stocks of Chapter 10 all integrate a flow imbalance. In Chapter 8 the frequency integrates the power imbalance, held back only by load damping, so a finite droop gain leaves an offset that only integral control removes.

1.4.5 Second-Order Responses

Let TK-1 drain freely into a second tank TK-2 with $A_2 = 0.5\,\mathrm{m}^2$ and the same orifice, also at 1 m. The outflow of TK-1 does not depend on the level of TK-2, so the tanks are non-interacting and the transfer function from the feed to the lower level is the product of two first-order lags, Equation (1.10).

$$\frac{L_2'(s)}{F'(s)} = \frac{K_L}{(\tau_1 s + 1)(\tau_2 s + 1)}, \qquad \tau_1 = 10\,\mathrm{min}, \quad \tau_2 = 5\,\mathrm{min} \tag{1.10}$$

The response now starts with zero slope and has an inflection, the mark of more than one lag. Any second-order model can be written in the standard form of Equation (1.11).

$$G(s) = \frac{K_p}{\tau_n^2 s^2 + 2\zeta\tau_n s + 1}, \qquad \tau_n = \sqrt{\tau_1 \tau_2}, \quad \zeta = \frac{\tau_1 + \tau_2}{2\sqrt{\tau_1 \tau_2}} \tag{1.11}$$

where τ_n is the natural period and ζ the damping factor. The expressions on the right hold for two real lags. For the two tanks $\tau_n = 7.07\,\text{min}$ and $\zeta = 1.06$. Two real lags always give $\zeta \geq 1$, so tanks in series never overshoot.

A response that rings needs $\zeta < 1$, which takes two stores that trade their contents, such as a mass on a spring or rotating inertia against a restoring force. The overshoot of a step response is then Equation (1.12).

$$\text{overshoot} = \exp\left(-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}\right), \qquad \text{period} = \frac{2\pi\tau_n}{\sqrt{1-\zeta^2}} \tag{1.12}$$

With $\zeta = 0.3$ and the same τ_n, the overshoot is 37.2 % and the period 46.6 min, as panel (b) of Figure 1.4 shows.

Chapter 3 is Equation (1.10) with a splitter that sends part of the feed past the upper tank. The splitter scales the upper time constant by the share of flow it still carries and adds a zero, so the same vessels range from second order to first order. Chapter 4 has four lags, two heaters and two sensors, and the sensor lag of 140 s dominates its step test. Chapter 8 has three lags in its primary loop and turns oscillatory as the droop gain rises. Chapter 10 finds a damped pair of eigenvalues near $-3.85 \pm 3.47\,j$ per year beside two slow modes, and Chapter 12 offers the second-order family with $\zeta < 1$ for a record that rings.

1.4.6 Dead Time and Lags That Look Like It

The water seen by TT-1 left the tank one pipe volume earlier. With plug flow the measurement is the tank temperature delayed by the transport time of Equation (1.13).

$$T_m(t) = T(t-\theta), \qquad \theta = \frac{V_p}{\bar{F}} = 1\,\text{min}, \qquad \mathscr{L}\{T'(t-\theta)\} = \mathrm{e}^{-\theta s}\,T'(s) \tag{1.13}$$

where V_p is the pipe volume. A dead time is an interval during which the output shows nothing of a change already made. At half the feed it doubles to 2 min, a variation that Volume 5 treats. The thermowell adds a lag of $\tau_s = 1\,\text{min}$, so the heater reaches the reading through the model of Equation (1.14).

$$\frac{T_m'(s)}{W'(s)} = \frac{K_W\,\mathrm{e}^{-\theta s}}{(\tau_T s+1)(\tau_s s+1)} \tag{1.14}$$

This is a second-order-plus-dead-time (SOPDT) model. Most tuning rules want the simpler first-order-plus-dead-time (FOPDT) model of Equation (1.15).

$$G(s) = \frac{K_p\,\mathrm{e}^{-\theta_p s}}{\tau_p s+1} \tag{1.15}$$

A small lag in series with a large one delays the response much as a dead time does, so the reduction folds the small lags into θ_p. The half rule [17] gives the largest lag half of the second largest and the dead time the other half, as shown in Equation (1.16).

$$\tau_p = \tau_1 + \frac{\tau_2}{2}, \qquad \theta_p = \theta + \frac{\tau_2}{2} + \sum_{i\geq 3}\tau_i \tag{1.16}$$

For the heater path the rule gives $\tau_p = 5.5\,\text{min}$ and $\theta_p = 1.5\,\text{min}$. For the two tanks it gives 12.5 and 2.5 min, while the two-point fit of Section 1.4.8 on their exact response gives 12.4 and 3.5 min. An effective dead time depends on the method that reads it.

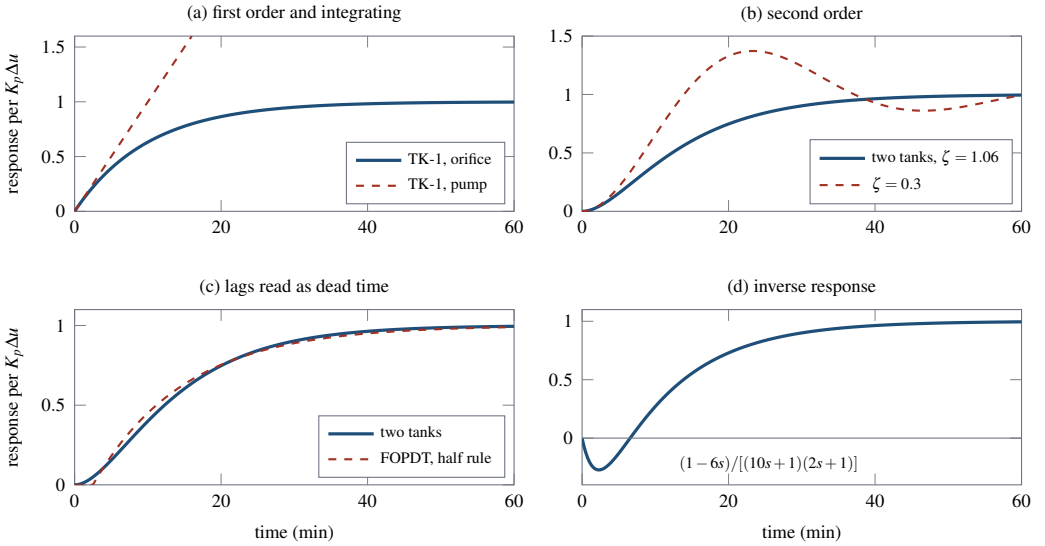


Figure 1.4 Response Shapes of the Running Example, Scaled by $K_p\Delta u$. (a) TK-1 With Its Orifice (Solid) and With a Pump on the Outlet (Dashed), Whose Ramp Reaches 1 at $t = \tau_L$. (b) Two Tanks in Series (Solid) and an Underdamped Plant With the Same τ_n (Dashed). (c) Two Tanks (Solid) and Their FOPDT by the Half Rule (Dashed). (d) Two Opposing Paths With an Inverse Response.

> **DID YOU KNOW? One Delay in Nerves, Networks and Pipes**
>
> A standard text names nerve propagation, communication systems and mass transport as typical sources of time delay, each with the transfer function $e^{-s\tau}$ [2]. The outlet pipe of TK-1 is mass transport, and the router of Volume 2, Chapter 7 carries a dead time equal to the round trip of its path.

Effective dead times appear in most chapters. In Chapter 4 the fitted FOPDT has $\tau_p = 141\,\text{s}$, the sensor lag, and $\theta_p = 18\,\text{s}$, the heater dynamics folded into a delay. Chapter 9 reads an apparent dead time of 0.16 h from a step whose transport delay is 0.06 h, and accounts for the rest as two small lags. The apparent dead time of Chapter 3 falls from 18.5 s to below zero as the splitter opens and the second lag leaves. Chapter 5 lumps a sample period, a servo lag and a flow lag into one dead time of 88 ms.

1.4.7 Inverse Response

TK-1 has no input that moves its level or its temperature the wrong way first. A plant does when one input acts through two paths of opposite sign and different speed, as in Equation (1.17).

$$G(s) = \frac{K_1}{\tau_1 s + 1} - \frac{K_2}{\tau_2 s + 1} = \frac{(K_1 - K_2) + (K_1\tau_2 - K_2\tau_1)\,s}{(\tau_1 s + 1)(\tau_2 s + 1)} \tag{1.17}$$

The final gain is $K_1 - K_2$, set by the slow path. The first slope is $K_1/\tau_1 - K_2/\tau_2$, set mostly by the fast path. When the fast path's initial slope is the larger, the response starts in the wrong direction and the numerator has a zero in the right half of the plane [9]. With the time scales of TK-1, $K_1 = 2$, $\tau_1 = 10\,\text{min}$, $K_2 = 1$ and $\tau_2 = 2\,\text{min}$ give $G = (1-6s)/[(10s+1)(2s+1)]$. Its step response falls to -0.27 at 2.29 min, crosses zero at 6.56 min and settles at $+1$, as panel (d) of Figure 1.4 shows.

Chapter 7 has the classical case. Opening the gate of a water turbine admits more flow, but the water in the penstock must first be accelerated by head taken from the turbine. The ideal model is $(1 - T_w s)/(1 + T_w s/2)$, where T_w is the water starting time. It equals $-2 + 3/(1 + T_w s/2)$, an instant

path of gain -2 in parallel with a lag of gain $+3$, so the power first moves twice as far the wrong way as it moves in the end. On the chapter's plant the first response is 1.71 to 1.94 times the final one, and the ratio of the water starting time to the mechanical starting time caps the governor's gain. Chapter 6 has the other classical case. A rise in steam demand lowers the pressure, the water boils and its bubbles lift the surface three points while the vessel loses water. A zero in the right half-plane limits how fast any controller can act, a limit that Volume 7 develops.

1.4.8 Step Tests and Regression

A balance gives the form of a model. Its numbers often come from a test. The plant is held steady in Manual, the MV is stepped and the PV is recorded until it settles. The step must stand clear of the noise, stay small enough for the linear model to hold and last long enough to settle, which for the pool of Chapter 9 means most of a day.

> **+ DID YOU KNOW? A Plant Test of 5 to 15 Days**
>
> In a survey of industrial predictive control, a formal plant test runs 24 hours a day with engineers watching. Each manipulated variable is stepped 8 to 15 times, with a signal-to-noise ratio of at least six, and the test may take up to 5 to 15 days [12].

TK-1's heater test steps W from 50 to 60 % at $t = 2\,\text{min}$. TT-1 is sampled every $T_s = 0.2\,\text{min}$ for 40 min, 201 samples, with measurement noise of standard deviation 0.03 °C. The plant is Equation (1.14) exactly, so every fitted model can be compared with the truth.

The two-point method reads a FOPDT model from two times, as shown in Equation (1.18).

$$K_p = \frac{\Delta y}{\Delta u}, \qquad \tau_p = 1.5\,(t_{63} - t_{28}), \qquad \theta_p = t_{63} - \tau_p \tag{1.18}$$

where t_{28} and t_{63} are the times after the step at which the PV has covered 28.3 and 63.2 % of its change. A first-order response reaches those levels $\tau_p/3$ and τ_p after it starts to move. On the record, with the final value taken as the mean of the last five minutes, $K_p = 0.200\,^\circ\text{C}/\%$, $t_{28} = 3.71\,\text{min}$ and $t_{63} = 7.11\,\text{min}$, so $\tau_p = 5.11\,\text{min}$ and $\theta_p = 2.00\,\text{min}$. The method uses two samples of 201, and the noise on those two moves the result.

Least squares uses every sample. The parameters ϑ of a model are chosen to minimize the sum of squared residuals of Equation (1.19) [10].

$$J(\vartheta) = \sum_{k=1}^{n} \varepsilon_k^2, \qquad \varepsilon_k = y_k - \hat{y}_k(\vartheta) \tag{1.19}$$

where n is the number of samples and $\hat{y}_k$ is the model's output computed from the input record alone. The FOPDT model fitted this way has $K_p = 0.201$, $\tau_p = 5.18\,\text{min}$ and $\theta_p = 1.86\,\text{min}$. The SOPDT model has $K_p = 0.200$, lags of 4.76 and 1.20 min and a dead time of 0.94 min, close to the true 5, 1 and 1 min. Both models are nonlinear in their time constants and dead time, so J is minimized by an iterative search. Chapter 11 uses the Nelder–Mead simplex method [16] and minimizes the sum of absolute errors in place of squares, and Chapter 12 uses a Levenberg–Marquardt search [15].

A sampled model avoids the search. When the input is held between samples, a first-order model with a dead time of d samples is exactly the difference equation of Equation (1.20).

$$y'_k = \phi\, y'_{k-1} + \gamma u'_{k-1-d}, \qquad \phi = \mathrm{e}^{-T_s/\tau_p}, \qquad \gamma = K_p\,(1 - \phi) \tag{1.20}$$

With the measured y'_{k-1} on the right, ϕ and γ enter linearly, and one linear least-squares solve gives both. This is the simplest autoregressive model with an exogenous input (ARX). The continuous constants follow from Equation (1.21).

$$\tau_p = -\frac{T_s}{\ln \phi}, \qquad K_p = \frac{\gamma}{1 - \phi} \tag{1.21}$$

Table 1.1 Models Fitted to the Heater Step Test of TK-1. Errors Are Root Mean Square Residuals of a Simulation From the Input, in °C. The Noise Has a Standard Deviation of 0.030 °C.

Model	Parameters (K_p in °C/%, times in min)	Training	Validation	κ_1	Runs
FOPDT, two-point	0.200, τ_p 5.11, θ_p 2.00	0.0372	0.0484	0.217	85
FOPDT, least squares	0.201, τ_p 5.18, θ_p 1.86	0.0340	0.0429	0.191	77
SOPDT, least squares	0.200, lags 4.76 and 1.20, θ_p 0.94	0.0311	0.0331	0.035	105
ARX, $d = 9$	0.199, τ_p 5.04, θ_p 1.8	0.0376	0.0508	0.306	72
True model	0.2, lags 5 and 1, θ_p 1	0.0319	0.0317	0.061	89

Scanning d and solving at each gives $d = 9$ samples, or 1.8 min, with $\tau_p = 5.04$ min and $K_p = 0.199$. The ARX fit minimizes the error of a one-step prediction made from the last measurement, which is a different objective from the error of a simulation run from the input alone. Table 1.1 shows the cost: run as a simulation, this model has the largest validation error of the four. Chapter 12 fits ARX models by least squares and then refines them on the simulation error for this reason. It also fits continuous transfer functions to the same sampled records [18].

1.4.9 Model Structure, Validation and Residuals

Table 1.1 compares the fitted models on two records. The training record is the step test. The validation record is a second test with fresh noise, in which W steps from 50 to 60, then to 45 and then to 55 %, at 2, 14 and 26 min. Each model is fitted once to the training record and then run unchanged on the validation record.

The training errors differ little, from 0.0311 to 0.0376 °C against a noise of 0.030. A small training error says only that the model can follow the record it was fitted to, and any model with enough parameters can. The validation column separates the models more clearly. The SOPDT model keeps its error at 0.0331 °C, near the true model's 0.0317, while the FOPDT fitted by least squares rises to 0.0429 °C on the new steps.

The residuals show why. If a model has the right structure, its residuals are the measurement noise, which is independent from sample to sample. If it does not, the residuals carry the part of the response the model cannot follow, which changes slowly. Figure 1.5 plots both sets with their one-minute moving mean. The FOPDT residuals run positive at the start of the response, with a mean of +0.030 °C from 2 to 4 min, negative at the shoulder, −0.024 from 6 to 9 min, and positive again, +0.016 from 14 to 20 min. A single lag cannot bend the way two lags do. The SOPDT residuals stay within 0.01 °C of zero over the same windows.

Two numbers test the pattern without a plot. The lag-one autocorrelation of the residuals, Equation (1.22), is near zero for independent noise.

$$\kappa_1 = \frac{\sum_{k=2}^{n} \varepsilon_k \, \varepsilon_{k-1}}{\sum_{k=1}^{n} \varepsilon_k^2} \tag{1.22}$$

For $n = 201$, noise alone keeps κ_1 within $\pm 2/\sqrt{n} = \pm 0.141$ about 19 times in 20. The FOPDT gives 0.191 and the ARX model 0.306, outside the band, and the SOPDT gives 0.035, inside it. The second number counts runs, the stretches of residuals of one sign. Independent noise gives about $(n+1)/2 = 101$ runs with a standard deviation of about 7. The FOPDT residuals form 77 and the SOPDT residuals 105. The Ljung–Box portmanteau test extends the first check to many lags at once [19].

More parameters always lower the training error, so the comparison must charge for them. The Akaike information criterion [11], for residuals that are independent and normal, is Equation (1.23).

$$\mathrm{AIC} = n \ln \frac{J}{n} + 2m \tag{1.23}$$

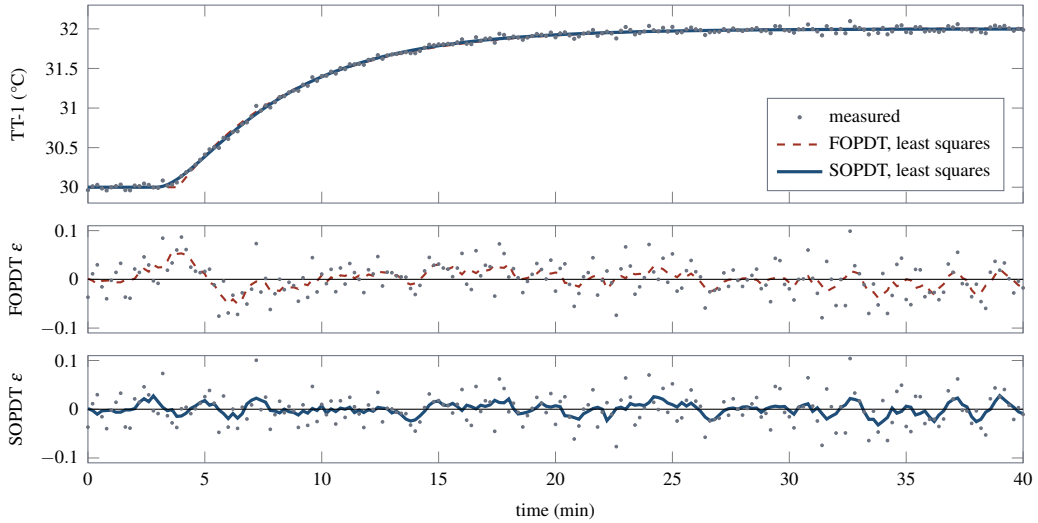


Figure 1.5 Heater Step Test on TK-1 (W From 50 to 60 % at 2 min) With Two Least-Squares Fits. Top: Measurements (Dots), FOPDT (Dashed) and SOPDT (Solid). Middle and Bottom: Residuals in °C (Dots) and Their One-Minute Moving Mean (Dashed for the FOPDT, Solid for the SOPDT).

where m is the number of fitted parameters. The lower value wins. The FOPDT, with $m = 3$, scores -1352.9, and the SOPDT, with $m = 4$, scores -1387.0. The fourth parameter more than pays for itself. A criterion is a guide and not a verdict. Validation on data the fit has not seen is the stronger test, and Chapter 12 builds its workflow around it, with chronological training, validation and final-test intervals and the test scores hidden until a model is chosen.

> ✦ DID YOU KNOW? **A Fit That Beats the Truth on Its Own Record**
>
> From the simulator: on the training record of Table 1.1, the SOPDT model fitted by least squares scores 0.0311, below the 0.0319 of the true model that made the data. On the validation record the order turns round, 0.0331 against 0.0317. Part of what the fit learned was noise.

None of this makes the FOPDT useless. Its validation error is 0.043 °C on steps of 2 °C, and it is the model that the tuning rules of Volume 2 consume. The residual tests say that its structure is incomplete, and how much a purpose can tolerate is a separate question. Chapter 11 shows the opposite error. Adding radiation to an energy balance lowers the fitting objective on both of its step tests, by 5.4 and 11.8 %, yet the same model loses by 10 to 26 % on three other runs and claims that 70.1 % of the heat left by radiation in data generated with none.

A test can only determine what it excites. The heater step fixes τ_T and K_W, but $\tau_T = A_1\bar{L}/\bar{F}$ and K_W depends only on $\rho c_p \bar{F}$. Doubling A_1 while halving $\bar{L}$, or doubling ρ while halving c_p, leaves the record unchanged. Only the groups are identifiable [20]. Chapter 11 demonstrates exactly this: doubling the area and halving the heat transfer coefficient of its convective model leaves the objective unchanged in every displayed digit. Chapter 12 faces the multivariable form of the problem, where two inputs that moved together cannot be credited separately.

The chapters keep three models apart. The simulation model is the full balance, nonlinear and exact within its assumptions. The identified model is a fitted FOPDT or transfer function, good near one operating point. The controller's model is whatever the controller holds, which for a PID is nothing but its settings and for a predictive controller is one identified model, often fitted at a single point. Chapter 2 and Chapter 3 name the three explicitly. Chapter 13 turns residuals into diagnosis: a balance that does not close points to a meter in error, and a statistic that leaves its limits flags a fault.

Table 1.2 The Inventory, Model Form and Source of the Model in Each Chapter

Ch.	Inventory	Model form	How the chapter obtains it
2	water, one tank	first order with K_p and τ_p both $\propto \sqrt{L}$, integrating with the outlet shut	balance linearized, step tests and a least-squares time constant
3	water, two tanks	second order with a zero, first order with the splitter open	balances linearized, two-point fits at three splitter positions
4	heat, two heaters	four lags, 2×2 gain matrix, FOPDT for tuning	energy balances, manual step tests
5	momentum of a ball	integrator with a lag and dead time, gain set by the valve map	balance about hover, drop and step tests
6	water and energy	integrator with swell and shrink	balances with steam tables, open-loop cases
7	water momentum, rotor energy	right-half-plane zero, oscillatory modes	balances linearized, computed stability map
8	rotating energy	third order in the primary loop, offset set by stiffness	swing equation and closed-form analysis
9	water and heat	level integrating, temperature FOPDT or integrating	balances and one step test
10	oil stocks	damped pair and slow modes, two gains at two speeds	balances linearized, eigenvalues, steps
11	heat, one heater	three nested energy balances	fitted to records, checked by cross-validation
12	any recorded plant	FOPDT, SOPDT, ARX, FIR and state space for several inputs	fitted to records with held-out intervals
13	plant records	balances and statistical models of normal data	reconciliation, principal components, charts

1.4.10 Which Model Form a Plant Needs

Three kinds of model appear in the chapters. A model from balances alone uses physical constants and needs no test. A model from a test alone fits a standard form, FOPDT or ARX, without asking why the plant has that form. A model between the two takes its form from the balances and its constants from data, as in Chapter 11. Table 1.2 lists, for each chapter, the inventory that sets the dynamics, the form of model the chapter works with and how the chapter obtains it. The final chapter compares the plants along these columns, with the measured results of each.

1.5 Instruments and Final Elements in This Volume

A model is fitted to what an instrument reports, not to the plant itself. Figure 1.6 lists what the model of each chapter reads and what it moves. Level and temperature are read most often, by four chapters each. Speed or frequency serves the two power chapters, and three chapters work on a stored record in place of a live signal. The final elements run from a control valve and a pump to a heater, a turbine gate and a production quota, and Chapter 13 moves nothing at all.

A thermowell protects a temperature element and adds a first-order lag through its own mass and heat capacity. An air gap between the element and the bottom of the well adds a second lag, and such lags can compromise the stability of a feedback loop [21]. The thermowell of TT-1 is the lag τ_s of Equation (1.14), drawn in Figure 1.7, and the sensor lag of Chapter 4 sets the time constant fitted to its step test.

A pressure transmitter at the floor of a vessel infers the level from the weight of the column above it. The reading gives the height only while the density of the liquid is known and constant, although in a vessel of constant cross-section it measures the mass of liquid whatever its density [21]. In Chapter 6 such a reading sits closer to the chapter's mass-equivalent level than to the boiling surface. Figure 1.8 shows the two taps.

Three chapters start from records. A data historian polls process data from the control system at regular intervals and keeps it over long periods, typically years [21], and a standards glossary

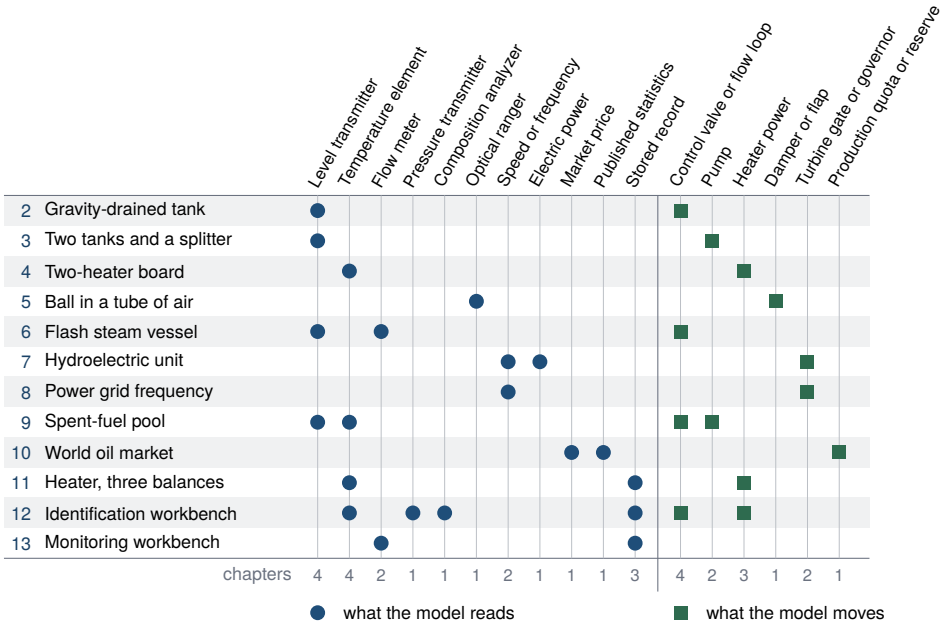


Figure 1.6 What the model of each chapter reads and moves: a round mark where the chapter's model reads the source and a square mark where it moves the element, as its variables table and its technology section name them. A chapter that works on stored records is also marked under the instruments its records hold. Notice that level, temperature and the control valve lead the counts, with four chapters each.

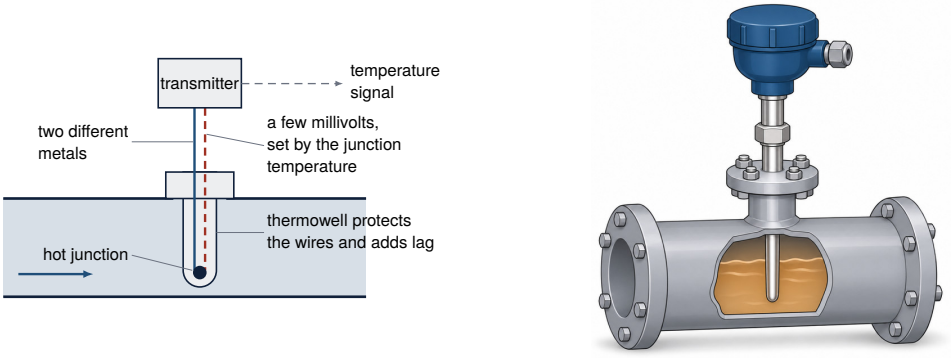


Figure 1.7 Principle of a thermocouple in a thermowell. Notice that the heat must cross the wall of the well before it reaches the junction, which is why TT-1 reads the outlet through a lag of its own (Section 1.4.6).

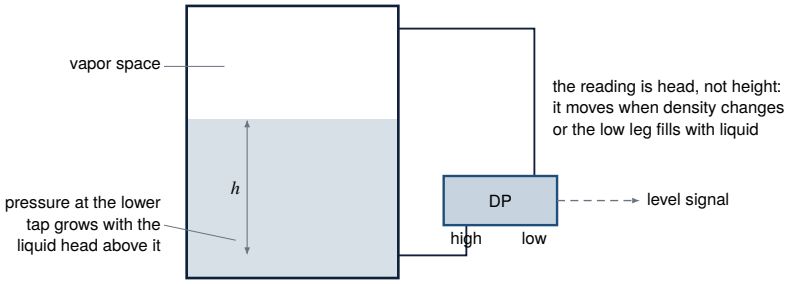


Figure 1.8 Level by differential pressure. Notice that the transmitter weighs the liquid above its lower tap, so it reads a height only while the density is known (Chapter 2 and Chapter 6).

Layer	What it does in this volume	Where it runs	How often
People who decide	engineers step the plant in Manual for a test, and a conference of governments sets a quota	**By hand**, in a plant test or at a meeting	a test of up to 5 to 15 days; meetings at least twice a year
Supervisory and predictive	predictive control holds an identified model, and generation control restores the grid	**Supervisory level** of a distributed control system, or an energy management system	0.1 s to 1 min in the apps
PID loop	the velocity-form PID of each chapter, with limits and anti-windup	**Distributed control system**, PLC, or a browser page beside a laboratory board	20 ms to 0.01 h; weekly in the oil market
Inner loop	a feed flow loop under level control, a gate servomotor, a flap servo	**Flow controller**, positioner or servo drive	0.25 s in the flash vessel
Sensors, final elements, records	transmitters, rangers and statistics; valves, pumps, heaters and gates; the archive of readings	wires, a serial link or a network, and a **data historian**	records from every 0.1 s to every 15 min, kept for years

Figure 1.9 The control layers of this volume, the platform that hosts each and how often each acts. Platforms are from [12], [22] and [21], and the app periods from the chapters' technology sections. Notice that the historian at the bottom keeps the records from which the models of Section 1.4.8 are fitted.

describes it as a central database that supports analysis by statistical process control [22]. That archive is the raw material of Chapter 11, Chapter 12 and Chapter 13.

Predictive control sits one level above the PID control of valves [12], and it holds an identified model in place of three settings. A test for such a model must excite both the steady state and the dynamics of the process [12]. Loops at every level run with cycle times from milliseconds to minutes [22]. Figure 1.9 places the loops of this volume on their platforms. The apps go from the 20 ms governor of Chapter 7 to the oil market of Chapter 10, whose quota changes only at meetings held at least twice a year.

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 1.3 lists what a reader meets.

Table 1.3 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.

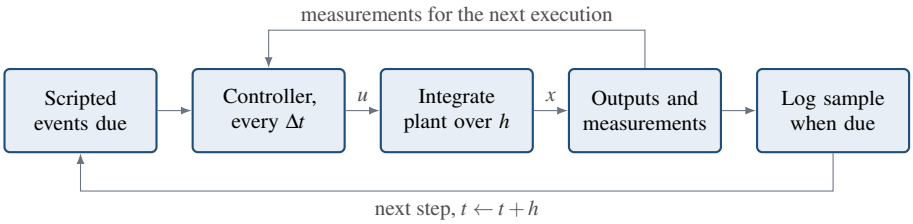


Figure 1.10 Work of the Engine in One Integration Step

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.

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 (1.24) to (1.27) for execution k and control period Δt.

$$e_k = a\,(\mathrm{SP}_k - \mathrm{PV}_k), \qquad P_k = K_c\,e_k, \qquad D_k = -a\,K_c\,\tau_D\,\frac{\mathrm{PV}^f_k - \mathrm{PV}^f_{k-1}}{\Delta t} \tag{1.24}$$

$$S_k = S_{k-1} + (P_k - P_{k-1}) + (D_k - D_{k-1}) \tag{1.25}$$

$$v_k = u_b + S_k + I_{k-1}, \qquad u_k = \mathrm{sat}(v_k) \tag{1.26}$$

$$I_k = I_{k-1} + \left[\frac{K_c}{\tau_I}\,e_k + \frac{u_k - v_k}{T_t}\right]\Delta t \tag{1.27}$$

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 (1.27) drives I toward the value that reproduces the output applied, which is back-calculation [23]. With no integral action, Equation (1.27) 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 $I_k = u_{\mathrm{man}} - u_b - S_k$, 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 (1.28).

$$x_{k+1} = A\,x_k + B\,u_k + E\,w_k + c, \qquad y_k = C\,x_k \tag{1.28}$$

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 (1.29) over a horizon of N steps.

$$J = \sum_{k=1}^{N} (y_k - r_k)^{\top} Q (y_k - r_k) + \sum_{j} \Delta u_j^{\top} R \, \Delta u_j + \sum_{i} \rho_i s_i^2 \tag{1.29}$$

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 [24]. 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 [25]. 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 [26].

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.

- *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 14.

1.8 Chapter Roadmap

Table 1.4 lists the chapters with the modeling 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. Figure 1.11 places the same chapters by where their models come from and by the response their lessons center on.

	From balances alone	Form from balances, constants from tests	From records alone
First order and integrating		2 gain moves with level 5 an integrator behind a valve 9 one pool, two plants	
Second and higher order	8 inertia and droop offset 10 two gains at two speeds	3 order set by piping	
Lags read as dead time		4 sensor lag dominates	
Inverse response	6 swell and shrink 7 a zero from water inertia		
Structure and validation		11 three nested balances	12 scored on held-out data 13 residuals flag faults

Figure 1.11 Method map of the volume. Each chapter is placed by how it obtains its model, from the last column of Table 1.2, and by the response its lesson centers on, with that lesson shortened from Table 1.4. The placement is the editor's reading of the two tables. Notice that most chapters take the form of the model from a balance and its numbers from a test.

Table 1.4 Roadmap of the Volume

Ch.	Plant	The modeling lesson	Level	Time
2	Gravity-drained tank	Linearization: gain and time constant both follow the square root of the level. The integrating limit.	introductory	2.5 h
3	Two tanks and a splitter valve	Tanks in series, plant order set by piping, the apparent dead time of a second lag.	introductory	3 h
4	Two-heater laboratory board	Energy balances, a gain matrix and a step test dominated by the sensor.	introductory	3 h
5	Ball in a tube of air	An integrator behind a valve whose gain changes thirtyfold.	intermediate	2.5 h
6	Flash steam vessel	Swell and shrink: a level that moves against its inventory.	advanced	3 h
7	Hydroelectric unit	A right-half-plane zero from water inertia and the gain it allows.	advanced	2 h
8	Power grid frequency	An energy inventory: inertia, droop offset and integral restoration.	intermediate	2.5 h
9	Spent-fuel pool	One measurement, two plants: self-regulating or integrating with the pumps.	intermediate	3 h
10	World oil market	A tank of stocks with a short-run and a long-run gain.	intermediate	2.5 h
11	Heater with three balances	Model structure, identifiability and cross-validation.	intermediate	3 h
12	Identification workbench	Multivariable models from records, scored on held-out data.	advanced	3 h
13	Monitoring workbench	Residuals of balances and statistics as evidence of faults.	advanced	2 h

Chapters 2 to 4 form a first course and are best read in order: one tank, two tanks, then heat in place of water. Chapter 5 and Chapters 8 to 10 each study one inventory in depth and can be read in any order after Chapter 2. Chapters 6 and 7 need Section 1.4.7. Chapters 11 and 12 need Sections 1.4.8 and 1.4.9, and Chapter 11 is best read after Chapter 4, whose board it models. Chapter 13 uses records from plants across the series and is best read last. A reader who has taken a first course in process dynamics can take a shorter path through Chapters 3, 6, 9, 11 and 12.

KEY POINTS

- A model comes from balances, from a test or from both. The balance gives its form and how its constants move, and the test supplies the numbers.
- Gain and time constant follow the operating point, so a step test returns the values of the range it crosses.
- Small lags in series look like dead time, and an effective dead time depends on the method that reads it.
- A small training error says only that a model can follow its own record. Validation on new data and residuals that look like noise are the stronger tests.

Where else this applies. Tuning from these models in Volume 2, dead time in Volume 5, nonlinear models in Volume 7.

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

[3] L. Ljung, "Perspectives on system identification," *Annual Reviews in Control*, vol. 34, no. 1, pp. 1–12, 2010.

[4] O. Tramis, E. Merlin-Anglade, G. Paternoster, M. Rabaud, and N. M. Ribe, "Torricelli's curtain: morphology of horizontal laminar jets under gravity," *Physics of Fluids*, vol. 33, no. 6, p. 062116, 2021.

[5] International Association for the Properties of Water and Steam, "Obsolete IAPWS recommendations," IAPWS, Information document, Mar. 2005. [Online]. Available: https://iapws.org/documents/release/Obsolete.download

[6] W. F. Guthrie, "NIST/SEMATECH e-Handbook of Statistical Methods (NIST Handbook 151)," National Institute of Standards and Technology, 2020, section 6.3.2, What are Variables Control Charts? [Online]. Available: https://www.itl.nist.gov/div898/handbook/

[7] J. G. Ziegler and N. B. Nichols, "Optimum settings for automatic controllers," *Transactions of the ASME*, vol. 64, no. 8, pp. 759–765, 1942.

[8] G. Pillonetto, T. Chen, A. Chiuso, G. De Nicolao, and L. Ljung, *Regularized System Identification: Learning Dynamic Models from Data*, ser. Communications and Control Engineering. Cham: Springer, 2022.

[9] K. Iinoya and R. J. Altpeter, "Inverse response in process control," *Industrial & Engineering Chemistry*, vol. 54, no. 7, pp. 39–43, 1962.

[10] K. J. Åström and P. Eykhoff, "System identification—a survey," *Automatica*, vol. 7, no. 2, pp. 123–162, 1971.

[11] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[12] 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.

[13] Health and Safety Executive, "Better alarm handling," Health and Safety Executive, HSE Information Sheet Chemicals Sheet No 6 (CHIS6), 2000, information current at 3/00. [Online]. Available: https://www.hse.gov.uk/pubns/chis6.pdf

[14] A. Rasheed, O. San, and T. Kvamsdal, "Digital twin: Values, challenges and enablers," arXiv preprint 1910.01719 (version 1); published in revised form as IEEE Access 8, 21980–22012, 2020, 2019.

[15] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[16] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.

[17] S. Skogestad, "Simple analytic rules for model reduction and PID controller tuning," *Journal of Process Control*, vol. 13, no. 4, pp. 291–309, 2003.

[18] H. Garnier, M. Mensler, and A. Richard, "Continuous-time model identification from sampled data: Implementation issues and performance evaluation," *International Journal of Control*, vol. 76, no. 13, pp. 1337–1357, 2003.

[19] G. M. Ljung and G. E. P. Box, "On a measure of lack of fit in time series models," *Biometrika*, vol. 65, no. 2, pp. 297–303, 1978.

[20] R. Bellman and K. J. Åström, "On structural identifiability," *Mathematical Biosciences*, vol. 7, no. 3–4, pp. 329–339, 1970.

[21] 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.

[22] 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, Gaithersburg, MD, Special Publication 800-82r3, 2023.

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

[24] 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.

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

[26] 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 · 5 min

A gravity-drained tank has a cross-section of $2\,\mathrm{m}^2$ and an outflow $\beta\sqrt{L}$ with $\beta = 0.5\,\mathrm{m}^{2.5}/\mathrm{min}$. It operates at a level of 4 m. Find the steady feed, then the gain and the time constant of the linear model from the feed to the level.

1.2 Theory and calculation · 4 min

TK-1 of the running example rests at 1 m with a feed of $0.2\,\mathrm{m}^3/\mathrm{min}$ and $\beta = 0.2\,\mathrm{m}^{2.5}/\mathrm{min}$. The feed is raised to $0.24\,\mathrm{m}^3/\mathrm{min}$. Find the new steady level from the nonlinear balance and the change the linear model predicts. By how many percent does the linear model miss the change?

1.3 Concept check · 2 min

The orifice of TK-1 is replaced by a pump that draws a fixed $0.2\,\mathrm{m}^3/\mathrm{min}$. With the tank at 1 m, the feed is raised from 0.2 to $0.22\,\mathrm{m}^3/\mathrm{min}$ and left there. What does the level do?

(a) It settles 0.21 m higher after about 30 min

(b) It rises at 0.02 m/min until the tank overflows or the feed changes

(c) It rises 0.20 m and then returns to 1 m

(d) It falls first and then rises, because the pump lags the feed

1.4 Theory and calculation · 5 min

A well-mixed tank holds $2\,\mathrm{m}^3$ of water with a throughflow of $0.4\,\mathrm{m}^3/\mathrm{min}$ at a fixed level, and $\rho c_p = 4200\,\mathrm{kJ}/(\mathrm{m}^3\,\mathrm{K})$. Its heater power is raised by 14 kW. Find the time constant of the outlet temperature, its steady change and the time at which it has covered 63.2 % of that change.

1.5 Theory and calculation · 6 min

A plant has the model $G(s) = 3/[(8s+1)(2s+1)]$ with time in minutes.

(a) Write it in the standard second-order form and give τ_n and ζ. Can its step response overshoot? *Checkpoint:* $\tau_n = 4\,\mathrm{min}$, $\zeta = 1.25$, no overshoot.

(b) Reduce it to a first-order-plus-dead-time model with the half rule. *Checkpoint:* $K_p = 3$, $\tau_p = 9\,\text{min}$, $\theta_p = 1\,\text{min}$.

1.6 Theory and calculation · 3 min

The thermometer of a heated tank sits at the end of a pipe that holds $0.3\,\text{m}^3$. Find the dead time at flows of 0.2, 0.1 and $0.3\,\text{m}^3$/min, assuming plug flow.

1.7 Concept check · 3 min

An ideal water turbine has the model $(1 - T_w s)/(1 + T_w s/2)$ from gate to power, with $T_w = 2\,\text{s}$. The gate is opened by a small step. What does the power do, per unit of its final change?

(a) It rises at once to its final value and stays there

(b) It rises with no overshoot and a time constant of 1 s

(c) It overshoots to twice its final value and then settles

(d) It jumps to -2, rises through zero near 1.1 s and settles at $+1$

1.8 Theory and calculation · 4 min

A least-squares fit of the ARX model $y'_k = \phi\, y'_{k-1} + \gamma u'_{k-1-d}$ to a record sampled every $T_s = 1\,\text{min}$ returns $\phi = 0.9$ and $\gamma = 0.03$. Find the time constant and the gain of the equivalent first-order model.

1.9 Theory and calculation · 5 min

Three steady tests on a plant give PV changes of 1.1, 1.9 and 3.1 units for MV changes of 5, 10 and 15 units. Fit the gain of the model $\Delta y = K_p \Delta u$ by least squares and give the three residuals.

1.10 Concept check · 3 min

Two models are fitted to the same step test, with measurement noise of standard deviation $0.030\,^\circ\text{C}$. Model A has a training error of $0.034\,^\circ\text{C}$, a validation error of $0.043\,^\circ\text{C}$ and a lag-one residual autocorrelation of 0.19 over 201 samples. Model B has 0.031, 0.033 and 0.035. Which reading is right?

(a) Model A, because fewer parameters are always preferred

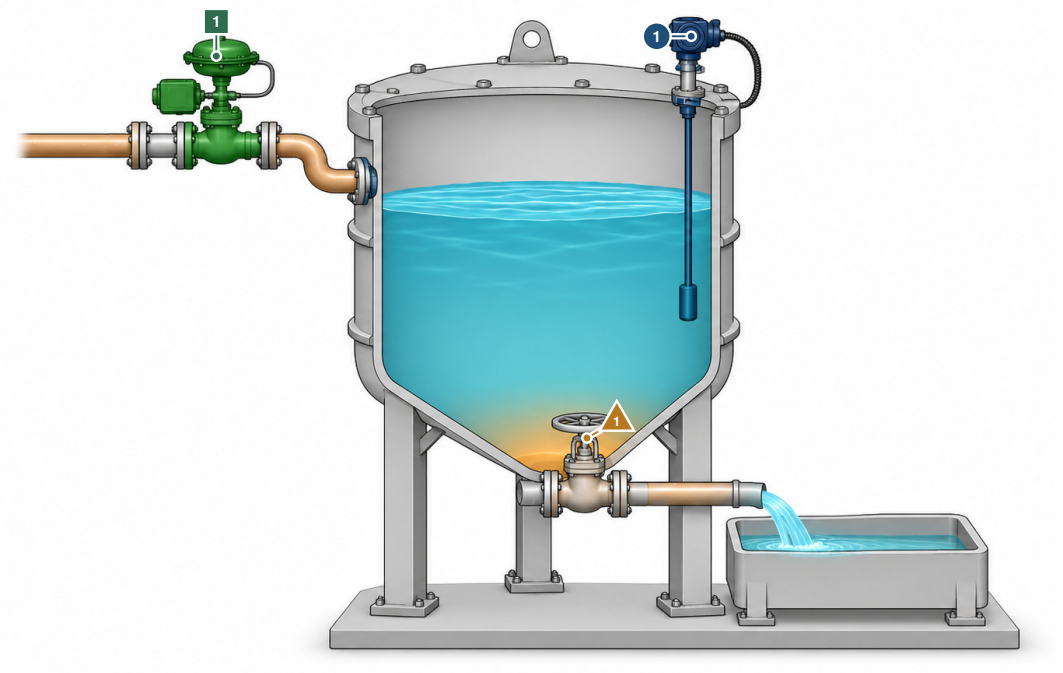
(b) Neither, because the training errors differ by less than the noise

(c) Model B, because its residuals look like noise and it holds on new data

(d) Model A, because a positive autocorrelation shows it follows the dynamics

CHAPTER 2

Gravity-Drained Tank Laboratory for Level Loops With a Moving Time Constant



The controller moves 1 the inlet valve on the supply line

The controller reads 1 the level sensor through the roof

What disturbs it 1 the hand valve on the drain

Notice that the drain sits under the full depth of water, so a fuller tank pushes water out faster, although the model's tank has straight sides and a flat bottom rather than a cone. Illustration, not to scale.

IN PLAIN TERMS **A Tank That Drains Under Its Own Weight**

The main idea. Water flows into a tank through a valve that a controller moves. It drains out through a hand valve at the bottom, and deeper water pushes out faster, so the tank settles at its own level for each inflow.

The challenge. A fuller tank reacts more strongly and more slowly to the same valve move, so settings tested at one depth behave differently at another. With the drain shut, the controller can only fill the tank and never lower it.

Why it matters. A controller that misjudges its tank can overfill it or pass a flow change on to the next unit.

THREE WAYS TO READ THIS CHAPTER

Explore	11 min	the first page, the history of a misread tower level (Section 2.2), the fact on the flowmeter (Section 2.3), the try-it segments of cases 2 and 4 and the fact on derivative action (Section 2.6) and the key points
Study	70 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	30 min	the first page, the results (Section 2.6), Technology in Practice (Section 2.7), Limitations (Section 2.8) and the open challenge

The control question A tank that drains through its own bottom valve is a first-order plant whose gain and time constant both follow the square root of the level, and closing the valve makes it an integrator. What does a step test at one level predict about the loop at another, and what does one proportional loop do on each side?

Learning objectives

- Derive the resting level, the process gain, the time constant and the disturbance gain of the tank from its mass balance.
- Explain why every inlet step starts at the same slope, and measure the time constant by draining the tank.
- Predict the offset of a proportional loop from the bias it carries, and explain why it vanishes when the outlet is shut.
- Compute the valve noise that the velocity-form PID passes from the transmitter, with and without derivative action.
- Explain how a predictive controller with a model five times too sensitive still ends without offset.

Prerequisites Sections 1.4.2 and 1.4.3 (balance and linear model of a drained tank), 1.4.4 (first-order and integrating responses), 1.4.8 (step tests) and 1.6.2 (velocity form). Section 1.6.3 is needed only for the predictive controller and its estimators.

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

See also Volume 2, Chapter 2 for a gain and a time constant that fall together with speed, Volume 2, Chapter 5 for tuning an integrating level loop and Volume 2, Chapter 6 for a gain that moves too far for one tuning. Volume 7, Chapter 3 for a plant that is stable, integrating or unstable according to its operating point.

2.1 Introduction

Liquid level is the most heavily instrumented variable in a process plant and the one most often detuned on purpose. Nearly every vessel carries a level measurement because inventory has to be accounted for, pumps have to be protected against running dry, and tanks have to be protected against spilling, and a large share of those loops are detuned deliberately so that the vessel absorbs a flow variation instead of passing it downstream [1–3]. A first course introduces the subject with the pumped-outflow tank, in which discharge is set by a positive-displacement pump and is independent of the liquid above it. That vessel is a pure integrator. It has no steady state at any inflow other than the exact outflow, it does not return to a level on its own, and proportional feedback holds it at an offset fixed by the difference between the controller bias and the valve position the plant requires. A tank that drains through a valve in its own bottom behaves in the opposite way, because the head driving the discharge is the level being controlled. The vessel reaches its own resting level, and the gain and the time constant of that approach both depend on which level it is.

The two cases appear in the same chapter of every textbook and are almost never placed in the same apparatus [4, 5]. That separation obscures the behavior that governs commissioning. A gravity-drained tank is not one plant with one model. It is a family of first-order plants indexed by the operating level and by the outlet valve position, and the pumped-outflow integrator is the limiting member of that family as the outlet closes. However, the laboratory rigs and simulations used to teach level control usually fix the discharge path, so the transition between the two behaviors is described rather than measured [6]. This chapter works with a gravity-drained tank simulation in which one outlet valve moves the same vessel continuously between the self-regulating and the integrating case. It quantifies by how much a controller commissioned at one level differs from the controller the plant requires at another.

The fundamentals the experiments rest on are the ones a first course establishes, and Chapter 1 develops them on a heated tank that drains through an orifice (Section 1.4.1). A mass balance on a constant-area vessel makes the level the integral of the flow imbalance, which is conservation of mass in its simplest useful form. Bernoulli's principle applied between the free surface and the discharge makes the driving force for outflow the head itself, which is what converts an integrator

into a first-order self-regulating process [4, 5]. The experiments therefore cover integrating dynamics, process order, operating-point-dependent gain, proportional offset and its removal by integral action, the role of the controller bias, anti-windup at an actuator limit, sampled and noisy measurement, derivative action against measurement noise, and open-loop step identification of a model that is only locally valid [7]. The industrial consequence is direct. A level controller tuned at one inventory is the loop that overfills a tank or passes a disturbance downstream at another, and both outcomes have safety and economic weight [3, 8].

Section 2.2 traces level control from the float regulator to averaging level control. Section 2.3 states every relation and constant needed to rebuild the tank from this text alone: the outflow law, the two hard limits, the instrument model and the numerics, specified closely enough that a second implementation reproduces the same trajectories. Section 2.4 gives a closed-form account of the effect of the square-root outflow on a level loop. The identified gain and the identified time constant scale identically, so their ratio is a constant of the vessel, and step tests in the running application confirm it. Section 2.5 describes the velocity-form PID as implemented, with a noise propagation expression checked against the application, and three options beside it: a linear model predictive controller, a bounded moving horizon estimator for that controller and a relay experiment that writes proportional-integral constants into the PID. Section 2.6 reports the experiments. The offset of a proportional loop, the removal of that offset under integral action, and the collapse of the plant into an integrator on which proportional control has no offset at all are each produced by changing one input. The comparison of the predictive controller with the shipped proportional loop and with a proportional-integral reference separates what a plant model contributes from what integral action contributes, and the estimator and the relay are judged against the shipped choices on the same runs. Section 2.7 describes the devices and the platform of a plant level loop, and Sections 2.8 to 2.10 state the limitations, discuss the safety weight of level loops and close the chapter.

2.2 Background

The oldest surviving continuous feedback device regulated a level, and it did so in order to exploit the outflow law this application implements. The water clock attributed to Ktesibios used a float valve to hold a constant head in a supply vessel, because an orifice discharging under a constant head passes a constant flow, and only a constant flow makes the rising float in the receiving vessel a linear measure of time. Mayr traces that device and the long gap before the idea reappeared in European practice [9]. The float regulator returned as the boiler feedwater regulator and as the ball valve, and it is still the mechanism in a steam trap and in a domestic cistern. What changed in the twentieth century was the separation of the measurement from the actuator, so that a displacer or a differential-pressure transmitter reports the level to a controller that positions a valve somewhere else. The history box below tells how one such displacer misled the operators of a refinery tower.

> ⧖ HISTORY **A Level That Fell While the Tower Filled, 2005**
>
> On 23 March 2005 a refinery restarted a splitter tower 52 m tall after maintenance. For over three hours liquid was pumped in and none was taken out, and the tower overfilled. Its level transmitter was a displacer, which weighs the buoyant force on a body hanging in the liquid, and it read only a span of about 1.5 m within the bottom 2.7 m of the tower. It had been calibrated for a denser liquid than the hot liquid in the tower, so as the bottoms heated the reading fell. When the liquid stood at 48 m, the operators' screen showed 78 percent of span. Liquid released through the relief system formed a vapor cloud, and its explosion killed 15 people and injured 180 [10].

The control literature treated level as its own class of problem early, because the answers that work on temperature do not transfer. Proportional and proportional-integral level controllers were compared against nonlinear and nonconventional alternatives, and the conclusion drawn was that the right objective is often not to hold the level at all [11]. That observation became averaging level

Table 2.1 Classification of Variables. Nominal Values Are Those of the 10 m Operating Point With the Outlet Valve Half Open.

Symbol	Role	Nominal	Range
h	state, true level	10.00 m	0 to $h_{\max}$
$\hat{h}$ (LT)	CV, measured level	10.00 m	0 to $h_{\max}$
h_{SP}	setpoint	10.0 m	0 to $h_{\max}$, step 0.1 m
u_{in}	MV, inlet valve	11.07 %	0 to 100 %, step 0.1 %
u_{out}	DV, outlet valve	50 %	0 to 100 %, step 1 %
q_{in}	output, inflow	0.553 m^3/s	0 to 5 m^3/s
q_{out}	output, outflow	0.553 m^3/s	≥ 0
σ_n	DV, transmitter noise 1σ	0.05 m	0 to 0.30 m
overflow	output, spill indicator	false	—

control. A surge vessel exists to absorb a flow disturbance, so a controller that removes the level deviation quickly has passed the disturbance to the next unit and defeated the purpose of the vessel. The trade between level excursion and outlet flow variation has since been posed as an optimization problem [1], solved with a predictive controller under explicit constraints on the tank limits [2], and derived directly from the mass balance so that the tuning parameter is the allowed level swing rather than a gain [12]. Luyben returned to the question forty years after the first of those papers and argued that the simple cases are simpler and the awkward cases more awkward than the literature suggests [3]. Plantwide, the requirement is weaker than regulation and stronger than nothing, because each inventory must be controlled by a flow somewhere in a consistent direction for the plant to be able to change throughput at all [13]. Tuning rules reflect the same split. An integrating process has no steady-state gain to identify, so the rules that apply to it are written in terms of a slope and a desired closed-loop time constant rather than a gain and a lag, and a self-regulating level loop and a pumped-outflow level loop are therefore commissioned by different arithmetic on the same vessel [14, 15]. Volume 2, Chapter 5 identifies and tunes such an integrating level loop on a cooling tower basin.

Tanks became a standard teaching apparatus for the same reason they are awkward. The quadruple-tank rig places two square-root outflows in a configuration whose multivariable zero can be moved from the left half plane to the right by a manual split of two flows, which is a level experiment used to teach something that has nothing to do with level [6]. The hydraulics themselves are equally standard. Draining a vessel through an orifice is the textbook application of Bernoulli's principle, and the discharge coefficient that turns the ideal Torricelli velocity into a real flow has been measured in the undergraduate laboratory for decades [16].

2.3 Process Model

The simulated vessel is an open, vertical, constant-area tank filled through a control valve at the top and drained through a valve in the bottom. The controlled variable is the level reported by transmitter LT and the manipulated variable is the inlet valve position. The outlet valve is set by the operator in both control modes, so it serves as the load disturbance and as the switch between the two dynamic characters of the vessel. There is one state, one actuator and no cascade, so every mechanism described below acts inside a single feedback path. Figure 2.1 places each symbol where the equations use it.

The signals are sorted into their roles in Table 2.1. Two hard limits bound the state. The level is clipped at the overflow height $h_{\max}$, where an overflow indicator is raised and the excess leaves the vessel, and it is clipped at zero, where the tank is dry and the outflow term vanishes on its own. Neither limit stops the simulation. The plant is written in SI units with the level in meters, both valve positions in percent of travel and time in seconds. One state is integrated.

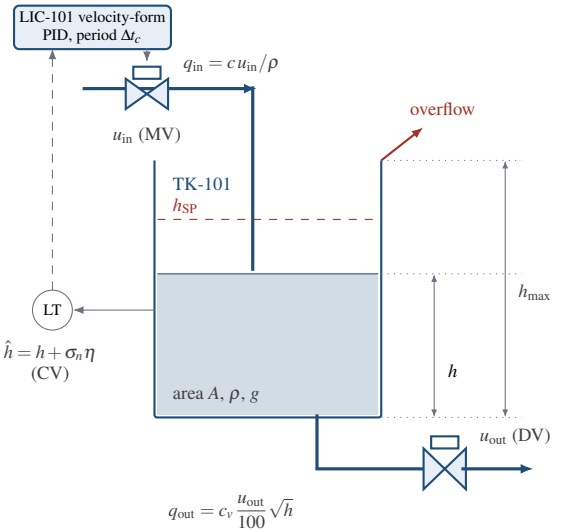


Figure 2.1 Gravity-drained tank TK-101 with its inlet control valve, its bottom outlet valve and the level transmitter. Inflow is proportional to the inlet valve position through the coefficient c, while outflow follows the square root of the level h through the outlet coefficient c_v and the fractional opening $u_{out}/100$. The controller LIC-101 reads the noisy level $\hat{h}$ and positions the inlet valve. The outlet valve is set by the operator and is never controlled.

2.3.1 Mass Balance and Outflow Law

The tank is a single well-mixed inventory whose volume is the level times a constant cross-sectional area, filled at a rate proportional to the inlet valve position and drained under its own head, as shown in Equation (2.1). It is the balance of Equation (1.2) with the two valve positions written out.

$$A\frac{\mathrm{d}h}{\mathrm{d}t} = \frac{c\,u_{in}}{\rho} - c_v\frac{u_{out}}{100}\sqrt{h}, \qquad 0 \le h \le h_{max} \tag{2.1}$$

where A is the tank cross-sectional area in m^2, c is the inlet valve coefficient in kg/s per percent of opening, ρ is the liquid density in $\mathrm{kg/m}^3$, c_v is the outlet coefficient in m^3/s per $\mathrm{m}^{1/2}$ at full opening, and u_{in} and u_{out} are the two valve positions in percent. Both sides of Equation (2.1) carry units of m^3/s. The inlet term is written as a mass flow divided by density because the problem statement on which the application is based specifies the inlet valve in kg/s per percent, while the outlet term is volumetric and carries no density. With the default constants the balance reduces to $\mathrm{d}h/\mathrm{d}t = 0.05\,u_{in} - 0.35\,(u_{out}/100)\sqrt{h}$ in m/s.

The outflow term is the Torricelli form obtained from Bernoulli's principle applied between the free surface and the discharge, with the ideal velocity $\sqrt{2gh}$ multiplied by an orifice area and a discharge coefficient, as shown in Equation (2.2).

$$q_{out} = C_d A_o\sqrt{2gh} \;\equiv\; c_v\sqrt{h} \qquad \text{at } u_{out} = 100\,\% \tag{2.2}$$

where C_d is the discharge coefficient, A_o is the discharge area in m^2 and g is gravitational acceleration. Neither C_d, A_o nor g appears anywhere in the code. The three are lumped into the single adjustable constant $c_v = C_d A_o\sqrt{2g}$, which is the only form a controller can distinguish. The default $c_v = 0.35\,\mathrm{m}^3/\mathrm{s}$ per $\mathrm{m}^{1/2}$ implies $C_d A_o = 0.0790\,\mathrm{m}^2$, which is 7.9 % of the $1\,\mathrm{m}^2$ tank cross-section and is a larger drain than a tank of this footprint would carry. The choice is deliberate, because it empties a full tank in about 20 s and makes every experiment fit inside a laboratory period. Such a drain would not be fitted to a 12 m vessel of $1\,\mathrm{m}^2$, but it compresses the whole self-regulating response

into a window that can be watched in real time without a speed multiplier, and the scaling results of Section 2.4 do not depend on it. The valve opening enters as a linear multiplier on c_v, which is a linear installed characteristic on the effective discharge area. Downstream pressure, pipe friction and the elevation of the outlet nozzle above the tank floor are absent, so the head that drives the discharge is the level itself.

> **+ DID YOU KNOW? The Square Root Inside a Flowmeter**
>
> The square root that drains this tank also governs a common flowmeter. Flow through an orifice plate is proportional to the square root of the pressure drop divided by the density, so the square root of the differential pressure signal is taken before it reads as flow [17]. Two such drains in series form the plant of Chapter 3.

Setting the derivative in (2.1) to zero gives the resting level at any pair of valve positions, as shown in Equation (2.3).

$$\bar{h} = \left[\frac{100\, c\, \bar{u}_{\text{in}}}{\rho\, c_v\, \bar{u}_{\text{out}}} \right]^2 \tag{2.3}$$

Equation (2.3) is the whole difference between this vessel and a pumped-outflow tank. A resting level exists for every inlet setting as long as the outlet valve is open at all, and it grows as the square of the inlet position, so the operating map is steep. Holding 10 m against a half open outlet requires $\bar{u}_{\text{in}} = 11.07\,\%$, and holding the same level against a fully open outlet requires 22.14 %. Beyond $\bar{u}_{\text{in}} = 24.25\,\%$ with the outlet wide open the resting level lies above $h_{\max}$ and the tank overflows instead of settling.

2.3.2 Measurement

The controller never reads the state. It reads the transmitter output of Equation (2.4).

$$\hat{h} = h + \sigma_n\, \eta \tag{2.4}$$

where σ_n is the noise standard deviation in m and η is a standard normal deviate drawn once per control execution by the Box-Muller transform from a seeded mulberry32 generator. The noise is white at the 10 Hz control rate, with no transmitter lag, no bias and no drift. The default of 0.05 m is pessimistic for a modern level instrument and is set that way so that the cost of derivative action is visible in a single run. The instrument chain contains no nonlinearity at all, and in particular the reported value is not clipped at zero, so an empty tank reads negative half the time. Half-rectifying the reading instead would add a bias of $\sigma_n/\sqrt{2\pi} = 0.399\,\sigma_n$ near an empty tank, which is 0.12 m at the largest noise setting and is a systematic error introduced by the instrument model rather than by the instrument. The true level and the measured level are reported together, so the effect of the measurement on the loop is separable from the state that produced it.

Every constant needed to rebuild the simulator is listed in Table 2.2. The initial level is the steady state of the settings a case starts with, and the initial integral is the manual inlet position at load, as described in Section 2.5. Operator-entered constants are sanitized before use, with A floored at machine epsilon, c and c_v floored at zero, $h_{\max}$ floored at 1 m and σ_n floored at zero, and the level is re-clipped to a new tank height at the next control execution. The setpoint range follows the tank height, and lowering $h_{\max}$ below the setpoint pulls the setpoint down to it.

2.3.3 Numerics

Equation (2.1) is integrated with explicit Euler on a fixed 0.05 s step, one of the app-supplied steps that Section 1.6.1 allows in place of the framework's Runge–Kutta method. A first-order right-hand side with a square root and two clipping bounds does not reward a higher-order integrator, because the truncation error is dominated by the two limits rather than by the smooth part. The level is clipped to $[0, h_{\max}]$ after every step, and the overflow indicator records whether the unclipped value exceeded

Table 2.2 Model and Controller Parameters. Every Value in the First Two Blocks Is Adjustable in the Application.

Symbol	Meaning	Default	Range	Unit
Plant constants, operator adjustable				
A	tank cross-section	1.0	0.3–5.0	m^2
c	inlet valve coefficient	50	10–100	kg/(s %)
c_v	outlet coefficient	0.35	0.10–0.80	m^3/s per $m^{1/2}$
h_{max}	overflow height	12	5–20	m
σ_n	transmitter noise	0.05	0–0.30	m
Controller, setpoint and outlet, operator adjustable				
K_c	proportional gain	10	0–100	%/m
K_I	integral gain	0	0–20	%/(m s)
K_D	derivative gain	0	0–50	% s/m
h_{SP}	level setpoint	10	0–h_{max}	m
u_{out}	outlet valve	10	0–100	%
—	speed multiplier	1	0.5–10	—
—	MPC estimator	Kalman	or MHE	—
Fixed in the source				
N_e	MHE window	30	—	executions
ε_r	relay band	0.15	—	m
ρ	liquid density	1000	—	kg/m^3
Δt	integration step (Euler)	0.05	—	s
Δt_c	controller sample period	0.1	—	s
I_0	initial integral	u_m	—	%
T_t	anti-windup tracking time	K_c/K_I	—	s
—	historian rate	5	—	samples/s
$h(0)$	initial level	steady	—	m

h_{max} on that step. The controller executes every 0.1 s of simulated time, which is two integration steps, and one row is recorded every 0.2 s, at every second control execution. A row carries the measured level that the controller acts on at the stamped time and the valve position held over the control period that has just ended. All three clocks run on simulated time, and the noise generator is reseeded whenever a case is loaded or reset, so a run at any speed between 0.5 and 10 gives the same trajectory with noise enabled (Section 1.6.1).

2.3.4 Assumptions

The model keeps only the mechanisms that change how the loop must be tuned. Each assumption below is stated with the range over which it holds and the condition under which it fails.

1. *Quasi-steady outflow.* The discharge is the steady Torricelli expression evaluated at the instantaneous level, so the liquid in the outlet line has no inertia and responds to a level change without delay. This is accurate while the drain time is long compared with the acceleration time of the outlet column, which holds throughout the default parameter set. It fails for a long discharge line, for a rapid valve stroke, and for any question about water hammer or oscillation of the outlet column.

2. *Constant discharge coefficient.* C_d is folded into c_v and does not vary with level or with valve opening. Measured discharge coefficients move by several percent over a wide Reynolds range and with the geometry of the opening [16]. The consequence is that the simulated outflow follows $\sqrt{h}$ more exactly than a real drain does, which makes the scaling results of Section 2.4 cleaner than a physical rig would show them.

3. *Linear valve characteristics.* The inlet flow is strictly proportional to the inlet position and the outlet coefficient is strictly proportional to the outlet position. There is no installed characteristic, no rangeability limit, no seat leakage, no stroking time, no hysteresis and no stiction. A real level loop frequently oscillates because of the last of these rather than because of its tuning, and no tuning in this application produces that symptom.

4. *Constant area and a flat bottom.* The cross-section A does not vary with height, so the level and the inventory are proportional. A conical, spherical or horizontal cylindrical vessel adds a second, purely geometric nonlinearity on top of the square-root outflow, and none of them is represented.

5. *Overflow discards liquid.* Above $h_{\max}$ the state is clipped and the excess leaves the model. Mass is therefore not conserved during a spill, the spilled rate is not reported, and the level returns from the limit as soon as the balance turns negative. This is adequate to show that a loop has lost control of the inventory. It cannot be used for any question about containment volume or spill rate.

6. *Adequate integration step.* The fastest plant reachable from Table 2.2 is $A = 0.3\,\mathrm{m}^2$ with $c_v = 0.8$ and the outlet fully open, which gives $\tau = 0.75\sqrt{\bar{h}}$ in seconds, so the 0.05 s step is at most 6.7 % of the local time constant above 1 m of level. Accuracy degrades close to an empty tank, where $\partial\sqrt{h}/\partial h$ is unbounded, and the clip at zero absorbs the resulting undershoot.

7. *Ideal transmitter apart from noise.* LT has no lag, no bias, no drift, no resolution limit and no rangeability, and its noise is white and Gaussian. Real level transmitters are dominated by low-frequency effects such as density change and turbulence at the surface, which are correlated rather than white, and a differential-pressure transmitter reads a head that depends on the density it was calibrated for.

8. *Constant density.* ρ is fixed at $1000\,\mathrm{kg/m}^3$ in the source and is not exposed. A change in composition or temperature that alters the density would move the inlet mass flow, the head at the outlet and the transmitter calibration together, and none of the three is available.

9. *Omitted entirely.* Inlet flow disturbances upstream of the valve, downstream pressure on the discharge, vapor space pressure, agitation, heat effects, cascade to a flow controller and any second vessel are all absent. Each is omitted because none of them changes the single behavior the application exists to expose.

2.4 Dynamic Analysis

2.4.1 Linearization About an Operating Level

Expanding Equation (2.1) about a resting point $(\bar{h}, \bar{u}_{\mathrm{in}}, \bar{u}_{\mathrm{out}})$ and keeping first-order terms, as Section 1.4.3 does for the running example, gives the first-order deviation model of Equation (2.5).

$$\tau \frac{\mathrm{d}h'}{\mathrm{d}t} + h' = K_u\, u'_{\mathrm{in}} + K_d\, u'_{\mathrm{out}} \tag{2.5}$$

where the primes denote deviations from the resting point and the three constants are given by Equation (2.6).

$$\tau = \frac{2A\sqrt{\bar{h}}}{c_v\, \bar{u}_{\mathrm{out}}/100}, \quad K_u = \frac{c}{\rho A}\,\tau, \quad K_d = -\frac{2\bar{h}}{\bar{u}_{\mathrm{out}}} \tag{2.6}$$

where τ is the process time constant in s, K_u is the process gain in m/% of inlet travel and K_d is the disturbance gain in m/% of outlet travel. Both τ and K_u carry a factor $\sqrt{\bar{h}}$, so they grow together as the tank fills, and both carry $1/\bar{u}_{\mathrm{out}}$, so they grow together as the drain is throttled. At 10 m with the

outlet half open the tank is a first-order plant with $K_u = 1.807\,\mathrm{m}/\%$ and $\tau = 36.1\,\mathrm{s}$, and at 3 m with the same outlet setting the same tank is a first-order plant with $K_u = 0.990\,\mathrm{m}/\%$ and $\tau = 19.8\,\mathrm{s}$. The two descriptions differ only in the inventory the vessel holds. The running example of Section 1.4.3 shows the same scaling in Equation (1.5).

Two consequences follow directly. The first is that the loop gain of a fixed controller is proportional to $\sqrt{\bar{h}}$, so a proportional gain set at the bottom of the tank produces a loop gain $\sqrt{h_{\max}/h_{\min}}$ times larger at the top. Over the usable range of the default tank, from 1 m to 12 m, that factor is 3.46. The neutralizer of Volume 2, Chapter 6, whose gain moves by orders of magnitude over one valve stroke, is the case in which a gain schedule becomes necessary. The second is that τ is exactly the time the same tank takes to empty from $\bar{h}$ with the inlet shut, because integrating (2.1) with $u_{\mathrm{in}} = 0$ gives $t_{\mathrm{empty}} = 2A\sqrt{h_0}/(c_v \bar{u}_{\mathrm{out}}/100)$. The identified time constant of the level loop and the observed drain time of the vessel are the same number, so the model is verified by a stopwatch measurement.

2.4.2 Invariance of the Gain to Time Constant Ratio

Dividing the gain by the time constant in (2.6) removes every operating-point term, as shown in Equation (2.7).

$$\frac{K_u}{\tau} = \frac{c}{\rho A} = 0.05\,\mathrm{m}/(\mathrm{s}\,\%) \tag{2.7}$$

The right-hand side of Equation (2.7) is the initial slope of the level response to an inlet step, and it depends on neither the level nor the outlet position. Every step test on this tank therefore starts at the same slope and differs only in its final value and in the time taken to reach it. That slope is also the integrating gain of the vessel, which is the connection to the second configuration. The cruise control of Volume 2, Chapter 2 has the same property: drag moves its gain and its time constant together, and their ratio is the thrust-to-mass ratio of the car.

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

2.4.3 The Integrating Limit

Closing the outlet valve removes the only level-dependent term in (2.1) and leaves Equation (2.8).

$$\frac{\mathrm{d}h}{\mathrm{d}t} = \frac{c\,u_{\mathrm{in}}}{\rho A} = 0.05\,\mathrm{m}/(\mathrm{s}\,\%) \times u_{\mathrm{in}} \tag{2.8}$$

The vessel is now a pure integrator with no resting level at any inlet setting except zero, which is the pumped-outflow tank of the textbook with the pump stopped. The application offers no separate integrating configuration and requires none, because $u_{\mathrm{out}} = 0\,\%$ produces it exactly and the outlet valve position moves continuously between the two cases. Taking $\bar{u}_{\mathrm{out}} \to 0$ in (2.6) sends τ and K_u to infinity with their ratio fixed at (2.7), which is the standard limiting argument by which a first-order lag with a very long time constant is treated as an integrator [7] and Equation (1.9) is reached from the running example of Chapter 1. The spent-fuel pool of Chapter 9 moves between the same two characters when its cooling trains stop. The consequence for control is asymmetric. The inlet valve can add liquid and no actuator can remove it, so a controller that has driven the level above setpoint with the outlet shut cannot return the level to setpoint.

2.4.4 Identification From the Running Application

Table 2.3 reports three step tests performed in manual mode from the first shipped case with the outlet valve wide open and the transmitter noise set to zero, each a 1 % increase in the inlet valve at $t = 300\,\mathrm{s}$ from the resting level at which the case starts once the two valve positions are entered before the run. The gain is the secant slope over the step, and the time constant is a one-parameter least-squares fit of a first-order response to the 300 s of logged history that follows the step, with the initial and final levels fixed at their measured values.

Table 2.3 Step Tests at Three Operating Levels, Outlet Valve at 100 % and $\sigma_n = 0$. The Predicted Columns Come From Equation (2.6) Evaluated at the Starting Level.

u_{in} (%)	$\bar{h}$ (m)	$\hat{K}_u$ (m/%)	K_u (m/%)	$\hat{\tau}$ (s)	τ (s)	$\Delta h/\bar{h}$ (%)	$\varepsilon_{\max}$ (m)
$10 \to 11$	2.041	0.429	0.408	8.73	8.16	21.0	0.0015
$17 \to 18$	5.898	0.714	0.694	14.44	13.88	12.1	0.0015
$22 \to 23$	9.878	0.918	0.898	18.53	17.96	9.3	0.0015

The identified gain rises by a factor of 2.14 between 2.04 m and 9.88 m against the factor of 2.20 predicted by $\sqrt{9.878/2.041}$, and the identified time constant rises by a factor of 2.12 against the same prediction. Both fitted values sit above the linearized prediction because a 1 % step moves the level by 0.43 m at 2.04 m, which is 21.0 % of the starting level, so the response is measurably faster at the end than at the beginning and the fit splits the difference. The bias is largest at the lowest level for that reason, at 7.0 % of τ against 3.2 % at the highest, while the largest residual of the fitted first-order response is 0.0015 m in all three tests, so the identified pair is a good description of a response that is nonetheless not first order. A drain test from a full 12 m tank with the inlet shut and the outlet wide open empties the vessel in 19.65 s against the 19.79 s of the analytic expression, and passes three quarters, one half and one quarter of the initial level at 2.65, 5.80 and 9.90 s against 2.65, 5.80 and 9.90 s, all read at the 0.05 s integration step. The 0.14 s shortfall arises close to the empty tank, where $\partial\sqrt{h}/\partial h$ is unbounded and the explicit Euler step is least accurate.

Three distinct models appear in this chapter, the three that Section 1.4.9 separates. The simulation model is Equation (2.1), nonlinear and exact within its assumptions. The identified control model is the first-order pair $(\hat{K}_u, \hat{\tau})$ of Table 2.3, valid near one level. The PID holds no model at all, because the algorithm of Section 2.5 carries three gains and a sample period and no representation of the plant. The predictive controller of Section 2.5.5 holds the third model, a first-order pair identified by a step test at 10 m with the outlet 10 % open and never updated as the plant moves away from that point.

2.5 Controller Design

2.5.1 Velocity Form With Derivative on the Measurement

LIC-101 runs the velocity-form PID of Section 1.6.2 in gain form [15]. Each cycle it computes an increment to the inlet valve rather than the position itself, and the operating point is carried by an integral state rather than by an operator-entered bias. The source stores the manual valve position at case load as a fixed bias and starts its integral at zero, and the equations below fold that constant into the integral, which gives the same recursion. The error is the setpoint minus the measurement, as shown in Equation (2.9), and the three increments are given by Equations (2.10) to (2.12).

$$e_k = h_{\text{SP}} - \hat{h}_k \tag{2.9}$$

$$\Delta u_k^P = K_c\, e_k - K_c\, e_{k-1} \tag{2.10}$$

$$\Delta u_k^I = K_I\, e_k\, \Delta t_c \tag{2.11}$$

$$\Delta u_k^D = -K_D\, \frac{\hat{h}_k - 2\hat{h}_{k-1} + \hat{h}_{k-2}}{\Delta t_c} \tag{2.12}$$

where K_c is in %/m, K_I in %/(m s), K_D in % s/m and $\Delta t_c = 0.1$ s. The increments are differences of the complete proportional and derivative terms, so a gain edited during a run acts on the whole current error at once. The proportional and derivative increments are accumulated, the demand is clipped to

the valve travel, and the integral is updated afterwards, as shown in Equation (2.13).

$$\begin{aligned} s_k &= s_{k-1} + \Delta u_k^P + \Delta u_k^D \\ c_k &= s_k + I_{k-1} \\ u_k &= \min(100,\ \max(0,\ c_k)) \\ I_k &= I_{k-1} + \Delta u_k^I + \frac{\Delta t_c}{T_t}\,(u_k - c_k) \end{aligned} \tag{2.13}$$

Here s_k accumulates the proportional and derivative increments without limit, c_k is the output the algorithm demands (the sum S_k and the unlimited output v_k of Equations (1.25) and (1.26), with the bias folded into I), the integral used at execution k is the one accumulated through execution $k-1$, and the last term is a back-calculation correction with tracking time $T_t = K_c/K_I$, which is 10 s at the tuning used in Section 2.6. The integral is updated only when $K_I \neq 0$. Away from both limits $u_k = c_k$, the correction vanishes, and the recursion is algebraically identical to the positional law $u_k = I_0 + K_c e_k + K_I \sum_{j<k} e_j \Delta t_c$ plus the derivative term, with I_0 the initial integral, so every tuning quoted in this chapter has its positional meaning. At a limit the two differ, because only the integral is corrected. Keeping s_k unclipped matters most at $K_I = 0$, where there is no integral to absorb the clipped part, for the reason Section 1.6.2 gives, and Section 2.6.3 shows what clipping it would do on this tank. The derivative acts on the measurement with no filter, so a setpoint move reaches the valve only through the proportional term, as a step of $K_c \Delta h_{\mathrm{SP}}$. The loop is reverse acting, because opening the inlet raises the level and e_k falls, and no action switch is offered. The gains are entered as K_c, K_I and K_D rather than as K_c, τ_I and τ_D, so a tuning correlation expressed in time constants has to be converted through $K_I = K_c/\tau_I$ and $K_D = K_c\tau_D$ first.

2.5.2 Initialization, Transfer and the Effective Bias

Every case starts at steady operation. In automatic the tank is placed at the setpoint and the manual station at the inlet position that holds it there, $\bar{u}_{\mathrm{in}}$ from Equation (2.3) with the outlet setting and plant constants in force, and in manual the tank is placed at the resting level of the manual inlet position. With the outlet shut no open inlet has a resting level, and the case starts at the setpoint, with the inlet shut in automatic. A setting changed before the run re-initializes the case exactly as Reset would, so a setpoint, outlet setting or inlet position entered before Run moves the starting steady state rather than producing a step. When a case is loaded the integral is set to the manual inlet position, so a case that opens in automatic issues $\bar{u}_{\mathrm{in}}$ plus K_c times its first error, which is transmitter noise alone. In manual the controller keeps executing and tracks the operator, setting $I_k = u_m - s_k$ so that its demand equals the manual valve position u_m. In automatic the manual setting follows the controller output at every execution without rounding, so both transfers are bumpless, as Section 1.6.2 describes.

The tracking gives the transfer a consequence that a positional law with an operator-entered bias does not have. Summing (2.10) from the last manual execution with $K_I = K_D = 0$ leaves $u_k = u_0 + K_c(e_k - e_0)$, where u_0 and e_0 are the valve position and the error at that execution. The loop therefore behaves as a positional law whose bias $u_0 - K_c e_0$ is fixed by the transfer condition. A shut-outlet tank holds still only with the valve shut, so a proportional-only loop transferred there rests at $\hat{h}_0 + u_0/K_c$, where $\hat{h}_0$ is the level read at transfer, whatever the setpoint. Measured on the shut-outlet tank at $K_c = 10\,\%/\mathrm{m}$ with a 10 m setpoint and no transmitter noise, starting from an empty tank (the first case with the inlet shut and the outlet wide open before the run, the outlet closed and the inlet opened at the execution at $t = 0.1$ s), filling by hand with the inlet at 0, 4, 20 and 60 % and switching to automatic as the level read 5.20, 5.16, 5.10 and 5.10 m settled the loop at 5.20, 5.56, 7.10 and 11.10 m, which matches $\hat{h}_0 + u_0/K_c$ to better than 0.001 mm in all four cases. For $u_0 = 0$ the tank was filled at 20 % and the inlet shut 1 s before the transfer. A case that opens in automatic carries the bias of its own starting inlet position, 2.21 % in the cases that start at 10 m with the outlet 10 % open.

2.5.3 Back-Calculation as the Anti-Windup Mechanism

Anti-windup is the last term of Equation (2.13) rather than a separate stage, and the running sum s_k is never truncated. With $K_D = 0$ the sum s_k telescopes to $K_c e_k$, so on a limit $u_{\lim}$ the clipped part is $u_{\lim} - c_k = u_{\lim} - K_c e_k - I_{k-1}$, and with $T_t = K_c/K_I$ the error part of the correction cancels Δu_k^I exactly. What remains is Equation (2.14).

$$I_k - u_{\lim} = \left(1 - \frac{\Delta t_c}{T_t}\right)(I_{k-1} - u_{\lim}) \tag{2.14}$$

where $u_{\lim}$ is 0 % or 100 %. The integral relaxes to the limit itself with time constant T_t whatever the error, and the demand exceeds the limit only by the proportional term $K_c e_k$. Once the integral has relaxed, the valve leaves the limit when the error changes sign, and no stored integral has to be worked off. Case 8 run with $K_I = 1\,\%/(\mathrm{m\,s})$ holds the valve at 100 % from the setpoint move at $t = 5\,\mathrm{s}$ to the end of a 60 s run (inlet coefficient 10 kg/s per percent, outlet wide open, setpoint moved from 6 to 10 m), and it reproduces the factor $(1 - \Delta t_c/T_t)^{100} = 0.366$ over 10 s to five digits. The application offers no free integrator, so it demonstrates the absence of windup rather than the failure mode.

2.5.4 Noise Propagation Through the Velocity Form

The derivative increments telescope in the same way. Summing (2.12) leaves a first difference rather than a running sum, so the noise contribution to the output is bounded rather than a random walk, as shown in Equation (2.15).

$$\sigma_u = \sigma_n \sqrt{\left(K_c + \frac{K_D}{\Delta t_c}\right)^2 + \left(\frac{K_D}{\Delta t_c}\right)^2} \tag{2.15}$$

where σ_u is the standard deviation of the valve position about its resting value in percent, valid while the loop is quiet and the output is off both limits. Equation (2.15) is obtained by writing the accumulated noise contribution as $-(K_c + K_D/\Delta t_c)\, n_k + (K_D/\Delta t_c)\, n_{k-1}$ with n_k the independent noise sample at step k. The short sample period amplifies the derivative term. With the default $\Delta t_c = 0.1\,\mathrm{s}$, a derivative gain of $K_D = 1\,\%\,\mathrm{s/m}$ contributes a factor $K_D/\Delta t_c = 10\,\%/\mathrm{m}$, which equals the whole default proportional gain. Integral action contributes a slow random walk of standard deviation $K_I \sigma_n \Delta t_c \sqrt{k}$ after k samples, which at $K_I = 1\,\%/(\mathrm{m\,s})$ reaches 0.17 % after two simulated minutes and is negligible beside the other two terms.

2.5.5 Model Predictive Control

A model predictive controller (MPC) is offered as an alternative to the PID in automatic mode, and the PID remains the default of every case. Every $T_s = 0.5\,\mathrm{s}$, or five controller executions, the MPC plans the inlet valve over a horizon, applies the first move and holds it until the next decision, the receding-horizon scheme of Section 1.6.3 [18, 19]. Its model is Equation (2.16), fitted with the noise off to the averaged responses to inlet steps of ± 0.2 percentage points at the design point of the shipped cases, $\bar{h}_m = 10\,\mathrm{m}$ with the outlet 10 % open and $\bar{u}_m = 2.214\,\%$. The fitted dead time is zero and the root-mean-square residual is 0.0004 m.

$$\tau_m \frac{\mathrm{d}x}{\mathrm{d}t} = -x + K_m\,(u_{\mathrm{in}} - \bar{u}_m + d)\,, \qquad h_m = \bar{h}_m + x \tag{2.16}$$

where x is the level deviation of the model in m, $K_m = 9.030\,\mathrm{m}/\%$ and $\tau_m = 180.7\,\mathrm{s}$ are the fitted gain and time constant, d is an unmeasured disturbance in percent of inlet travel and h_m is the level the model predicts. Equation (2.6) gives 9.035 m/% and 180.7 s at that point, and $K_m/\tau_m = 0.0500\,\mathrm{m}/(\mathrm{s}\,\%)$ reproduces the invariant of Equation (2.7). The model contains neither the square-root outflow nor the outlet valve. With the outlet half open the same tank at 10 m has $K_u = 1.807\,\mathrm{m}/\%$ and $\tau = 36.1\,\mathrm{s}$, a factor of five below both.

The disturbance d is estimated by a steady-state Kalman filter on Equation (2.16), sampled for a zero-order hold at Δt_c and augmented with d as a random walk, which is the standard arrangement for offset-free predictive control [20]. At every execution, whichever controller is in charge, the filter predicts and corrects as shown in Equation (2.17).

$$\begin{aligned} \bar{x}_k &= a\,\hat{x}_{k-1} + b\left(u_{k-1} - \bar{u}_m + \hat{d}_{k-1}\right) \\ \hat{x}_k &= \bar{x}_k + \ell_x\left(\hat{h}_k - \bar{h}_m - \bar{x}_k\right) \\ \hat{d}_k &= \hat{d}_{k-1} + \ell_d\left(\hat{h}_k - \bar{h}_m - \bar{x}_k\right) \end{aligned} \tag{2.17}$$

where $a = e^{-\Delta t_c/\tau_m} = 0.999447$, $b = K_m(1-a) = 0.004995\,\mathrm{m/\%}$, u_{k-1} is the inlet position applied over the period just ended, $\hat{h}_k$ is the reading the PID also uses, and $\ell_x = 0.0569$ and $\ell_d = 0.340\,\mathrm{\%/m}$ are the steady-state gains for a measurement variance $r = \sigma_n^2 = 0.0025\,\mathrm{m^2}$, a state noise variance $10^{-6}r$ and a disturbance noise variance $q_d = 10\,r/K_m^2 = 3.07\times10^{-4}\,\%^2$. A load that the model lacks appears in $\hat{d}$ as the inlet offset that would produce it, and the plan holds $\hat{d}$ over its horizon, so the closed loop settles without offset. The estimate is integral action obtained through the model rather than through the error.

The plan is a quadratic program on the model sampled at T_s, over a horizon $T_p = 360\,\mathrm{s}$ of $N = 720$ steps, or two model time constants. It chooses the inlet positions $v_1,\ldots,v_{12}$, held over move blocks of $1,1,1,1,2,2,3,4,6,9,15$ and 675 steps, that minimize the objective of Equation (2.18) subject to $0 \le v_j \le 100$.

$$J = \frac{T_s}{\tau_n}\sum_{k=1}^{N}\left(\frac{h_{m,k} - h_{\mathrm{SP}}}{y_s}\right)^2 + \lambda\sum_{j=1}^{12}\left(\frac{v_j - v_{j-1}}{u_s}\right)^2 \tag{2.18}$$

where $h_{m,k}$ is the level predicted k steps ahead with $\hat{d}$ held, v_0 is the inlet position applied now, $\tau_n = 180.7\,\mathrm{s}$ keeps the sum from growing with the horizon, $y_s = 0.5\,\mathrm{m}$ is the settling band of the scorecard, $u_s = 100\,\%$ is the valve span and $\lambda = 0.03$ is a dimensionless move suppression. The valve range, in percent, is the only constraint and it is hard. There is no level band, overflow limit, rate limit or terminal weight, and no preview, because the controller is not told the outlet position. The decisions are formed as deviations from $\bar{u}_m$, and a ridge of 10^{-9} on the scale of the weights keeps the program strictly convex. Its 12 variables and 24 bound rows are handled by a dual active-set solver [21] from the framework's shared modules. The shipped λ and the factor of 10 in q_d come from one rule for the seven single-loop studio apps, which takes the largest move suppression within 10 % of the best true-level IAE among settings with no more valve travel than the shipped PID on cases 3 and 4. For this tank the rule selects the smallest λ swept.

A plan that is not optimal, or whose first move is not finite or leaves the valve range, is rejected. The PID tracks the applied valve at every execution as it tracks the operator in manual, so it takes the valve without a bump, the history is marked, and the next valid plan takes the valve back. A rejection forced at 6 s of case 9 without noise hands over with a valve change of 0.38 %, the size of the next three moves of the PID (0.32 to 0.36 %), and the take-back moves the valve by 3.2 %. The first plan is solved when a case loads, so a run under the MPC is logged from $t = 0$. The controller reads only the transmitter, the applied inlet position, the setpoint and the valve limits, which keeps the plant, the instrument and the controller model apart.

2.5.6 Moving Horizon Estimation

While the MPC is the controller, its plans may start from a moving horizon estimator (MHE) instead of the filter of Equation (2.17), which keeps running either way. At execution k the MHE chooses z_s and $w_s,\ldots,w_{k-1}$ over the window $j = s,\ldots,k$, $s = k - N_e$, to solve the bounded least-squares problem

of Equation (2.19), in the form analyzed in [22].

$$\begin{aligned} \min\ & \|z_s - \bar{z}_s\|^2_{\bar{P}^{-1}} + \sum_{j=s}^{k-1} w_j^{\mathsf{T}} Q^{-1} w_j \\ & + \frac{1}{r} \sum_{j=s}^{k} \left(\hat{h}_j - \bar{h}_m - x_j\right)^2 \\ \text{s.t. } & z_{j+1} = \begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix} z_j + \begin{bmatrix} b \\ 0 \end{bmatrix} (u_j - \bar{u}_m) + w_j \\ & 0 \le \bar{h}_m + x_j \le h_{\max,0} \end{aligned} \tag{2.19}$$

Here $z_j = [x_j \;\; d_j]^{\mathsf{T}}$ augments the model state with the disturbance, w_j is the process noise, $Q = \mathrm{diag}(10^{-6} r, q_d)$ and r are the variances of the filter, u_j is the inlet position applied after execution j, $\bar{z}_s$ is the estimate of execution $s-1$ advanced one step, and $\bar{P}$, the steady a-priori covariance of the filter, is the fixed arrival cost. The window of $N_e = 30$ executions spans 3 s, which gives 62 variables and 62 bound rows for the dual active-set method of the plan. The bound holds the model level between the floor and $h_{\max,0} = 12\,\mathrm{m}$, the declared default rather than the height set for the run, and d is unbounded because no declared value limits it. Without an active bound the problem is the one the filter solves recursively. A window that ends without an optimum, or with an estimate that is not finite or not inside the bound, gives way to the Kalman correction of the MHE's own prediction one step ahead, projected onto the bound, and the history marks the substitution.

2.5.7 Relay Autotuning

The PID gains may instead be obtained from the closed loop itself through the relay feedback experiment of Åström and Hägglund [23]. It is an action started from the operating state of the moment under the PID in automatic, and no shipped case runs with it engaged. The relay holds the inlet valve δ_r above or below u_0, the mean of the valve positions applied during the 6 s before the test, with δ_r the smaller of 5 % of the span and 0.9 of the distance from u_0 to the nearer limit, and it switches each time the reading passes more than $\varepsilon_r = 3\sigma_{n,0} = 0.15\,\mathrm{m}$ beyond the setpoint, three times the declared noise $\sigma_{n,0}$ whatever noise the run carries. Cycles with a half shorter than four executions are not counted. When the periods and the swings $\pm\alpha$ of two successive counted cycles match to 5 %, the hysteresis-corrected describing function yields the ultimate gain K_{cu}, and the Tyreus–Luyben settings follow from it and the period P_u, as in Equation (2.20), where D is the part of each cycle that the valve spends above u_0, equal to one half when the cycle is symmetric.

$$K_{cu} = \frac{4\delta_r \sin(\pi D)}{\pi \sqrt{\alpha^2 - \varepsilon_r^2}}, \qquad K_c = \frac{K_{cu}}{3.2}, \qquad K_I = \frac{K_c}{2.2\,P_u} \tag{2.20}$$

The constants enter the PID to three significant figures with $K_D = 0$ and stay editable. Without convergence the relay stops after 30 cycles or 120 s, a second press cancels it, and the PID tracks the applied valve throughout so that it resumes without a bump. Once $\psi = \arcsin(\varepsilon_r/\alpha)$, the angle by which the hysteresis displaces the cycle from the -180° point, reaches 10°, the result is reported as set by the band rather than by the tank.

2.6 Results

The application ships ten cases. All ten start at steady operation, nine of them at 10 m with the outlet 10 % open and the inlet at the 2.21 % that holds that level, and the first scripted event of every case but the first falls at $t = 5\,\mathrm{s}$, so nothing moves before it except the response of the loop to transmitter noise. Case 1, Reference operating point, holds that state in manual. Case 2, Open-loop actuator step, steps the inlet from 2.21 to 1.21 % in manual. Cases 3, 6 and 7 move the setpoint in automatic, to 11.2 m at

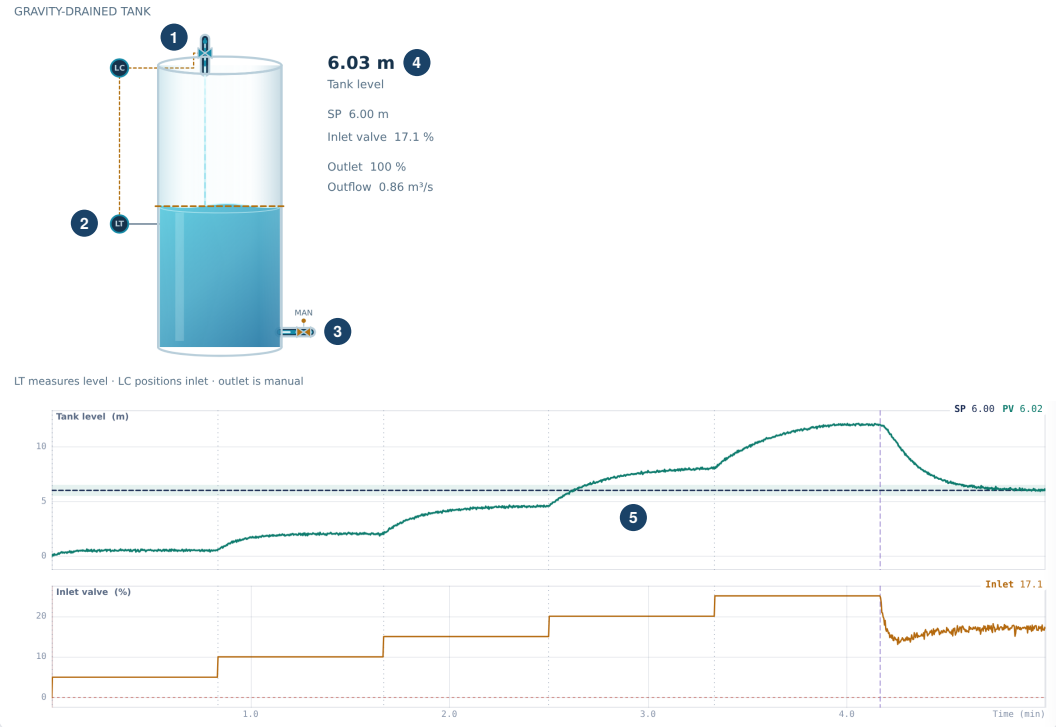


Figure 2.2 Equal inlet valve steps producing unequal level steps on the gravity-drained tank, run from the first shipped case with the inlet shut and the outlet wide open before the run, so that the case starts with the tank empty. The outlet valve stays wide open for the whole run and the transmitter noise is at its 0.05 m default. Manual inlet settings of 5, 10, 15, 20 and 25 %, the first entered at $t = 0.1$ s and each later one 50 s after the one before, end at levels of 0.51, 2.04, 4.55 and 8.01 m and then at the 12 m overflow limit, reached at $t = 235.8$ s, so the same 5 % increment raises the level by 0.51 m near the bottom of the tank and by 3.46 m near the top, and each new level takes longer to reach than the one before it. At $t = 250$ s the loop is switched to automatic with $h_{SP} = 6$ m, $K_c = 10\,\%/\text{m}$ and $K_I = 1\,\%/(\text{m s})$, and at $t = 300$ s it has drained the tank to 6.03 m with the inlet valve at 17.1 %. The figure is a composite of the application window: the plant view above the two trend strips, with the header and the side panels cut. Callout 1 marks the inlet valve that LC moves, 2 the level transmitter LT, 3 the outlet valve set by hand, 4 the readout column with the level of 6.03 m and the inlet at 17.1 %, and 5 the level trend, whose steps grow at each equal valve step. The trend draws the levels without printing them, and their values are those given in Section 2.6.

the reference gain $K_c = 10\,\%/\text{m}$ in case 3 and to 10.96 m at $K_c = 4$ and 20 %/m in the conservative and the aggressive case. Cases 4 and 5 open the outlet from 10 to 50 %, under the proportional loop in case 4 and in manual in case 5, where the inlet has to be set by hand to hold the level once settled. Case 8, Actuator limit, opens the outlet wide, weakens the inlet valve to $c = 10\,\text{kg/s}$ per percent and moves the setpoint from 6 to 10 m. Case 9 runs the predictive controller of Section 2.5.5 through the outlet step to 50 % and a setpoint move to 11.2 m at 30 s, and is then repeated under the PID. Each case has a completion rule, as Section 1.6.4 describes.

Case 10, the challenge, applies the outlet step of case 4 with $K_c = 2\,\%/\text{m}$ and asks for the measured level within ± 1.5 m of 10 m over the last third of a 20 s run, from 13.2 s. It is judged on the case as given, so changing a plant constant or the outlet setting voids the attempt. With the shipped $K_c = 2\,\%/\text{m}$ the level is still falling, at 7.07 m, when the run ends, and the run fails with a worst error

of 2.97 m. The suggested $K_c = 10\,\%/\mathrm{m}$ with $K_I = K_D = 0$ passes with a worst error of 0.946 m, a margin of 0.55 m, and the app gives the same run whether the suggested tuning is entered before Run or carried through Reset. The band admits the 0.84 m proportional offset of case 4, and 0.10 m of the judged error is transmitter noise.

Case 2 runs open loop. The 1 % inlet step at $t = 5\,\mathrm{s}$ starts the level down at 0.05 m/s, the slope of Equation (2.7), and the level reaches 8.62 m at 35 s and 7.64 m at 60 s on its way to the 3.01 m resting level of the smaller inflow, with a time constant of 181 s at 10 m and a 10 % outlet. Cases 6 and 7 differ only in the proportional gain, and case 3 makes a larger move at the reference gain. Because each case starts with the bias that holds 10 m, a proportional loop is left short of a new setpoint only by the change in the required inlet position divided by K_c, and at $K_c = 4$, 10 and 20 %/m the level lies 0.033, 0.014 and 0.007 m below the stepped setpoint over 25 to 35 s (mean of the true level). The true level rises from 10 to 90 % of the move in 12.25, 4.5 and 2.0 s, read at the integration step, and over the 35 s run the scorecard valve travel of cases 6 and 7 is 0.39 and 2.06 times that of case 6 rerun at $K_c = 10\,\%/\mathrm{m}$, most of it driven by transmitter noise. Without noise the true level never reaches the setpoint, because a proportional loop on a first-order plant with no dead time does not overshoot. With the default noise it crosses the setpoint by 6 mm at $K_c = 10$ and 28 mm at $K_c = 20\,\%/\mathrm{m}$, and the overshoot tile of the scorecard, computed from the measured level, reads 9.3, 8.1 and 13.5 % for cases 3, 6 and 7, all of it noise. Case 4 parks 0.84 m low, as analyzed below. Case 5 starts from the same steady state in manual, so the outlet step drains the tank toward the 0.40 m resting level of the unchanged 2.21 % inlet, through 7.94 m at 10 s to 0.48 m at 60 s, unless the inlet is raised by hand, and 11.1 % holds 10.06 m. Case 8 starts at 6 m with the weaker inlet at 85.7 %. The setpoint move to 10 m drives the inlet valve to 100 % at the execution at $t = 5.0\,\mathrm{s}$, where it stays, and the level climbs toward 8.16 m, the most the valve can hold, reading 8.09 m at 60 s. No tuning closes the remaining 1.84 m.

> **TRY IT IN THE APP** **2 · Open-Loop Actuator Step**
>
> Load the case and press Run. At 5 s the inlet valve closes by one percentage point with the controller in manual. Watch the tank level trend: the level starts down at 0.05 m/s, passes 8.62 m at 35 s and 7.64 m at 60 s, and keeps falling toward the 3.01 m resting level of the smaller inflow.

The experiments of this section start from those cases and change the valve positions, the control mode, the three gains or the plant constants, as listed in Table 2.4. Every number below was computed from the logged history, which carries one row every 0.2 s, except that crossing times are read at the 0.05 s integration step and the valve statistics of Table 2.6 at every 0.1 s controller execution.

2.6.1 Manual Operation and the Steady-State Map

Equal increments of the inlet valve produce level steps that grow with the inventory already held, as shown in Figure 2.2. Five manual settings at 5 % intervals, each held for 50 s with the outlet wide open, end at levels of 0.51, 2.04, 4.55 and 8.01 m and then at the overflow limit, against the 0.51, 2.04, 4.59, 8.16 and 12.76 m resting levels of Equation (2.3). The fourth reading falls short of its prediction because 50 s is only 3.1 time constants at that level, and the fifth prediction lies above $h_{\max}$, so the tank spills instead of settling. The same run carries the time-constant result, because each successive approach takes longer than the one before it while the valve increments that caused them are identical.

2.6.2 Offset Under Proportional and Proportional-Integral Control

Table 2.5 reports the identical load disturbance, the outlet valve opened from 50 % to 70 %, applied to the same controller at two setpoints. Each run starts from case 4 with the transmitter noise set to zero and $K_I = 1\,\%/(\mathrm{m\,s})$ entered before the run, so that integral action finds the inlet position that holds the setpoint against the half-open outlet, which is the value listed in the *inlet before* row. At $t = 300\,\mathrm{s}$ the integral gain is set to the value of the column, which leaves the valve where it is, and

Table 2.4 Experiment Sequence and the Mechanism Each One Exposes

Experiment	Setup	Mechanism exposed
Fill and drain	case 1: inlet 100 % with the outlet shut, then inlet shut and outlet 100 %	$t_{\text{empty}} = 2A\sqrt{h_0}/c_v$ and the overflow limit
Step test	case 1, outlet 100 %: 1 % inlet step at three settled levels	$\hat{K}_u$ and $\hat{\tau}$ both scale as $\sqrt{\bar{h}}$
Outlet upset, P only	case 4 settled with $K_I = 1$, then $K_I = 0$ and outlet $50 \to 70$ %	offset proportional to the local process gain
Outlet upset, PI	same upset, $K_I = 1$	offset removed for under 1 % of added valve travel
Same tuning, two levels	repeat both at 3 m and 10 m	one tuning producing two closed-loop responses
Outlet closed	case 3 with $u_{\text{out}} = 0$ %, P only	proportional control with no offset on an integrator
Outlet open, carried bias	case 4 as shipped, outlet $10 \to 50$ %	offset $(u_{ss} - u_b)/K_c$ with the bias of the steady start
Noise and derivative	case 4, $K_I = 1$, K_D raised, $\sigma_n = 0.05$ m	valve activity from Equation (2.15)
Predictive control	cases 3, 4, 8, 9 and 10, MPC chosen before the run	offset removed through a disturbance estimate

Table 2.5 Response to an Outlet Valve Step From 50 % to 70 % With $K_c = 10\,\%/\text{m}$ and $K_D = 0$, $\sigma_n = 0$, Measured Over the 240 s Following the Upset

	$h_{\text{SP}} = 10$ m		$h_{\text{SP}} = 3$ m	
K_I (%/(m s))	0	1	0	1
K_u at the setpoint (m/%)	1.807	1.807	0.990	0.990
τ at the setpoint (s)	36.1	36.1	19.8	19.8
Level if uncontrolled (m)	5.10	5.10	1.53	1.53
Inlet before (%)	11.07	11.07	6.06	6.06
Inlet after (%)	15.17	15.50	8.18	8.49
Final offset (m)	0.411	0.000	0.212	0.000
Maximum deviation (m)	0.411	0.323	0.212	0.170
IAE (m s)	97.84	4.43	50.53	2.42
Valve travel (%)	4.1	5.0	2.1	2.5
Settling within 0.02 m of the final level (s)	5.50	30.10	4.05	26.20

the outlet is opened to 70 %. Every run therefore starts on setpoint with no inherited error whether or not it has integral action, and the offset reported is the deterministic one.

The two proportional columns differ only in how full the tank was. The same $K_c = 10\,\%/\text{m}$ leaves a standing error of 0.411 m at 10 m and 0.212 m at 3 m, a ratio of 1.94. The linearized prediction of that offset is $K_d\,\Delta u_{\text{out}}/(1 + K_u K_c)$ with both constants taken from Equation (2.6), which evaluates to 0.420 m and 0.220 m and to a ratio of 1.90, so the first-order model accounts for the measured offsets to within 2.1 % and 3.7 % even though the disturbance itself is far outside the range in which that model was derived. Without any control the same upset would take the level from 10 to 5.10 m and from 3 to 1.53 m by Equation (2.3), so proportional feedback alone recovers 92 % of the excursion at 10 m and 86 % of it at 3 m, and integral action recovers the rest. Figure 2.3 sets the offsets, errors and valve travel of the table side by side.

🕮 **Pause.** Try Exercise 2.2 before reading on.

Integral action carries almost no cost on this plant. Adding $K_I = 1\,\%/(\text{m s})$ removes the offset

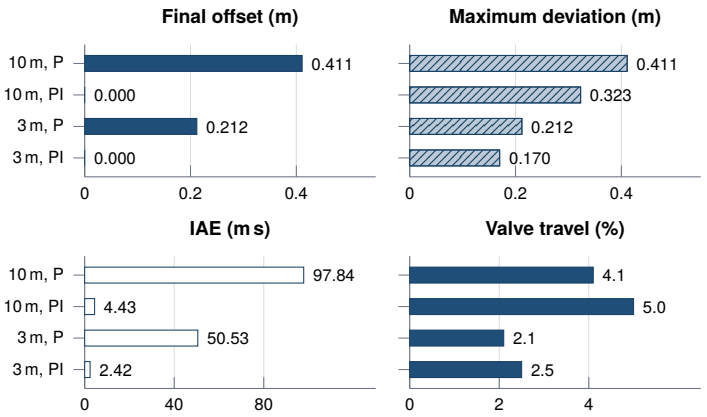


Figure 2.3 The outlet step of Table 2.5 at the two setpoints, proportional (P) and proportional-integral (PI) with the same K_c: final offset and valve travel (solid bars), largest deviation (hatched) and integrated absolute error (open). Notice that the same proportional gain leaves about twice the offset at 10 m as at 3 m, and that integral action removes both offsets for less than one percentage point of added valve travel.

entirely, cuts IAE from 97.84 m s to 4.43 m s at 10 m and from 50.53 m s to 2.42 m s at 3 m, reduces the maximum deviation in both cases, and spends 0.9 % and 0.4 % more valve travel. The trade that averaging level control is built around does not appear here, because the disturbance is a change in the discharge path itself rather than an inlet flow that has to be passed on. On this vessel the cost of tight regulation is measured in valve activity driven by the transmitter rather than in flow variation handed to a downstream unit, which is the subject of Section 2.6.4.

2.6.3 Proportional Control of the Integrating Configuration

Closing the outlet valve changes the character of the loop rather than its tuning, and it removes the offset rather than enlarging it. Setting $u = u_b + K_c e$ equal to the valve position u_{ss} the plant needs to hold still gives the steady-state error of a proportional loop directly, as shown in Equation (2.21), where u_b is the effective bias of Section 2.5.

$$e_\infty = \frac{u_{ss} - u_b}{K_c} \tag{2.21}$$

A vessel with no discharge holds still only with the inlet shut, so $u_{ss} = 0$. A loop that opens in automatic on the shut-outlet tank starts with the inlet shut, the only position that holds it still, so it carries $u_b = 0$, and the offset is therefore zero. Proportional action alone drives the level to setpoint and holds it there with $K_I = K_D = 0$, and it reaches the new setpoint after a setpoint move. Measured on case 3 with the outlet shut before the run, $K_c = 10\,\%/\mathrm{m}$ and no transmitter noise, the case starts at 10 m with the inlet shut, follows the scripted move to 11.2 m at $t = 5\,\mathrm{s}$, and holds the level within 1 mm of the new setpoint from $t = 18.85\,\mathrm{s}$ with the inlet valve shut again.

This property is the reason proportional-only control is accepted on integrating level loops, and it does not hold on the self-regulating side of the same vessel. Case 4 starts at 10 m with the outlet 10 % open and the inlet at the 2.21 % that holds it, which becomes the bias u_b. Once the outlet opens to 50 % at $t = 5\,\mathrm{s}$, holding 10 m requires 11.07 % of inlet valve, so Equation (2.21) predicts an offset of $(11.07 - 2.21)/K_c = 0.886\,\mathrm{m}$, and the loop settles 0.838 m low (mean of the true level from 60 to 300 s). The remaining discrepancy is accounted for by the square-root outflow law rather than by error in the prediction. At the depressed level of 9.162 m the required valve is only 10.59 %, which returns $(10.59 - 2.21)/K_c = 0.838\,\mathrm{m}$ exactly. No single bias is correct at every level and outlet setting

Table 2.6 Valve Activity of a Settled Loop at 10 m With the Outlet Half Open, $\sigma_n = 0.05\,\text{m}$, $K_c = 10\,\%/\text{m}$ and $K_I = 1\,\%/(\text{m s})$, Measured Over 600 s

K_D (% s/m)	0	1	5
σ_u measured (%)	0.505	1.167	4.582
σ_u from Eq. (2.15) (%)	0.500	1.118	3.905
$\sigma_{\hat{h}}$ measured (m)	0.051	0.051	0.053
σ_h of the true level (m)	0.009	0.009	0.017
Valve travel (% per s)	5.8	15.8	66.3

on that side of the vessel, which is the direct argument for integral action, and it is why the runs of Table 2.5 reach their starting point with integral action on.

> **TRY IT IN THE APP** **4 · Disturbance Rejection**
>
> Load the case and run it for 60 s. When the outlet opens from 10 to 50 % at 5 s, the proportional loop opens the inlet but parks the level about 0.84 m below the 10 m setpoint. Then Reset, set the Integral gain to 1 before Run and run again: the level ends 0.001 m above the setpoint (Table 2.7).

Restoring the default 0.05 m of transmitter noise degrades the offset-free result, and the mechanism is the instrument rather than the controller. With the outlet shut the valve opens on any sample reading below setpoint, and the admitted liquid cannot be removed through a closed drain, so the level rises until such a reading has become rare. The level approaches 3.2 standard deviations above setpoint. In case 3 with the outlet shut it stands 0.12 m above the 11.2 m setpoint at five minutes and 0.16 m above it at fifty minutes with $\sigma_n = 0.05\,\text{m}$, and 0.24 m and 0.32 m above it with $\sigma_n = 0.10\,\text{m}$. Across five noise seeds the fifty-minute value spans 0.006 m and 0.012 m at the two noise levels. Opening the outlet removes the bias, because the vessel then discharges the excess.

Pause. Try Exercise 2.3 before reading on.

The bound itself is a property of keeping s_k unclipped. An implementation that clips the running sum instead discards the negative increments computed while the valve already sits at 0 % while applying the positive ones in full, and the resulting rectification leaves no resting point at all. The same experiment, run as a model counterfactual with the controller replaced by that variant, reached the 12 m overflow limit at $t = 30.1\,\text{s}$, 25.1 s after the setpoint move, with no disturbance applied, and between 27.05 and 33.65 s across five noise seeds. The contrast makes the experiment a compact demonstration of why an integrating inventory with a one-way actuator is treated as a different control problem rather than as a slow first-order one.

2.6.4 Derivative Action and Measurement Noise

Table 2.6 reports the valve activity of a settled loop at 10 m with the default transmitter noise and three derivative settings, measured from case 4 with K_I and K_D entered before the run, over the 600 s from $t = 200\,\text{s}$ with no further disturbance. The statistics are taken at every controller execution. The 0.2 s log gives the same standard deviations to within 2.5 % but only 2.8, 6.3 and 23.4 %/s of valve travel, because it sees every second valve move.

Equation (2.15) predicts the measured valve noise to within 1.0 % at $K_D = 0$ and 4.4 % at $K_D = 1\,\%\,\text{s/m}$, and underpredicts it by 17 % at $K_D = 5\,\%\,\text{s/m}$, because the expression keeps only the direct path from the transmitter to the valve. A valve deviation u' held for one execution moves the level by $c\,u'\Delta t_c/(\rho A) = 0.005\,u'$ in meters, and the derivative term returns that motion at the next execution as $-K_D c\,u'/(\rho A)$, which is $-0.25\,u'$ at $K_D = 5\,\%\,\text{s/m}$. Adding that one-step autoregression to the noise path of Equation (2.15) raises the prediction to 4.50 %, within 2 % of the measurement. The true level is disturbed far less than the valve. Its standard deviation rises from 0.009 m, the residual of a loop that is only reacting to the instrument, to 0.017 m on the sampled integration phase, a third of the

Table 2.7 Shipped PID, Predictive Controller and a Proportional-Integral Reference on Five Cases With the Same Noise Seed

Case	Controller	IAE (m s)	IAE_h	$e_{\max}$ (m)	$\bar{e}_{\text{end}}$ (m)	Travel (%)	Limit (%)
3, setpoint	P	4.43	2.86	1.073	0.010	188.5	0
	PI[a]	4.98	3.47	1.072	−0.004	187.2	0
	MPC	3.07	0.98	0.916	−0.007	82.6	0
4, outlet	P	44.52	44.61	0.856	0.835	176.0	0
	PI[a]	9.65	8.89	0.660	−0.001	175.4	0
	MPC	5.87	4.16	0.649	−0.006	33.2	0
8, limit	P	133.61	133.68	3.969	1.918	29.2	91.7
	MPC	133.59	133.70	3.968	1.918	16.6	91.7
9, mismatch	P	48.42	48.51	1.928	0.899	191.1	0
	PI[a]	11.85	11.35	1.159	−0.015	190.1	0
	MPC	6.32	4.77	0.916	−0.006	91.6	0
10, challenge	P, $K_c = 2$	29.23	29.06	2.931	2.735	13.2	0
	P, $K_c = 10$[b]	11.28	11.12	0.846	0.837	61.3	0
	MPC	4.14	3.84	0.649	−0.005	17.6	0

IAE uses the measured level and IAE_h the true level. $e_{\max}$ is the largest true deviation after the first event, $\bar{e}_{\text{end}}$ the mean of $h_{\text{SP}} - h$ over the last 5 s, and Limit the share of the run with the valve at a limit. P is the shipped proportional-only tuning. [a]Reference, not shipped, $K_I = 1\,\%/(\text{m s})$. [b]Suggested, $K_I = K_D = 0$.

transmitter noise. Sampling the other integration phase gives about 0.012 m, so the apparent increase depends on the phase. Derivative action provides no compensating benefit on this plant. A first-order process with no dead time has no phase lag for derivative action to recover, so the only effect of K_D is to convert transmitter noise into valve travel at 66.3 % per s, which is 11 times the travel of the $K_D = 0$ case, and to increase the sampled variation of the true level. The near-doubling applies to the integration phase used in the table. That result is the reason derivative action is switched off in the large majority of industrial loops [15, 24], and the reason both K_I and K_D default to zero in the application.

> **+ DID YOU KNOW? Derivative Action That Only Moves the Valve**
>
> From the simulator: on the settled tank, raising the derivative gain from 0 to 5 % s/m raises the valve travel from 5.8 to 66.3 % per second, while the standard deviation of the true level grows only from 0.009 to 0.017 m (Table 2.6).

2.6.5 Comparison With the Predictive Controller

The controllers are compared on cases 3, 4, 8, 9 and 10 in Table 2.7, each case run for 60 s, or 20 s for the challenge, with the controller chosen before the run, so that both see the same events and the same noise sequence. The PID keeps the shipped tuning of each case and the MPC keeps the settings of Section 2.5.5 throughout. The shipped PID is proportional only, with $K_c = 10\,\%/\text{m}$ in cases 3, 4, 8 and 9, while the challenge ships $K_c = 2\,\%/\text{m}$ and suggests $K_c = 10\,\%/\text{m}$, both with $K_I = K_D = 0$, so the disturbance estimate gives the MPC an advantage of structure over every shipped tuning. A proportional-integral reference with $K_I = 1\,\%/(\text{m s})$, the setting of Table 2.5, is added on cases 3, 4 and 9 to separate the contribution of the model from that of integral action.

On the setpoint move of case 3, at the point where the model was fitted, every controller ends within 0.01 m of 11.2 m. The MPC opens the inlet to 31.7 % against 14.5 %, the true level rises from 10 to 90 % of the move in 0.85 s against 4.5 s with an overshoot of 0.041 m against 0.006 m, and the true-level IAE is 0.98 m s against 2.86 m s. Over the quiet first 5 s of cases 3, 4 and 9 the proportional loop moves the valve by 15.95 % and the MPC by 2.65 %, because the filter gain $\ell_x = 0.0569$ admits a small fraction of each reading and a plan is made only at every fifth execution, which accounts for most of the difference in travel.

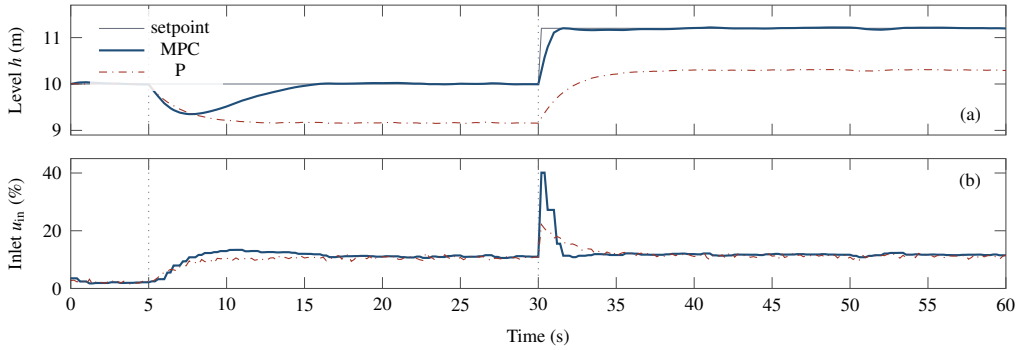


Figure 2.4 Case 9 under the predictive controller and under the shipped proportional loop, with the same noise seed. The outlet valve opens from 10 to 50 % at 5 s and the setpoint steps to 11.2 m at 30 s (dotted lines). (a) True level and setpoint. (b) Inlet valve position. Thick solid: MPC. Dash-dotted: shipped proportional loop. Thin gray: setpoint.

The outlet step of case 4 separates the structures. The proportional loop settles 0.835 m low, the offset of Equation (2.21), while the MPC ends 0.006 m above the setpoint with $\hat{d} = -8.79\,\%$, close to the −8.85 % that Equation (2.3) gives as $\bar{u}_m$ minus the 11.07 % that the half-open outlet requires. The proportional-integral reference also removes the offset, but with 8.89 m s of true-level IAE against 4.16 m s and 175.4 % of travel against 33.2 %. Most of the gain of the MPC over the shipped loop therefore comes from removing the offset, and the model with its filter adds a factor of 2.1 in IAE and 5.3 in travel.

Case 8 is a tie. Both controllers hold the inlet at 100 % from the execution at 5.0 s to the end, 91.7 % of the run, both leave the level at 8.094 m, and their IAEs differ by less than 0.02 m s, although the initial slope of the model is 5.0 times and its gain 64.5 times those of this tank with its weaker inlet. Every plan was optimal, and 110 of the 120 had the upper bound active, with up to 12 active-set iterations. The lower travel of the MPC comes from the 5 s before the step, 2.27 % against 15.76 %.

Case 9 joins the outlet step of case 4 and the setpoint move of case 3, so the move is made on a tank whose gain and time constant are a fifth of those of the model, as shown in Figure 2.4. The model keeps the initial slope of the vessel, so its response to a held 1 % inlet step after one decision period is 0.0249 m against 0.0248 m for the linearized tank, and the mismatch grows only with time, to 7.80 m against 1.81 m at the end of the horizon. Before the move $\hat{d}$ reaches −8.82 %. After it the MPC opens the inlet to 40.1 %, holds the true level within 0.1 m of 11.2 m from 1.0 s on and ends 0.006 m above the setpoint, while the proportional loop ends 0.899 m short, against the 0.902 m of Equation (2.21) at the depressed level, and the reference stays within 0.1 m from 4.05 s. The true-level IAE is 4.77, 48.51 and 11.35 m s and the travel 91.6, 191.1 and 190.1 % for the MPC, the shipped loop and the reference. A gain wrong by a factor of five costs the MPC little, because the estimator turns the mismatch into a load that the plan can see. Raising the move suppression from the shipped 0.03 to $\lambda = 1$, the top of its range, raises the IAE of the same run from 6.32 to 8.64 m s.

On the challenge the MPC repeats the first 20 s of its case 4 run, since it does not read K_c. Its worst measured error over the judged last third is 0.241 m, against 2.97 m for the shipped $K_c = 2\,\%/\mathrm{m}$, which fails, and 0.946 m for the suggested $K_c = 10\,\%/\mathrm{m}$, which passes but still ends 0.837 m low.

The MPC solved 120 plans in each 60 s case and 40 in the challenge, and none of the 520 was rejected. Every plan of cases 3, 4, 9 and 10 was the unconstrained optimum, so the advantage there comes from prediction and estimation rather than from constraint handling. The mean logged solve time was 0.3 to 0.6 ms per plan under Node.js and 0.54 to 0.64 ms in Chrome on case 9, at most 5.8 ms, or 0.11 to 0.13 % of one processor core in real time. These are wall-clock values on one desktop computer and vary between runs. The MPC loses none of the five cases and ties case 8, and

Table 2.8 Level Loops on Cases 3 and 4 Over a 180 s Run: Shipped Proportional, Counterfactual and Relay-Tuned PI, and the MPC, One Noise Seed

Case	Controller	IAE (m s)	IAE_h	e_{max} (m)	Travel (%)	$Travel^0$
3, setpoint	P, shipped	9.49	4.53	—	527.9	22.7
	PI[c]	9.92	4.49	—	526.0	23.0
	PI, relay	11.34	5.80	—	339.2	15.7
	MPC	8.06	2.16	—	123.3	60.0
4, outlet	P, shipped	144.7	145.1	0.860	515.4	8.4
	PI[c]	18.01	13.66	0.694	513.9	9.7
	PI, relay	18.93	13.93	0.916	333.3	10.7
	MPC	10.86	5.23	0.650	73.8	13.4

P, shipped: $K_c = 10$, $K_I = 0$. PI[c]: $K_c = 10$ with the relay's $K_I = 0.675$, a counterfactual that is neither shipped nor written. PI, relay: $K_c = 6.53$, $K_I = 0.675$. IAE uses the measured level, IAE_h the true level on the 0.05 s step, e_{max} is the largest true deviation after the outlet step, and $Travel^0$ repeats Travel with $\sigma_n = 0$. The two estimators give the same MPC rows.

most of its gain on cases 4, 9 and 10 is the integral action that the shipped tunings lack.

2.6.6 Estimator Choice and Relay Autotuning

These runs last 180 s, three trend windows, or 20 s for the challenge, with every setting entered before the run and the shipped noise seed, so their integrals exceed those of Table 2.7. Over the 1799 executions of case 9 under the MPC, the moving horizon and Kalman estimates differ by at most $6.1 \times 10^{-9}\sigma_n$ in the level and 3.2×10^{-9} % in $\hat{d}$. The residue is not rounding. The Riccati iteration that designed the filter stopped after 365 steps with its covariance a relative 5.4×10^{-10} short of the fixed point, which the window refines while the filter keeps its gain, and with the fixed-point covariance the difference is $8.9 \times 10^{-14}\sigma_n$. With each estimator in turn, the MPC runs of the ten cases give valve records within 5.1×10^{-9} % of each other, with no window rejected and no bound active. The Kalman estimate stays inside 0 to 12 m in every case, the lowest true level being the 0.40 m of the manual case 5, so no case separates the estimators and no trend of them is shown.

Against the true level, read only by the measuring script, both estimators follow case 3 with a root-mean-square error of 0.012 m and a largest error of 0.078 m, and case 9 with 0.070 m and 0.452 m. The largest error stays at or below 0.078 m in the five cases that keep the outlet setting and inlet coefficient of the fit, and reaches 0.45 to 0.61 m in the five that change either. An estimate took a median of 247 µs and a 99th percentile of 538 µs under Node.js. In Chrome, measured during case 9 on the animation loop that drives the application, the estimator and the planner together occupied at most 0.95 percent of a single core at normal speed and 7.3 percent at the tenfold speed, the higher reading of two runs taken at a load average of 2.5 to 3.1, which uses nearly three quarters of the budget of one tenth of a core.

The relay was started 6 s into case 3 with its setpoint step removed, so the tank held 10 m on the 2.21 % valve that holds that level. The relay was centered on $u_0 = 2.25$ %, and δ_r was limited to 2.03 %, 0.9 of the way to the closed valve. The experiment converged after 20 cycles and 48.0 s with $\alpha = 0.193$ m, $D = 0.55$, $P_u = 4.40$ s and $K_{cu} = 20.9$ %/m, and it wrote $K_c = 6.53$ %/m and $K_I = 0.675$ %/(m s) in place of the shipped 10 and 0. With $\varepsilon_r/\alpha = 0.776$, ψ is 51° and the result is flagged. Ten further noise seeds, 1001 to 1010, all converged, with K_{cu} from 14.9 to 51.1 %/m, P_u from 2.60 to 5.96 s and ψ from 42° to 72°. Started instead at 60 s of case 3, after the level has settled at 11.2 m, the relay converges in five cycles on $K_{cu} = 17.5$ %/m and $P_u = 3.41$ s with $\psi = 45°$, and it writes $K_c = 5.48$ %/m and $K_I = 0.732$ %/(m s), again a weaker proportional gain than the shipped one. Repeated after Reset with $\sigma_n = 0$, it returns 90.6 %/m at 5.80 s with $\psi = 79°$ and writes $K_c = 28.3$ %/m.

None of these values is the ultimate point of the tank. The model of Equation (2.16) is a lag with no dead time, and sampled with a zero-order hold at Δt_c its phase reaches −180° only at the Nyquist frequency, where $K_{cu} = (1 + a)/b = 400$ %/m and $P_u = 2\Delta t_c = 0.2$ s. A relay with hysteresis cycles

where the imaginary part of the sampled frequency response G of the model equals $-\pi\varepsilon_r/(4\delta_r)$, and the same relay on the noise-free sampled model returns 107 %/m at 6.0 s, so the band places the cycle. Noise widens the swing, and at $\varepsilon_r/\alpha = 0.776$ Equation (2.20) turns a one percent error in α into a 2.5 percent error in K_{cu}, and the seeds spread K_{cu} threefold. The measured K_{cu} is 0.052 of the sampled ultimate gain and 0.195 of the noise-free relay result. The same mechanism separates the two runs started at 60 s: their swings differ by a factor of only 1.38, yet their gains differ by a factor of 5.2, because the noise-free swing lies just outside the band, where $\sqrt{\alpha^2 - \varepsilon_r^2}$ in Equation (2.20) is small.

The written constants improve only the outlet step in Table 2.8. On the setpoint step the true-level IAE rises from 4.53 to 5.80 m s, and the shipped K_c with the written K_I gives 4.49 m s, so the loss comes from the smaller K_c. On the outlet step the IAE falls from 145.1 to 13.93 m s, and the run ends 0.005 m from the setpoint where the shipped loop parks 0.841 m low. Integral action accounts for all of that gain, since the shipped K_c with the written K_I reaches 13.66 m s, and the smaller K_c raises the peak deviation from 0.694 to 0.916 m for 35 percent less valve travel. That travel is mostly transmitter noise passed through K_c, as Equation (2.15) predicts. It falls to 0.64 and 0.65 of the shipped loop's travel on the two cases, close to the ratio $6.53/10$ of the gains, and with $\sigma_n = 0$ it falls from 339 to 15.7 % on case 3 and from 333 to 10.7 % on case 4. On the outlet step the relay-tuned loop has 2.7 times the true-level IAE of the MPC, 13.93 against 5.23 m s, with 4.5 times its valve travel.

2.7 Technology in Practice

The loop of Section 2.3 reads LT through white noise and moves the inlet valve the instant it is commanded. Figure 2.5 draws it with the devices a plant tank would carry in the place of the schematic's symbols.

What Measures It

Three principles serve a vessel like TK-101. A hydrostatic transmitter near the floor reads the pressure of the liquid column. That pressure gives the height only while the density is known and constant, although in a vessel of constant area it measures the mass of liquid whatever its density [17]. A displacer weighs a body hanging in the liquid and cannot report a level above the top of that body [10]. An echo transmitter times a wave reflected from the surface [17]. No range, accuracy or response time is given here, because none was found in an open source that names no maker. The noise σ_n of Equation (2.4) stands for the turbulence of the surface. The model's transmitter has no density error, no span limit and no fouling, so $\hat{h}$ and h differ only by that noise.

> **TECHNOLOGY** **Sensor: A Guided-Wave Radar Level Transmitter**
>
> A guided-wave radar transmitter sends microwave pulses down a metal probe that hangs into the liquid and times the echo from the surface. The wave travels through the vapor above the liquid, so the calibration rests on the speed of light in that vapor, which shifts when the vapor's pressure or temperature changes. A probe may be a single rod, a pair of rods or a rod inside a tube. The coaxial form guides the energy best, and the single rod tolerates fouling best. Sticky deposits on a probe can return echoes that the transmitter takes for the liquid surface, so the reading reports the deposit as the level [17].

What Moves It

The inlet valve of Equation (2.1) is linear and has no friction. A plant inlet is a throttling valve with a pneumatic actuator, usually fitted with a positioner that corrects the difference between the commanded and the actual stem position [17]. Its failure position follows from the hazard, not from habit: an air-to-open valve fails closed, and the risks of the process should decide which way it fails [17]. On a tank that can overfill, a valve that shuts when its air fails stops the inflow. The outlet of TK-101 is a hand valve, the disturbance of every case.

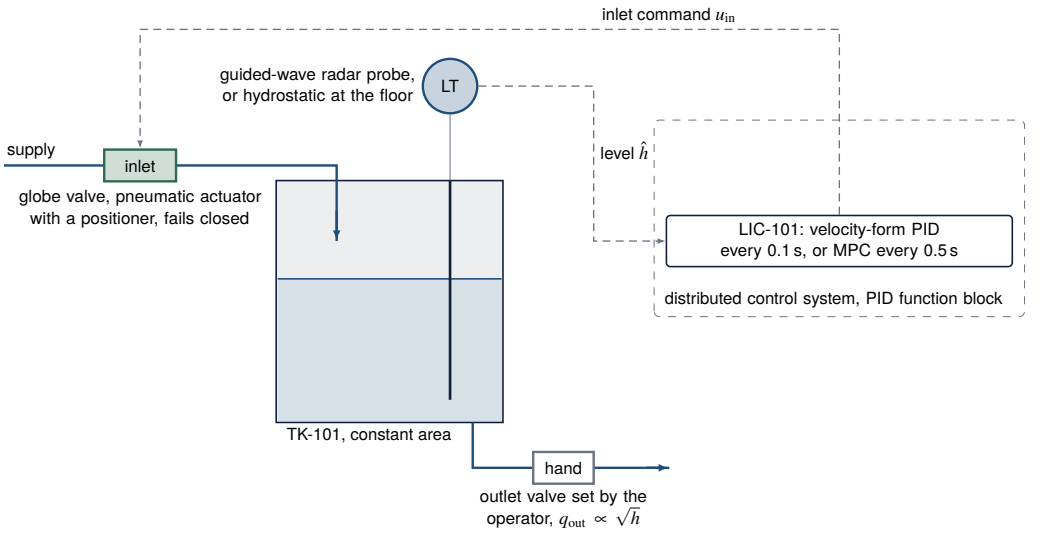


Figure 2.5 The level loop of the chapter with plant devices in place of the schematic's symbols. LT reads the level from the roof with a guided-wave radar probe, or from the floor by hydrostatic pressure, and LIC-101 on the control platform moves the inlet valve through its positioner. Notice that the outlet is a hand valve the loop never moves, so with the outlet shut the loop can add liquid but has no way to remove it.

> **TECHNOLOGY** **Final Element: A Globe Valve and Its Stem Friction**
>
> A globe valve is built for fine throttling and is widely used to regulate flow [25]. Packing around its stem adds friction, and friction at rest exceeds friction in motion. The stem of a pneumatic valve therefore sticks until the actuator pressure is high enough to break it free, jumps, and sticks again, so a smooth command becomes a staircase of stem positions. A reversal must first relax the friction in one direction and then overcome it in the other, which acts as a dead band [17]. On a trend, a sticky valve shows a sawtooth controller output and a process variable that cycles around its setpoint [25].

Where the Controller Runs

A distributed control system runs production at one site in refining, chemical manufacturing, water treatment and power generation. Its loops run with cycle times from milliseconds to minutes [26]. LIC-101 executes every 0.1 s and the predictive controller plans every 0.5 s, both inside that range. The velocity-form PID of Equation (2.13) is the PID function block of such a platform, with the manual station and the bumpless transfer of Section 2.5.

> **IN PRACTICE** **Test the Valve Before Tuning the Loop**
>
> An instrument handbook advises a test of the valve before any tuning. With the controller in manual, the output is stepped up and down by small amounts, and a valve with excess friction shows up as a process variable that fails to move after a reversal. Valves with poor maintenance, or without a positioner to correct the stem, are the worst. Tuning cannot remove the cycling that friction causes: a detuned loop cycles more slowly but still cycles. An integral dead band stops the cycling at the price of a standing error no larger than the band [17].

What Is Modernizing It

Three developments touch this loop. The first is the echo instrument. Some industries replace old displacer level instruments with guided-wave radar, even at high pressure, and a common retrofit

installs the radar probe in the cage that once held the displacer [17]. The second is the smart valve. Many modern refineries use valves with position feedback, self-diagnostics and health monitoring, although most valves still lack them [25]. The third is data-driven diagnosis of the loop from the controller output and the process variable alone, the subject of the open challenge at the end of Section 2.9 [27].

2.8 Limitations

The simulator is a teaching instrument and not a design tool. A single well-mixed inventory with a quasi-steady discharge cannot represent outlet line inertia, surface turbulence, vortexing at the drain or any two-phase behavior, so it must not be used to size a drain or a surge volume. The vessel is a vertical cylinder of constant area, which removes the geometric nonlinearity that a horizontal or a conical vessel adds on top of the hydraulic one. The actuators are ideal, so stiction, hysteresis and stroking time are absent, and the limit cycle they cause in industrial level loops cannot be produced. The transmitter carries white noise and nothing else, so the low-frequency errors that dominate a real level measurement are not represented. Mass is not conserved during an overflow. There is no cascade to a flow controller, no second tank, and no way to make the outlet flow independent of the level other than closing the valve entirely, so the pumped-outflow case is available only as the zero-flow limit rather than as a nonzero fixed discharge. Two gravity-drained tanks in series are the plant of Chapter 3. All ten shipped cases start at steady operation, so a start-up from an empty tank has to be set up by hand, by shutting the inlet and opening the outlet before the run as in Figure 2.2. The noise source is reseeded whenever a case is loaded or reset, so the noise realization is the same for every user, and a result should not be treated as an average over realizations. Each simplification that was retained, the square-root discharge, the two hard limits, the level-dependent gain and the short control period, was retained because removing it removes a behavior the loop must be tuned around. Section 2.3.4 states where each assumption fails.

The predictive controller plans on a first-order pair fixed at 10 m with the outlet 10 % open. It leaves out the square-root outflow, so its gain and time constant are five times too large with the outlet half open, and it leaves out the outlet valve, the overflow and the dry limit, so no plan protects the rim. Its Kalman gain assumes the default 0.05 m of noise and is not recomputed when σ_n or a plant constant changes, its λ sits at the lowest value of the shared tuning sweep, and its comparison uses one noise realization and excludes the shut-outlet tank, an integrator that the self-regulating model does not describe. The bounds of the moving horizon estimator apply to the states of the same first-order model, so where the model is wrong the estimate is wrong inside them, and no case on this tank activates them. The relay result describes the experiment as much as the vessel, because the phase of a delay-free lag approaches -180° only through the sampling, and the Tyreus–Luyben rule is one of several published conversions.

2.9 Discussion

Level loops carry a share of plant safety that is large relative to the attention they receive. The two disturbances the application offers, a drain that opens and a drain that closes, bracket a real hazard. A drain that closes while the inlet keeps filling is an overfill, and an overfill of a storage tank produced the vapor cloud at Buncefield, where the release rate and the cascade of liquid down the tank wall are what generated the flammable inventory rather than any subsequent process upset [8]. The refinery tower of the history box in Section 2.2 was overfilled while its level indication read within its span, and a dynamic reconstruction of the sequence shows how far the true inventory had diverged from the instrument before anything else went wrong [28]. Both events turn on a distinction the application makes visible in its simplest form, that the level a controller acts on is $\hat{h}$ and the level that overflows the tank is h.

The opposite requirement is just as real and pulls the other way. A surge vessel is installed to absorb variation, and a level controller tuned for tight regulation converts an inlet flow disturbance

into an outlet flow disturbance of the same size and hands it to the next unit. Averaging level control exists to spend the available tank volume instead, and the tuning parameter of a well-posed averaging controller is the level swing that is allowed rather than a gain [1, 12]. Consistency of the inventory loops across a plant, in the sense that each holdup is regulated by a flow in a direction compatible with how throughput is set, is a structural requirement that no amount of local tuning can supply [13]. The shut-outlet experiment of Table 2.4 is the smallest instance of that structural problem, in which one inventory has no flow available to regulate it in the required direction.

OPEN CHALLENGE **Finding the Sticky Valves in a Plant**

Valve stiction is a common fault in process plants, and it brings oscillation, actuator wear and higher maintenance cost. Many plants still run conventional valves without position feedback or self-diagnostics, so the only signals available are the controller output and the process variable, and detection rests on manual inspection and operator judgment [25]. Detectors trained on simulated loops often fail on real ones, because simulation cannot fully reproduce plant data [27]. Progress would be a detector that works on any loop from those two routine signals and warns well before the valve fails, which one study on refinery data has begun to show, predicting stiction up to four hours ahead [25].

2.10 Conclusion

This chapter studied a gravity-drained tank as an interactive laboratory in which the plant model is a function of the inventory. The simulator integrates one state under a mass balance whose discharge follows the Torricelli square-root law, with a linear inlet valve, hard limits at the overflow height and at the dry bottom, and a transmitter carrying 0.05 m of Gaussian noise. Linearization gives a first-order model whose gain and time constant both scale as the square root of the operating level while their ratio stays fixed at 0.05 m/s per percent of inlet travel, and step tests in the running application confirm the scaling with an identified gain rising by a factor of 2.14 and an identified time constant by 2.12 between 2.04 m and 9.88 m against a predicted factor of 2.20. The identified time constant equals the drain time of the same vessel, measured at 19.65 s from a full tank against the 19.79 s of the analytic expression.

The same outlet upset applied at two setpoints leaves a proportional loop 0.411 m short at 10 m and 0.212 m short at 3 m with identical gains, a ratio of 1.94 against the 1.90 of the linearized prediction, and integral action removes both at the cost of 0.9 % and 0.4 % of extra valve travel. The default controller is a velocity-form PID with derivative on the measurement, whose integral alone is corrected at the valve limits by back-calculation and whose manual tracking fixes the effective bias of a proportional loop at the moment of transfer. Its short 0.1 s sample period makes a derivative gain of 1 % s/m contribute as much valve noise as the entire default proportional gain.

A linear predictive controller fitted at 10 m with the outlet 10 % open removes the offset through its disturbance estimate when the outlet opens to 50 %, ending case 9 within 0.01 m of its new setpoint with a true-level IAE of 4.77 m s against 48.51 m s and less than half of the valve travel of the shipped proportional loop, which ends 0.899 m short. A moving horizon estimator offered in place of its filter gives valve records within 5.1×10^{-9} % of the filter's on all ten cases, because no bound becomes active, and a relay autotune lowers the true-level IAE of the outlet step from 145.1 to 13.9 m s through the integral action it adds.

Closing the outlet valve turns the same vessel into a pure integrator that no inlet valve can bring back down, which places both textbook level models in one apparatus. A fuller model would add a fixed-rate pumped discharge as a third configuration, a nonconstant vessel cross-section, valve stiction, a predictive controller that spends the surge volume against the overflow constraint, and a second tank in series so that averaging control has somewhere to send the disturbance. The tank is an accessible platform for level experiments in which the tuning the loop requires depends on the inventory the vessel holds.

KEY POINTS

- A tank that drains under its own head settles at a level of its own, and a fuller tank answers the same valve move more strongly and more slowly.
- Gain and time constant both grow as the square root of the level, and their ratio, the initial slope of every inlet step, stays at 0.05 m/s per percent.
- The same proportional loop leaves 0.411 m of offset at 10 m and 0.212 m at 3 m, and integral action removes both.
- With the outlet shut the tank is an integrator: a proportional loop has no offset, and no inlet valve can bring the level back down.

Where else this applies. Volume 2, Chapter 2 has the same fixed ratio of gain to time constant, and Chapter 9 moves between the same two characters.

References

[1] K. A. McDonald, T. J. McAvoy, and A. Tits, "Optimal averaging level control," *AIChE Journal*, vol. 32, no. 1, pp. 75–86, 1986.

[2] P. J. Campo and M. Morari, "Model predictive optimal averaging level control," *AIChE Journal*, vol. 35, no. 4, pp. 579–591, 1989.

[3] W. L. Luyben, "Liquid level control: Simplicity and complexity," *Journal of Process Control*, vol. 86, pp. 57–64, 2020.

[4] 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.

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

[6] K. H. Johansson, "The quadruple-tank process: a multivariable laboratory process with an adjustable zero," *IEEE Transactions on Control Systems Technology*, vol. 8, no. 3, pp. 456–465, 2000.

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

[8] G. Atkinson, S. Coldrick, S. Gant, and L. Cusco, "Flammable vapor cloud generation from overfilling tanks: Learning the lessons from Buncefield," *Journal of Loss Prevention in the Process Industries*, vol. 35, pp. 329–338, 2015.

[9] O. Mayr, "The origins of feedback control," *Scientific American*, vol. 223, no. 4, pp. 110–118, 1970.

[10] U.S. Chemical Safety and Hazard Investigation Board, "Investigation report: Refinery explosion and fire, BP Texas City, Texas, march 23, 2005," U.S. Chemical Safety and Hazard Investigation Board, Washington, DC, Report 2005-04-I-TX, 2007. [Online]. Available: https://www.csb.gov/assets/1/20/csbfinalreportbp.pdf

[11] T.-F. Cheung and W. L. Luyben, "Nonlinear and nonconventional liquid level controllers," *Industrial & Engineering Chemistry Fundamentals*, vol. 19, no. 1, pp. 93–98, 1980.

[12] D. Sbarbaro and R. Ortega, "Averaging level control: An approach based on mass balance," *Journal of Process Control*, vol. 17, no. 7, pp. 621–629, 2007.

[13] E. M. B. Aske and S. Skogestad, "Consistent inventory control," *Industrial & Engineering Chemistry Research*, vol. 48, no. 24, pp. 10 892–10 902, 2009.

[14] S. Skogestad, "Simple analytic rules for model reduction and PID controller tuning," *Journal of Process Control*, vol. 13, no. 4, pp. 291–309, 2003.

[15] K. J. Åström and T. Hägglund, *Advanced PID Control*. Research Triangle Park, NC: ISA – The Instrumentation, Systems, and Automation Society, 2006.

[16] J. Njock Libii, "Mechanics of the slow draining of a large tank under gravity," *American Journal of Physics*, vol. 71, no. 11, pp. 1204–1207, 2003.

[17] 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.

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

[21] 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.

[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] 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.

[24] ——, "The future of PID control," *Control Engineering Practice*, vol. 9, no. 11, pp. 1163–1175, 2001.

[25] N. Promsricha, C. Chatpattanasiri, N. Kerdgongsup, and S. Balabani, "Machine learning-based framework for real time detection and early prediction of control valve stiction in industrial control systems," arXiv preprint 2601.12362, 2026.

[26] 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, Gaithersburg, MD, Special Publication 800-82r3, 2023.

[27] S. K. Damarla, "Optimal transport image representation and deep covariance alignment (CORAL) for control valve stiction detection," arXiv preprint 2607.22486, 2026.

[28] D. Manca and S. Brambilla, "Dynamic simulation of the BP Texas City refinery accident," *Journal of Loss Prevention in the Process Industries*, vol. 25, no. 6, pp. 950–957, 2012.

2.11 Exercises

2.1 Concept check · 2 min

The inlet valve of the default tank is stepped up by 1 % three times: from a settled 2 m with the outlet wide open, from a settled 10 m with the outlet wide open, and from a settled 10 m with the outlet 10 % open. How does the rate at which the level starts to rise compare across the three tests?

(a) It is the same in all three, and only the final rise and the time to reach it differ

(b) It grows with the square root of the level, so it is largest in the two tests at 10 m

(c) It is largest with the outlet 10 % open, because the process gain is largest there

(d) It is the same at both levels with the outlet wide open, and smaller with the outlet nearly shut

2.2 Concept check · 2 min

The same proportional loop, $K_c = 10\,\%/\mathrm{m}$ with $K_I = K_D = 0$, holds the tank on setpoint at 3 m in one run and at 10 m in another, both with the outlet 50 % open. The outlet then opens to 70 %. How do the two standing errors compare?

(a) They are equal, because the controller gain and the outlet step are the same

(b) The error at 10 m is about twice that at 3 m, as K_d grows faster than K_u

(c) The error at 3 m is the larger one, because the loop gain $K_u K_c$ is smaller there

(d) Neither run keeps an error, because the drain makes the tank regulate itself

2.3 Concept check · 2 min

Case 3 is run with the outlet valve shut before Run, a proportional-only loop at $K_c = 10\,\%/\mathrm{m}$ and the default 0.05 m of transmitter noise. The setpoint moves to 11.2 m at 5 s. Where does the true level stand five minutes into the run?

(a) On the setpoint, because a proportional loop on an integrator has no offset

(b) A little below the setpoint, because the valve is shut for part of every second

(c) About 0.12 m above it, because low readings open the valve and nothing drains

(d) At the 12 m rim, because the noise winds up the running sum of the PID

2.4 Theory and calculation · 8 min

The default tank has $A = 1\,\mathrm{m}^2$, an inlet coefficient $c = 50\,\mathrm{kg/s}$ per percent, $\rho = 1000\,\mathrm{kg/m}^3$ and an outlet coefficient $c_v = 0.35\,\mathrm{m}^3/\mathrm{s}$ per $\mathrm{m}^{1/2}$. The outlet is 50 % open.

(a) Find the inlet position that holds the level at 3 m. *Checkpoint:* 6.06 %.

(b) Find the time constant, the process gain and the disturbance gain of the linearized model at that point. *Checkpoint:* 19.8 s, 0.990 m/% and −0.12 m/% of outlet travel.

(c) A proportional loop with $K_c = 10\,\%/\mathrm{m}$ holds the level at 3 m when the outlet opens to 70 %. Predict the standing error, and the level the tank would reach with the inlet left where it was. *Checkpoint:* 0.220 m, and 1.53 m.

2.5 Theory and calculation · 5 min

The default tank ($A = 1\,\mathrm{m}^2$, $c_v = 0.35\,\mathrm{m}^3/\mathrm{s}$ per $\mathrm{m}^{1/2}$) is full at 12 m. The inlet is shut and the outlet opened wide.

(a) Integrate Equation (2.1) to find the time to empty the tank, and compare it with the time constant of Equation (2.6) at the starting level. *Checkpoint:* 19.79 s for both.

(b) Find the times at which the level passes 9, 6 and 3 m. *Checkpoint:* 2.65, 5.80 and 9.90 s.

2.6 Theory and calculation · 7 min

Case 4 starts at 10 m with the outlet 10 % open and the inlet at the 2.214 % that holds it, under a proportional loop with $K_c = 10\,\%/\mathrm{m}$ and no integral action. At 5 s the outlet opens to 50 %, where holding 10 m needs 11.068 %.

(a) Predict the offset with Equation (2.21), taking the bias as the starting inlet position. *Checkpoint:* 0.885 m.

(b) The loop settles below 10 m, where less inlet is needed. With the outlet half open the inlet that holds a level h is $3.5\sqrt{h}$ percent. Find the level at which the proportional loop supplies exactly that inlet. *Checkpoint:* 9.162 m, an offset of 0.838 m.

(c) The same loop runs case 3 with the outlet shut before Run, and nothing opens it. Find the offset with no transmitter noise. *Checkpoint:* Zero.

2.7 App challenge · 7 min

Load scenario 2, Open-loop actuator step, and change no setting. The case starts in Manual at 10 m with the outlet 10 % open, and the inlet steps from 2.214 to 1.214 % at 5 s. Run to 35 s at the default seed and export the CSV file.

(a) From the true level, column `true_pv`, in the rows at 5.0 and 6.0 s, estimate the rate at which the level starts to fall. *Checkpoint:* 0.050 m/s, within 0.002 m/s.

(b) Read the true level at 35 s, and compare it with the linearized model of Equation (2.5) at this point. *Checkpoint:* 8.62 m, within 0.01 m, for both.

(c) Find the level at which the tank will settle, and the final level the linearized model predicts. *Checkpoint:* 3.01 m against 0.96 m.

2.8 App challenge · 8 min

Load scenario 4, Disturbance rejection. The outlet opens from 10 to 50 % at 5 s. Use the default seed, run each case for 60 s and export the CSV file of each.

(a) Run the scenario as shipped, a proportional loop with $K_c = 10\,\%/\mathrm{m}$. Read the IAE and the MV travel tiles of the scorecard, and find the mean true level over the last 5 s. *Checkpoint:* IAE 44.52 m s and travel 176.0 %, each within 0.2. Mean level 9.165 m, within 0.01 m.

(b) Press the reload button, set the Integral gain to 1 %/(m s) and run again. Read the same three values. *Checkpoint:* IAE 9.65 m s, travel 175.4 %, mean level 10.00 m, within 0.01 m.

Compare the offset of run (a) with the prediction of Equation (2.21), and say what the integral action cost.

2.9 App challenge · 9 min

Load scenario 9, MPC on a tank it was not fitted to. The outlet opens from 10 to 50 % at 5 s and the setpoint moves to 11.2 m at 30 s. Use the default seed and run each case for 60 s. The exported file begins with comment lines marked `#`.

(a) Run the scenario as shipped, with the Controller at MPC. Read the IAE tile, and the disturbance estimate, column `est_d`, in the row at 29.8 s. *Checkpoint:* 6.32 m s and −8.82 %.

(b) Set the Controller to PID, press Reset and run again. Read the IAE tile and the mean true level over the last 5 s. *Checkpoint:* 48.42 m s and 10.30 m.

(c) Set the Controller back to MPC, set the Move suppression (Advanced level) to 1, press Reset and run again. Read the IAE tile. *Checkpoint:* 8.64 m s.

Accept an IAE within 0.1 m s, an estimate within 0.1 % and a level within 0.01 m. Explain how a model five times too sensitive still ends without offset.

2.10 Application area · 4 min

The default outlet has $C_d A_o = 0.0790\,\mathrm{m}^2$, which is 7.9 % of the tank cross-section. A plant tank of the same $1\,\mathrm{m}^2$ drains through an orifice of 50 mm diameter with $C_d = 0.6$.

(a) Find its outlet coefficient c_v with Equation (2.2), taking $g = 9.81\,\mathrm{m/s^2}$. *Checkpoint:* $0.005\,22\,\mathrm{m^3/s}$ per $\mathrm{m}^{1/2}$.

(b) Find its time constant at 10 m with the outlet wide open, and compare it with that of the default tank. *Checkpoint:* About 1210 s, or 20 min, against 18.1 s.

2.11 Application area · 3 min

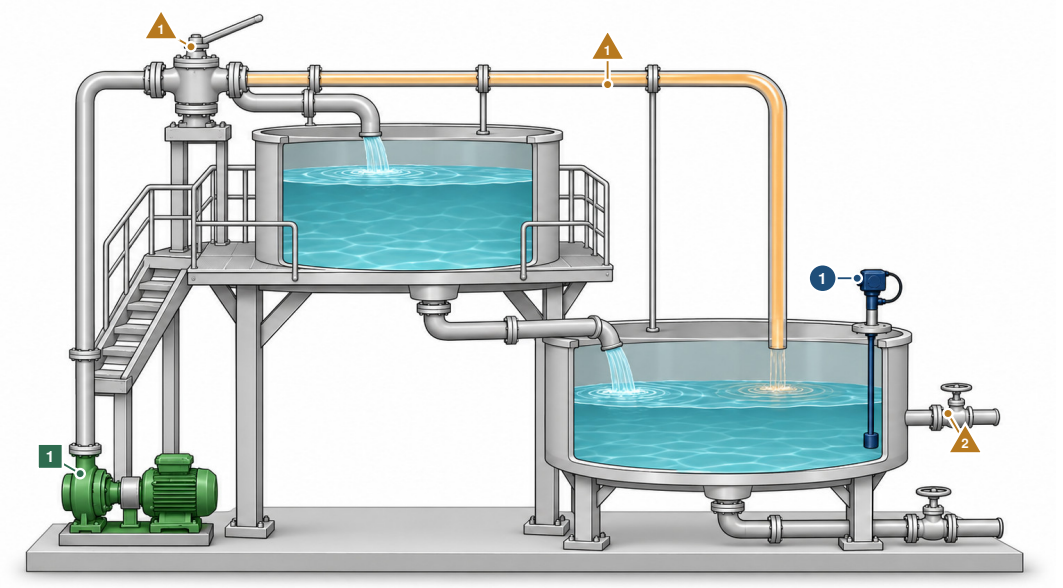
A surge drum feeds a downstream column, and a manager asks for its level loop to be tightened, citing Table 2.5: integral action removed the offset for under 1 % of extra valve travel. Give two reasons from this chapter why that table does not support the request, and name what the app would need in order to show the trade.

2.12 Review the agent · 5 min

An AI assistant is asked how to tighten the level loop of case 4, which runs with $K_c = 10\,\%/\mathrm{m}$ and $K_I = 1\,\%/(\mathrm{m\,s})$, and replies as follows. “Add derivative action, $K_D = 5\,\%\,\mathrm{s/m}$. Derivative action anticipates the error and adds phase lead, and the controller executes every 0.1 s, so the derivative always sees fresh data. The level will stay closer to its setpoint for about the same valve movement.” Find the flaw in the proposal and say how it could be checked.

CHAPTER 3

Two-Tank Level Control With a Splitter Valve That Sets the Plant Order



The controller moves 1 the pump that feeds both paths

The controller reads 1 the level in the lower tank

What disturbs it 1 the splitter, set by hand, and the bypass it feeds 2 a draw-off from the lower tank

Notice that the bypass reaches the lower tank at once, while the other path must first fill the upper tank, whose own level indicator is not drawn. Illustration, not to scale.

IN PLAIN TERMS **One Pump, Two Tanks and a Choice of Path**

The main idea. A pump fills a top tank that drains into a bottom tank, and a controller holds the bottom tank at its target depth. A hand valve can send part of the water straight to the bottom tank.

The challenge. Whichever path the water takes, the bottom tank ends at the same depth, so a slow look shows no change. But the path through the top tank is slower, and settings tuned for one path can make the level swing on the other.

Why it matters. A plant that reroutes its flow for repairs can keep a controller tuned for a path that is gone.

THREE WAYS TO READ THIS CHAPTER

Explore	11 min	the first page, the fact on five tanks in series (Section 3.2), the history of the draining tank (Section 3.3), the try-it segments of cases 2 and 9 (Sections 3.4 and 3.8), the fact on the draw-off (Section 3.7) and the key points
Study	70 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	20 min	the first page, the results (Sections 3.6 and 3.7), Technology in Practice (Section 3.9), Limitations (Section 3.10) and the open challenge

The control question A splitter valve that sends pump flow past the top tank leaves the bottom level and the process gain where they were, yet turns a second-order plant into a first-order one. What does a step test read at each splitter setting, and what does a tuning or a controller model carried from one lineup to the other do?

Learning objectives

- Derive the steady levels, the pump bias and the linearized time constants of two non-interacting tanks from their mass balances.
- Explain why the splitter changes the poles and the zero of the pump transfer function but not its gain.
- Fit a gain, time constant and apparent dead time to a step test, and tell a secant gain from a tangent gain.
- Predict the proportional offset and the reachable level under a draw-off, and the top-tank level that rejecting it requires.
- Judge a tuning or a predictive controller's model against the plant structure it was identified on.

Prerequisites Sections 1.4.2 and 1.4.3 (balances and linearization), 1.4.5 (two tanks in series), 1.4.6 (half rule) and 1.4.8 (two-point fit), and Chapter 2 for one gravity-drained tank. Sections 1.6.2 and 1.6.3 are needed for the controllers.

The app https://apmonitor.com/apps/2tank/

See also Volume 2, Chapter 4 for step tests on a plant whose model moves with load, Volume 2, Chapter 6 for gain scheduling and Volume 2, Chapter 3 for a bound set by the supply. Volume 9, Chapter 3 for predictive controllers that differ only in their step test, and Volume 9, Chapter 4 for offset-free estimation.



3.1 Introduction

A controller is commissioned against a plant, and the identified model reflects the piping configuration that was in service on the day of the step test. Operating-point dependence is taught carefully, because a nonlinear gain moves with throughput and every process control course shows how to recognize it. Structural change receives less attention. A bypass opens, a vessel is taken out of service, a recycle is lined up, and the loop that used to see two lags in series now sees one. The steady-state map may not move at all when that happens, so the operator has no reason to expect a change, the trend looks the same in the long run, and the only symptom is that a controller tuned against the previous structure exhibits oscillation.

Two tanks in series with a splitter valve are the smallest apparatus in which that structural change can be produced deliberately. One pump feeds a three-way valve. Part of the flow goes to a top tank that drains into a bottom tank, and the rest bypasses the top tank and enters the bottom tank directly. Whatever the split, all of the pump flow leaves through the same bottom drain, so the resting level in the bottom tank depends on the pump position alone. The split changes only the dynamic path from the pump to the bottom tank. With the bypass shut the pump must fill the top tank before the bottom tank moves at all, and with the bypass wide open the top tank empties and stays empty. The same pump step therefore produces the same final level through a plant that is second order in one case and first order in the other.

The model behind the apparatus is a mass balance on each vessel with gravity discharge through a fixed restriction as the driving force, which is the standard construction for tank dynamics in process control texts [1, 2]. The experiments built on it cover process order, apparent dead time from lags in series, graphical identification, internal model control tuning, proportional offset, integral action, anti-windup and gain scheduling, and the limits they expose follow from the bound that effective delay places on the bandwidth of any linear feedback loop [3, 4]. The behavior matters industrially because bypasses, spared vessels and seasonal lineups change plant structure without changing the material balance, so a tuning carried across such a change degrades without any steady-state symptom.

Coupled tanks are already the standard apparatus for teaching multivariable control, and the quadruple-tank rig makes exactly this kind of point by moving a transmission zero across the

imaginary axis with a pair of manual splits [5]. However, that experiment needs two pumps and two controlled levels before the effect appears, and the structural change it demonstrates is a property of the multivariable zero rather than of the plant order that a single loop has to be tuned against. This chapter works with a two-tank simulation in which the splitter fraction converts a second-order overdamped level plant into a first-order one without disturbing the operating point. It quantifies what a controller tuned at either extreme of that fraction does at the other.

Section 3.2 places two tanks in the plantwide and tuning literature. Section 3.3 writes down the simulator in enough detail (two states, two outflow laws, three disturbances, four hard limits, an instrument model and a numerical scheme) for an independent build to reproduce its trajectories. Section 3.4 shows that the pair is strictly lower triangular and non-interacting, that its upper time constant is the lower one scaled by the share of flow still passing the top tank, and that the splitter reaches the bottom level through a transfer function with a zero at the origin. It then reads one gain and three dynamic models from step tests at three splitter positions. Section 3.5 describes the velocity-form PID and the linear model predictive controller that can replace it in automatic, with its moving horizon estimator, its table of models selected by the splitter setting and a relay test that proposes PI constants from a limit cycle. Section 3.6 runs the tunings identified at the two ends of the splitter against both structures. Section 3.7 treats the draw-off and the leak, which move the steady state, and shows that the top tank caps the load a fixed lineup can reject. Section 3.8 turns the splitter into a test of a controller model as well as of a tuning. Section 3.9 names the devices a plant pair of tanks would carry. Sections 3.10 to 3.12 state the limitations, discuss the industrial setting and conclude, and Section 3.13 holds the exercises. One sequence of experiments thus produces identification, tuning, tuning mismatch and a controllability bound from the travel of a single valve.

3.2 Background

The reason two tanks appear in a control course at all is that process plants are built out of units that must be decoupled from one another. A vessel placed between two units is there so that a flow disturbance arriving from upstream does not have to be passed downstream at the moment it arrives, and the inventory in that vessel supplies the surge capacity that absorbs the disturbance. That requirement shapes the whole control structure of a plant rather than the tuning of one loop, because every holdup has to be regulated by some flow, the direction in which each holdup is regulated has to be consistent with how throughput is set, and the loops that do the regulating are chosen before any of them is tuned [6]. Level is where the structural question is most visible, because a level is an integral of the imbalance between two flows and cannot be left uncontrolled.

Placing vessels in series rather than in parallel has a specific dynamic cost. Each additional holdup contributes a lag, and lags in series accumulate into an apparent dead time that no controller can undo. The half rule of Equation (1.16) states that consequence quantitatively for tuning, and it is exactly the trade a second tank imposes on a level loop [4]. The limit is fundamental rather than a matter of tuning skill, because effective delay bounds the achievable bandwidth of any linear controller regardless of its structure [7]. The practical answer, when the plant structure itself is the thing that moves, is to schedule the controller against whatever variable indexes the change, and the gain-scheduling literature records both the appeal and the pitfalls of doing so [8, 9].

> **✦ DID YOU KNOW? Five Lags in Series in a Gold Circuit**
>
> A gold leaching circuit strings five stirred tanks in series, each a first-order lag on the one before it (Volume 15, Chapter 7). For tuning, such a chain is reduced to one lag and a delay by the half rule: the largest neglected time constant is split evenly between the effective delay and the smallest time constant kept [4].

Tank rigs became the laboratory vehicle for all of this because they are inexpensive, slow enough to observe in real time, and nonlinear. The quadruple-tank process uses two pumps and four vessels so that a pair of manual splits moves a multivariable zero between the left and right half planes [5], and

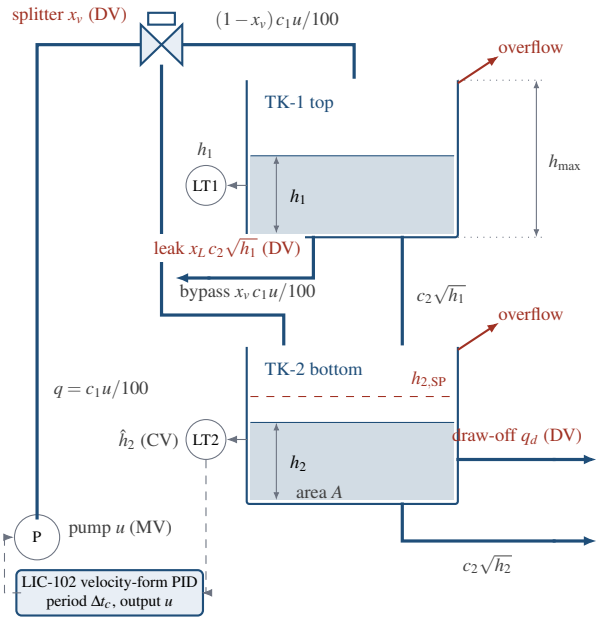


Figure 3.1 Two gravity-drained tanks in series behind one pump and a three-way splitter valve. The splitter sends the fraction x_v of the pump flow past TK-1 and straight into TK-2, TK-1 drains into TK-2 through $c_2\sqrt{h_1}$, and TK-2 drains to waste through $c_2\sqrt{h_2}$. LIC-102 controls the bottom level $\hat{h}_2$ with the pump. The splitter fraction, the bottom-tank draw-off q_d and the top-tank leak x_L are set by the operator and are never controlled.

it has served since as a plant for constrained and model-based control studies [10]. Two-tank rigs are equally standard, coupled through an orifice so that a backstepping controller has to be written for the pair [11], or driven over a network as a remote laboratory [12]. The textbook chapter that introduces tanks in series also separates interacting from non-interacting pairs [1], which decides whether the plant transfer matrix is full or triangular, and the apparatus of this chapter is the non-interacting case with a variable structure, a combination the existing rigs do not offer.

3.3 Process Model

The simulated plant is a pair of open, vertical, constant-area tanks. A single pump draws from an unlimited supply and discharges into a three-way splitter valve. The fraction $1 - x_v$ of the pump flow enters the top tank TK-1, the fraction x_v bypasses TK-1 and enters the bottom tank TK-2 directly, TK-1 drains by gravity into TK-2, and TK-2 drains by gravity to waste. Transmitter LT2 on the bottom tank reports the controlled variable, transmitter LT1 on the top tank gives an indication only, and the pump position is the manipulated variable, while the splitter fraction, an extra draw-off on TK-2 and an extra leak on TK-1 stay under operator control in manual and in automatic alike. Every symbol used below appears on the flowsheet of Figure 3.1.

Table 3.1 sorts the signals into their roles. There are two states and one actuator, so both tanks are driven by the same pump and no cascade exists. Four hard limits bound the states, an overflow limit and a dry limit on each tank. The plant is written with both levels in meters, the pump, the splitter and the leak in percent of travel, the draw-off as a level velocity in m/s and time in seconds. The splitter and leak settings enter the balances as the fractions x_v and x_L, each the setting divided by 100.

Table 3.1 Signal Roles, Ranges and Nominal Values at the 0.5 m Bottom-Tank Operating Point With the Splitter Shut and No Load

Symbol	Role	Nominal	Range	Unit
h_2	state, bottom level	0.5013	0–$h_{\max}$	m
h_1	state, top level	0.5013	0–$h_{\max}$	m
$\hat{h}_2$ (LT2)	CV, bottom level	0.5013	0–$h_{\max}$	m
$\hat{h}_1$ (LT1)	indication, top level	0.5013	0–$h_{\max}$	m
$h_{2,\mathrm{SP}}$	setpoint	0.50	0–$h_{\max}$, step 0.01	m
u	MV, pump position	35.4	0–100, step 0.1	%
$100\,x_v$	DV, splitter setting	0	0–100, step 1	%
q_d	DV, bottom draw-off	0	0–0.030, step 0.001	m/s
$100\,x_L$	DV, top-tank leak	0	0–100, step 1	%
σ_n	DV, LT2 noise 1σ	0.004	0–0.020	m
$\sigma_{n,1}$	DV, LT1 noise 1σ	0.004	0–0.020	m
overflow 1, 2	outputs, two flags	false	—	—

3.3.1 Level Balances

Both tanks have the same constant cross-sectional area A, and A is divided out of every term before it reaches the source, so the two balances are written directly as level velocities. Each is the mass balance of Equation (1.2) for one gravity-drained tank [1, 2], as shown in Equations (3.1) and (3.2).

$$\frac{\mathrm{d}h_1}{\mathrm{d}t} = c_1\,(1-x_v)\,\frac{u}{100} - c_2\,(1+x_L)\,\sqrt{h_1} \tag{3.1}$$

$$\frac{\mathrm{d}h_2}{\mathrm{d}t} = c_1\,x_v\,\frac{u}{100} + c_2\sqrt{h_1} - c_2\sqrt{h_2} - q_d \tag{3.2}$$

where c_1 is the pump coefficient in m/s at full pump, c_2 is the gravity-drain coefficient in $\mathrm{m}^{1/2}/\mathrm{s}$, u is the pump position in percent, $x_v \in [0,1]$ is the splitter setting over 100, $x_L \in [0,1]$ is the leak setting over 100 and $q_d \geq 0$ is the draw-off in m/s of bottom-tank level. Every term on both right-hand sides carries units of m/s. The outflow terms are the Torricelli form with the discharge coefficient, the orifice area, the tank area and $\sqrt{2g}$ all lumped into the single constant c_2, which is the only combination a controller can distinguish. Substituting the default constants gives $\mathrm{d}h_1/\mathrm{d}t = 0.0008\,(1-x_v)\,u - 0.04\,(1+x_L)\sqrt{h_1}$ and $\mathrm{d}h_2/\mathrm{d}t = 0.0008\,x_v\,u + 0.04\sqrt{h_1} - 0.04\sqrt{h_2} - q_d$ in m/s.

> ⧖ HISTORY **The Law of the Draining Tank, 1644**
>
> Evangelista Torricelli (1608–1648), the last student of Galileo, stated the law that drains both tanks of this chapter. Liquid leaving a small hole at the bottom of a vessel flows out as fast as a body that has fallen freely from the height of the liquid surface [13]. His work on water jets appeared in 1644, in a section on the motion of waters within a treatise on falling bodies, and it showed that every jet leaving a hole in a filled tube traces a parabola. It counts among the founding results of fluid mechanics [14]. Torricelli had no calculus. Written as a differential equation for the level, his law is now a standard exercise of a first course in differential equations [13].

The structure of Equations (3.1) and (3.2) establishes the property on which the rest of the analysis depends. The tanks are *non-interacting*. The discharge from TK-1 is $c_2\sqrt{h_1}$ and depends on the top level alone, because TK-1 drains into the vapor space of TK-2 rather than below its liquid surface, so nothing that happens in the bottom tank propagates back into the top one. The state matrix is lower triangular for that reason, and it stays lower triangular at every splitter setting. An arrangement in which a connecting orifice sits below both surfaces makes the discharge proportional to $\sqrt{h_1 - h_2}$ and produces a genuinely coupled pair with a full state matrix and a different control problem [11]. Neither vessel in this application is coupled in that sense.

3.3.2 Steady State and the Bias Term

Setting both derivatives to zero with $x_L = 0$ gives the resting levels of Equations (3.3) and (3.4).

$$\bar{h}_2 = \left[\frac{c_1\,\bar{u}/100 - q_d}{c_2}\right]^2 \tag{3.3}$$

$$\bar{h}_1 = (1 - x_v)^2 \left[\frac{c_1\,\bar{u}/100}{c_2}\right]^2 \tag{3.4}$$

The splitter fraction is absent from Equation (3.3) and appears squared in Equation (3.4). At steady state every unit of pump flow has to leave through the bottom drain no matter which path it took to get there, so the resting bottom level is a function of the pump position and the draw-off alone, while the resting top level falls as $(1 - x_v)^2$ and reaches zero when the bypass is wide open. Inverting Equation (3.3) at $q_d = 0$ gives the pump position that holds a given level, which is the bias of a positional control law and the value the integral term of Section 3.5 has to carry at rest, as shown in Equation (3.5).

$$\bar{u} = \frac{100\,c_2\sqrt{h_{2,\mathrm{SP}}}}{c_1} = 50\sqrt{h_{2,\mathrm{SP}}} \tag{3.5}$$

where the second form uses the default coefficients and returns a pump position in percent. Holding 0.50 m requires 35.36 % of pump travel whether the splitter is shut, half open or wide open, which states the splitter invariance of the operating point directly.

3.3.3 Overflow, Dry Tanks and the Instrument

Each tank stops at the rim. When a level has reached $h_{\max}$ and its balance is still positive, the balance is replaced by zero for that step and an overflow flag is raised for that tank, as shown in Equation (3.6).

$$\frac{\mathrm{d}h_i}{\mathrm{d}t} \leftarrow 0 \;\text{ if }\; h_i \geq h_{\max} \text{ and } \frac{\mathrm{d}h_i}{\mathrm{d}t} > 0, \quad i \in \{1,2\} \tag{3.6}$$

The spilled liquid leaves the model. It is not routed into the tank below and its rate is not reported. Both levels are additionally clipped to $[0, h_{\max}]$ after every integration substep, which is what enforces the dry limit, and at $h_i = 0$ the corresponding outflow term vanishes on its own because $\sqrt{0} = 0$. The draw-off does not vanish at $h_2 = 0$, so a dry bottom tank under draw-off is held at zero by the clip rather than by the physics.

Both transmitters add independent Gaussian noise to the true level and clip the result to the tank, as shown in Equation (3.7).

$$\hat{h}_i = \min(h_{\max},\ \max(0,\ h_i + \sigma_{n,i}\,\eta_i)), \qquad \sigma_{n,2} \equiv \sigma_n \tag{3.7}$$

where σ_n is the noise standard deviation of LT2, the transmitter the loop reads, $\sigma_{n,1}$ is that of LT1, each in m and set separately in the application, and η_i is an independent unit normal deviate obtained through a Box-Muller pair from the seeded generator of the simulation engine, redrawn for each of the two instruments on every controller execution. Nothing else is added. Neither instrument carries a lag, an offset or a drift, and its noise is white at the 2 Hz rate. The noise enters the readings only. The states h_1 and h_2, and with them the drains of Equations (3.1) and (3.2), are computed from the true levels. Only $\hat{h}_2$ is used, and the controller acts on it. The deviate η_1 is drawn first whatever $\sigma_{n,1}$ is, so $\hat{h}_2$, and every closed-loop result in this chapter, is independent of $\sigma_{n,1}$. Both readings are displayed and recorded beside the true levels, the top-tank trend above the bottom-tank trend as the tanks stand in the plant, so the error the loop reacts to and the indication LT1 gives can each be read against the state underneath it.

Table 3.2 Model, Controller and Numerical Parameters. Values in the First Two Blocks Are Adjustable in the Application.

Symbol	Meaning	Default	Range	Unit
Adjustable model parameters				
c_1	pump coefficient	0.08	0.02–0.20	m/s
c_2	drain coefficient	0.04	0.01–0.10	$m^{1/2}/s$
h_{max}	overflow height	1.0	0.5–2.0	m
σ_n	LT2 transmitter noise	0.004	0–0.020	m
$\sigma_{n,1}$	LT1 transmitter noise	0.004	0–0.020	m
Adjustable controller parameters				
K_c	proportional gain	300	0–1000	%/m
τ_I	integral time	60	0.05–500[a]	s
τ_D	derivative time	0	0–50	s
$h_{2,SP}$	level setpoint	0.50	0–h_{max}	m
T_p	MPC prediction horizon	300	40–600	s
λ	MPC move suppression	0.3	0–3	—
—	MPC estimator	Kalman	or MHE	—
—	speed multiplier	1	0.5–10	—
Fixed in the source				
u_{bias}	controller bias	$u(0)$	—	%
T_t	tracking time	τ_I	—	s
Δt	integration substep	0.05	—	s
Δt_c	controller sample period	0.5	—	s
T_s	MPC decision period	2	—	s
N_e	MHE window	40	—	executions
ε_r	relay hysteresis band	0.012	—	m
—	substeps per execution	10	—	—
—	historian rate	2	—	samples/s
—	noise seed	20260824	—	—
$h_i(0)$	initial levels	steady	—	m
A	tank area	—	—	divided out

[a] The top of the range, 500 s, switches integral action off.

3.3.4 Parameters and Numerics

Table 3.2 collects every constant the simulator needs, and the proportional gain moves in steps of 3 %/m. Every parameter edit is guarded before it reaches the model, c_1 and c_2 taking a lower bound of machine epsilon, h_{max} a lower bound of 0.1 m and both noise levels a lower bound of zero. A change of tank height re-clips both levels into the new geometry and moves the top of the setpoint range with it, pulling a setpoint above the new rim down to the rim. The three tuning parameters are bounded to their table ranges, and the top of the integral-time range switches integral action off.

Every case starts at steady operation rather than from empty tanks. In manual the two levels are set to the steady state of Equations (3.1) and (3.2) at the pump in force, and in automatic the pump that holds the setpoint is found by bisection on that steady state and the levels are set there, in both cases with the splitter, the draw-off, the leak and the parameters in force. A brim-full top tank passes on at most $c_2\sqrt{h_{max}}$, and a bottom tank whose inflow cannot meet the draw-off starts empty. A setting chosen before a run starts therefore changes the starting state instead of producing a step, and it gives the same run as the same setting carried through a restart. A step test is made after the run has started.

Equations (3.1) and (3.2) are integrated with explicit Euler on a fixed 0.05 s substep, with the overflow test of Equation (3.6) applied inside every substep and the $[0, h_{max}]$ clip applied after it. Nothing is gained from a higher-order integrator on this right-hand side, because the four clipping bounds contribute more error than the smooth dynamics do. This explicit Euler step replaces the framework's Runge–Kutta integrator. Every tenth substep, which is 0.5 s of simulated time, the

controller executes and the historian writes one row, on the simulated clocks of Section 1.6.1, so the trajectory at 10× is the trajectory at 1× and a noisy run repeats exactly. The export advances its time column in exact 0.500 s increments. A change to the pump, the splitter, the draw-off or the leak reaches the plant at the next controller execution, so every such step lands on the 0.5 s grid. One detail of the export matters for identification. A row stamped t carries the true levels at t and the measured level read at t, which is the reading the controller execution at t acts on, while its pump is the value held over $[t - \Delta t_c, t]$, the output of the execution at $t - \Delta t_c$. A fit of either level against the pump column therefore has to move the pump column back by Δt_c. Without that shift the shut-splitter test of Section 3.4 returns an apparent dead time of 18.0 s instead of 18.5 s.

3.3.5 Assumptions

The model keeps the mechanisms that change how the loop has to be tuned and discards the rest. Each entry below names what the simplification buys and the situation that breaks it.

1. *Non-interacting discharge.* The flow from TK-1 to TK-2 is $c_2\sqrt{h_1}$ and is independent of h_2, which requires the TK-1 outlet to discharge above the TK-2 liquid surface at all times. It fails as soon as TK-2 is flooded to the level of that nozzle, at which point the driving head becomes $h_1 - h_2$ and the pair becomes genuinely interacting [11]. Nothing in the source detects that condition, so the non-interacting form is used even when TK-2 is at the rim.

2. *Equal and constant tank areas.* The two balances share one c_1 and one c_2 and carry no area at all, which forces both tanks to have the same cross-section and the same drain geometry. Two vessels of different size would need separate coefficients, and a conical or horizontal cylindrical vessel would add a geometric nonlinearity on top of the hydraulic one.

3. *Quasi-steady outflow.* Each drain passes the steady Torricelli flow for whatever level stands above it at that instant, which gives the liquid in the outlet lines no momentum of its own. The approximation survives as long as a tank takes far longer to empty than its outlet column takes to accelerate, and the default constants satisfy that by a wide margin. A long discharge line, a fast valve stroke or any water-hammer question breaks it.

4. *Ideal splitter.* The split is a pure fraction applied to the pump flow, with no pressure drop, no installed characteristic, no stroking time, no hysteresis and no interaction between the two branches. The two paths of a real three-way valve have different resistances, so the split of a real valve depends on the levels it is feeding. The splitter also moves instantly, which makes every splitter experiment a true step.

5. *Ideal pump.* Flow is strictly proportional to the pump signal with no head dependence, no minimum flow, no slew limit and no deadband. A real centrifugal pump delivers less as the discharge head rises, which would couple the pump flow to h_1 and to h_2 and add a second interaction path that this model does not have.

6. *Overflow discards liquid.* Above $h_{\max}$ the excess leaves the model rather than cascading into the tank below or into a bund. A spill therefore breaks the mass balance, and its rate appears nowhere. What the flag does establish is that an inventory has escaped the loop, and that is all it establishes. Containment volume and spill duration lie outside what the model carries.

7. *Ideal transmitters apart from white noise.* Neither LT1 nor LT2 has lag, bias, drift, a resolution limit or a rangeability limit, and their noise is white, Gaussian and independent. The dominant errors in a real level measurement are slow ones, from a moving surface and from density, and they are correlated rather than white.

8. *Omitted entirely.* Inlet flow disturbances upstream of the pump, downstream pressure on either drain, vapor space pressure, temperature, composition, agitation, a cascade to a flow controller and any third vessel are absent. Each omission was made for the same reason, because none of these mechanisms changes the structural behavior the splitter produces.

3.4 Effect of the Splitter Fraction on Plant Order

3.4.1 Linearized State-Space Model

Taking first-order terms of Equations (3.1) and (3.2) around a resting point $(\bar{h}_1, \bar{h}_2, \bar{u})$, as Equation (1.4) does for one tank, produces the two-state deviation model of Equation (3.8), the state-space form of Equation (1.7).

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} h_1' \\ h_2' \end{bmatrix} = \begin{bmatrix} -1/\tau_1 & 0 \\ 1/\tau_1 & -1/\tau_2 \end{bmatrix}\begin{bmatrix} h_1' \\ h_2' \end{bmatrix} + \begin{bmatrix} c_1(1-x_v)/100 \\ c_1 x_v/100 \end{bmatrix} u' \tag{3.8}$$

in which a prime marks a departure from the resting point and the two time constants follow Equation (3.9).

$$\tau_1 = \frac{2\sqrt{\bar{h}_1}}{(1+x_L)\,c_2}, \qquad \tau_2 = \frac{2\sqrt{\bar{h}_2}}{c_2} \tag{3.9}$$

The zero in the upper right corner of the state matrix in Equation (3.8) is the whole content of the non-interacting claim, and it is a zero at every operating point rather than an approximation. The two poles of the plant are therefore $-1/\tau_1$ and $-1/\tau_2$ and are read straight off the diagonal. Each time constant is that of Equation (1.5) for one tank. Substituting Equation (3.4) into Equation (3.9) at $x_L = 0$ gives the central relation of the analysis, as shown in Equation (3.10).

$$\tau_1 = (1-x_v)\,\tau_2 \tag{3.10}$$

The upper time constant is the lower one scaled by the share of flow that still goes through the top tank. At the 0.5013 m operating point $\tau_2 = 35.40$ s, so shutting the splitter gives a repeated pole at –0.0282/s and opening it to 50 % moves one of the two poles to –0.0565/s while leaving the other exactly where it was. Opening it fully sends the upper pole to infinity, because an empty tank contributes no lag.

3.4.2 Pump and Splitter Transfer Functions

Eliminating h_1' from Equation (3.8) gives the transfer function from the pump to the controlled level, as shown in Equation (3.11).

$$G_p(s) = \frac{h_2'(s)}{u'(s)} = \frac{K_p\,(\tau_z s + 1)}{(\tau_1 s + 1)(\tau_2 s + 1)}, \qquad \tau_z = x_v\,\tau_1 \tag{3.11}$$

where $K_p = c_1\tau_2/100$ is the process gain in m/% of pump travel. With $x_v = 0$ it is the two non-interacting lags of Equation (1.10). Three properties of Equation (3.11) organize the results. The gain K_p contains no x_v at all, so the steady-state map is untouched by the splitter, as Equation (3.3) already showed. The numerator zero at $-1/\tau_z$ and the pole at $-1/\tau_1$ both travel outward together as the splitter opens, and at $x_v = 1$ they coincide and cancel exactly, leaving the first-order plant $K_p/(\tau_2 s + 1)$. Between the ends the zero is faster than the pole it partially cancels by the factor $1/x_v$, and combining Equations (3.10) and (3.11) gives $\tau_z = x_v(1-x_v)\tau_2$, which is largest at a half open splitter and reaches 8.85 s there. At the default operating point $K_p = 0.0283$ m/%, and that number is the same at $x_v = 0$, at $x_v = 0.5$ and at $x_v = 1$.

The splitter itself is a disturbance, and the transfer function it enters through is the reason the operating point does not move, as shown in Equation (3.12).

$$G_v(s) = \frac{h_2'(s)}{x_v'(s)} = \frac{c_1\bar{u}}{100}\,\frac{s}{(s+1/\tau_1)(s+1/\tau_2)} \tag{3.12}$$

Table 3.3 Manual Step Tests From 35.4 % to 45.4 % Pump at Three Splitter Positions, $\sigma_n = 0$, $q_d = 0$, $x_L = 0$. The Predicted Columns Come From Equations (3.4) to (3.11) at the Starting Point.

Splitter x_v	0 %	50 %	100 %
$\bar{h}_1$ before (m)	0.5013	0.1253	0.0000
$\bar{h}_1$ predicted (m)	0.5013	0.1253	0.0000
$\bar{h}_2$ before (m)	0.5013	0.5013	0.5013
$\bar{h}_2$ after (m)	0.8245	0.8245	0.8245
$\hat{K}_p$ (m/%)	0.0323	0.0323	0.0323
t_{28} (s)	42.4	21.1	13.7
t_{63} (s)	90.1	55.3	42.3
$\hat{\tau}_p$ (s)	71.6	51.3	42.9
$\hat{\theta}_p$ (s)	18.5	4.0	−0.6
τ_1 predicted (s)	35.40	17.70	0
τ_2 predicted (s)	35.40	35.40	35.40
τ_z predicted (s)	0	8.85	0
IMC K_c (%/m)	66.6	174.0	309.4
IMC τ_I (s)	71.6	51.3	42.9

The numerator of Equation (3.12) is a bare s, so the disturbance transfer function has a zero at the origin and its steady gain is exactly zero. A splitter step at a fixed pump position produces a transient in the bottom level and returns it to the level it started from with no offset and no control action of any kind. Starting from a shut splitter, where $\tau_1 = \tau_2 = \tau$, the response to a step of size Δx_v is $(c_1\bar{u}/100)\,\Delta x_v\, t\, e^{-t/\tau}$, which peaks at $t = \tau$ with a height of $(c_1\bar{u}/100)\,\Delta x_v\,\tau/e$. For the half open step at the default operating point that estimate is 0.184 m at 35.4 s, and the running application produces a peak of 0.176 m at 31.8 s before returning to 0.5013 m, which is the level it left. Two further disturbances act differently and are treated in Section 3.7, because the draw-off and the leak both move the steady state that the splitter leaves alone.

Pause. Try Exercise 3.1 before reading on.

3.4.3 Step Tests at Three Splitter Positions

Table 3.3 reports three step tests carried out in manual mode with the instrument noise set to zero, each raising the pump by 10 % from 0.5013 m at rest. The splitter position and a pump of 35.4 % are chosen before each run starts, so both tanks start at rest at that splitter position, and the pump is raised 10 s into the run. Times are counted from the controller execution at which the pump moves. The graphical fit takes the gain as the secant slope over the step, the time constant as $\hat{\tau}_p = 1.5\,(t_{63} - t_{28})$ and the apparent dead time as $\hat{\theta}_p = t_{63} - \hat{\tau}_p$, the two-point fit of Equation (1.18).

The three tests return the same gain to three decimal places and three quite different dynamic descriptions. The apparent dead time falls from 18.5 s with the splitter shut to 4.0 s half open and to −0.6 s wide open. The same construction applied to the exact first-order response gives −0.03 s, so the small negative value comes from the gain rising along the step rather than from any lead in the plant. The fitted time constant falls from 71.6 s to 42.9 s. The steady levels reached in the top tank, 0.5013, 0.1253 and 0.0000 m, match the $(1 - x_v)^2$ scaling of Equation (3.4) exactly. The vessels are unchanged across the three tests, and only the flow path differs.

Pause. Try Exercise 3.2 before reading on.

Two differences between the fitted and the predicted columns are worth naming, because both are informative rather than errors. The fitted gain 0.0323 m/% exceeds the linearized 0.0283 m/% by 14 %, because the step raises the bottom level from 0.5013 to 0.8245 m and the local gain $2c_1^2\bar{u}/(100\,c_2)^2$ grows with level, so the secant over a large step lies above the tangent at its start. The fitted time constant at the shut splitter, 71.6 s, exceeds the 58.1 s that the same graphical construction returns

when applied to the exact linear second-order response, for the same reason. Chapter 2 follows the same growth of gain and time constant with level in a single tank. The apparent dead time, on the other hand, is a structural quantity rather than an operating-point one, and the 18.5 s measured against the 17.8 s predicted from Equation (3.11) agrees to 4 %.

Pause. Try Exercise 3.3 before reading on.

Three distinct models have now appeared and none substitutes for another. Equations (3.1) and (3.2) are the simulation model, exact inside the assumptions of Section 3.3.5. Each column of Table 3.3 is an identified control model, good near one level at one splitter setting only. The PID of Section 3.5 holds no model, only a gain, two time constants, a tracking time and a sample period. The predictive controller of Section 3.5.4 carries one identified control model, fitted near 0.50 m with the splitter shut, and keeps it at every splitter setting unless it is given the table of Section 3.5.5, five such models fitted at fixed splitter settings.

> **TRY IT IN THE APP 2 · Open-Loop Actuator Step**
>
> Load the case and press Run. At 10 s the pump steps from 35.36 to 45.36 % with the controller in manual. Watch the two level trends. The upper tank rises at once, while the lower tank, fed only through it, starts slowly and has covered 63 percent of its rise only about 90 s after the step (Table 3.3).

3.5 Controller Design

3.5.1 Velocity Form

LIC-102 runs the velocity-form PID of Section 1.6.2 [15], written here in the loop's own symbols. Error is taken against the measured bottom level in Equation (3.13), and the three increments follow in Equations (3.14) to (3.16).

$$e_k = h_{2,\mathrm{SP}} - \hat{h}_{2,k} \tag{3.13}$$

$$\Delta u_k^P = P_k - P_{k-1}, \quad P_k = K_c\, e_k \tag{3.14}$$

$$\Delta u_k^I = \frac{K_c}{\tau_I}\, e_k\, \Delta t_c \tag{3.15}$$

$$\Delta u_k^D = D_k - D_{k-1}, \quad D_k = -K_c \tau_D\, (\hat{h}_{2,k} - \hat{h}_{2,k-1})/\Delta t_c \tag{3.16}$$

with K_c in %/m, τ_I and τ_D in s and $\Delta t_c = 0.5$ s. Both P_k and D_k are formed with the gains in force at scan k. Equation (3.17) combines the three increments and clips the sum to the actuator span.

$$\begin{aligned} s_k &= s_{k-1} + \Delta u_k^P + \Delta u_k^D \\ c_k &= u_{\mathrm{bias}} + s_k + I_{k-1} \\ u_k &= \min(100,\ \max(0,\ c_k)) \\ I_k &= I_{k-1} + \Delta u_k^I + \frac{\Delta t_c}{T_t}\,(u_k - c_k) \end{aligned} \tag{3.17}$$

In the symbols of Equations (1.24) to (1.27), $s_k = S_k$, $c_k = v_k$, $u_{\mathrm{bias}} = u_b$ and the action is $a = +1$. Here s_k accumulates the proportional and derivative increments without limit and telescopes to $P_k + D_k$ exactly, c_k is the output the algorithm demands, I_k is the integral state and $T_t = \tau_I$ is the tracking time of the back-calculation. The integral used at scan k was accumulated through scan $k-1$. The bias u_{bias} is the pump in force when the run starts, which is the pump that holds the starting steady state, and the integral starts from zero, so u_{bias} carries the starting operating point and I_k every change from it. Away from the limits $u_{k-1} = c_{k-1}$, and differencing the second line gives $c_k = u_{k-1} + \Delta u_k^P + \Delta u_k^D + \Delta u_{k-1}^I$, the velocity law with the integral increment of the previous scan. The recursion is therefore algebraically identical to the positional law $u = P + I + D$ only while the pump stays inside its range. The back-calculation runs whenever $K_c/\tau_I \neq 0$. Setting τ_I to the top of

its range, 500 s, switches integral action off, so neither Δu_k^I nor the back-calculation is applied, I_k keeps its last value and the law becomes proportional, or proportional-derivative, control about the fixed bias $u_{\text{bias}} + I$.

As in every framework app, the derivative acts on the measurement and a gain edit acts at once on the whole proportional term. A setpoint move therefore steps the pump by $K_c \Delta h_{2,\text{SP}}$ with no derivative kick, and a gain edit steps it by $\Delta K_c\, e_k$ and leaves the integral untouched. The controller is reverse acting, because raising the pump raises the level, so the pump has to fall as the measured level rises, and no action switch is offered. The gains are entered as K_c, τ_I and τ_D, so the integral time of a tuning correlation is typed in directly and the integral gain K_c/τ_I follows from it.

The gain units require attention, because K_c is expressed in percent of pump travel per meter of level and is therefore a number in the hundreds. A tuning quoted as $K_c = 2.0$ and $K_I = 0.03$ for a pump signal on a zero to one scale is $K_c = 200\,\%/\text{m}$ and $\tau_I = 66.7\,\text{s}$ in the units of Table 3.2. The two are the same controller, although the proportional gain moves in steps of 3 %/m, so 201 %/m is the nearest setting.

3.5.2 Bias, Transfer and Setpoint Edits

The velocity form carries its operating point in the bias and the integral state, so where the recursion starts decides them. In manual the integral is overwritten so that the demand equals the manual pump, as shown in Equation (3.18).

$$I_k = u_{\text{man},k} - u_{\text{bias}} - s_k \qquad \text{(manual)} \tag{3.18}$$

Both transfers between manual and automatic are therefore bumpless, as Section 1.6.2 explains. A run that starts in automatic starts from the steady state of its setpoint with u_{bias} at the pump that holds it and the integral at zero, so nothing moves before the first scripted event except the response to transmitter noise.

Editing the setpoint, the integral time or the proportional gain during a run changes the next increments and nothing else, so integral action built up against a standing load survives the edit. A loop with $K_c = 66\,\%/\text{m}$ and $\tau_I = 71.65\,\text{s}$ holding 0.500 m against a 0.010 m/s draw-off sits with the pump at 47.86 %, the 35.36 % bias of its starting steady state plus 12.50 % carried by the integral. Raising the setpoint by 0.01 m moves the pump to 48.52 % in one sample, which is the proportional step $K_c \Delta h_{2,\text{SP}} = 0.66\,\%$, and the level climbs to the new setpoint without first falling below 0.500 m.

3.5.3 Saturation and Windup

Anti-windup is the back-calculation term of Equation (3.17) [15]. While the pump is clipped, $(\Delta t_c/T_t)(u_k - c_k)$ drives I toward the value that reproduces the applied pump with a tracking time of τ_I, so the integral alone is corrected and the running sum s_k is never truncated. Setting the integral update to zero on a limit gives $c_k = u_k + K_c e_k$, so with $\tau_D = 0$ the sum $u_{\text{bias}} + I$ settles at the limit value itself and the demand returns inside the range as soon as the error changes sign. Held on the 100 % limit for 700 s by the 0.020 m/s draw-off of Section 3.7, the bias and the integral together reach 99.98 % while the demand stands at 116.5 %. Removing the draw-off releases the pump 16.5 s later, at the first execution after the level crosses the setpoint. That sum then still has to unwind from 100 % to the 35.4 % the unloaded plant needs, the pump stays above 32.5 %, and TK-1 stays at its rim for 104 s, so the bottom level climbs to 0.940 m before it returns. Back-calculation bounds the integral by the actuator range, and a load that holds the pump on a limit for longer than τ_I still leaves a full limit's worth of integral to unwind.

3.5.4 Model Predictive Control

A model predictive controller (MPC), the option of Section 1.6.3, can take the pump in automatic mode instead of the PID. Every case except cases 9 and 10 starts with the PID, and while the MPC acts the PID follows the applied pump so that either can hand over to the other. Every $T_s = 2\,\text{s}$, four controller executions, the MPC computes a pump trajectory over a horizon, sends its first move

and keeps it until the next computation, the receding-horizon arrangement of industrial predictive control [16] whose constrained form is analyzed in [17]. It plans on two lags in series about the design point $\bar{u}_m = 35.36\,\%$, $\bar{h}_m = 0.50\,\mathrm{m}$, as shown in Equation (3.19).

$$\begin{aligned} \tau_a \frac{\mathrm{d}z_1}{\mathrm{d}t} &= -z_1 + K_m\,(u - \bar{u}_m + d) \\ \tau_b \frac{\mathrm{d}z_2}{\mathrm{d}t} &= z_1 - z_2, \qquad h_m = \bar{h}_m + z_2 \end{aligned} \tag{3.19}$$

Here z_1 and z_2 are model states in m, h_m is the predicted bottom level, d is an unmeasured disturbance in percent of pump travel, $K_m = 0.028\,28\,\mathrm{m}/\%$, $\tau_a = 39.7\,\mathrm{s}$ and $\tau_b = 31.1\,\mathrm{s}$. The constants are a least-squares fit to the mean response to pump steps of ± 3 percentage points from rest at the design point, with the splitter shut, no load and the noise off, over 600 s. The residual is $7.8 \times 10^{-6}\,\mathrm{m}$ root mean square, under 0.01 % of the response, the fitted dead time of 0.14 s is dropped, and a first-order-plus-dead-time fit leaves 143 times that residual. At $x_v = 0$ Equation (3.11) gives the same gain and two lags of 35.36 s, whose 70.71 s sum the fit reproduces to 0.02 s. The model is the shut-splitter plant only, with no splitter, no zero, no top-tank level, no rim and no square-root drain.

Offset-free tracking comes from a steady-state Kalman filter on Equation (3.19), sampled with a zero-order hold at Δt_c and augmented with d as a random walk, the disturbance model of offset-free predictive control [18]. At every execution, whichever controller holds the pump, it updates as in Equation (3.20).

$$\begin{aligned} \bar{z}_k &= \Phi\,\hat{z}_{k-1} + \Gamma\left(u_{k-1} - \bar{u}_m + \hat{d}_{k-1}\right) \\ \varepsilon_k &= \hat{h}_{2,k} - \bar{h}_m - \bar{z}_{2,k} \\ \hat{z}_k &= \bar{z}_k + \ell_z\,\varepsilon_k, \qquad \hat{d}_k = \hat{d}_{k-1} + \ell_d\,\varepsilon_k \end{aligned} \tag{3.20}$$

where Φ and Γ are the sampled matrices of the two lags, u_{k-1} is the pump applied over the period just ended, $\hat{h}_{2,k}$ is the reading the PID also receives, and $\ell_z = [1.615,\ 0.202]^{\mathsf{T}}$ and $\ell_d = 315.9\,\%/\mathrm{m}$ are the converged filter gains obtained with a measurement variance $r = \sigma_n^2$, a state noise variance $10^{-6} r$ and a disturbance noise variance $q_d = 100\,r/K_m^2 = 2.00\,\%^2$. A load the model lacks, including an error in its own dynamics, appears in $\hat{d}$ as an equivalent pump offset that the plan holds constant, so a constant load leaves no offset. A 1 mm prediction error moves $\hat{d}$ by 0.32 %.

Each decision minimizes a quadratic cost over predictions of Equation (3.19) discretized at T_s, across $T_p = 300\,\mathrm{s}$, or $N = 150$ steps, about four times the sum of the lags. Twelve pump positions v_j, held over blocks of $1, 1, 1, 1, 2, 2, 3, 4, 6, 9, 15$ and 105 steps, minimize Equation (3.21) subject to the hard bounds $0 \le v_j \le 100$.

$$J = \frac{T_s}{\tau_n} \sum_{k=1}^{N} \left(\frac{h_{m,k} - h_{2,\mathrm{SP}}}{y_s} \right)^2 + \lambda \sum_{j=1}^{12} \left(\frac{v_j - v_{j-1}}{u_s} \right)^2 \tag{3.21}$$

In it $h_{m,k}$ is the level predicted k steps ahead with $\hat{d}$ held, v_0 is the pump applied now, the factor T_s/τ_n with $\tau_n = 70.7\,\mathrm{s}$ makes the first sum a time integral, $y_s = 0.03\,\mathrm{m}$ is the scorecard settling band, $u_s = 100\,\%$ and $\lambda = 0.3$. There is no level band, rim, rate limit, terminal weight or preview, because the controller is told neither the draw-off nor the leak, and the splitter fraction only through the schedule of Section 3.5.5. The decisions are departures from $\bar{u}_m$ with a ridge of 10^{-9} on the scale of the weights, solved by the dual active-set algorithm of [19]. The values of λ and q_d follow one rule for the seven single-loop studio apps, the largest λ within 10 % of the best true-level IAE on cases 3 and 4 among settings that move the pump no more than the shipped PID, and for this plant the rule takes the largest q_d of its grid.

The answer of the solver is discarded when it is not optimal, when its first move is not a finite number, or when that move lies outside 0 to 100 %. The PID, which tracks the applied pump as it tracks the operator in manual, then takes over without a bump until a valid plan returns. A rejection

Table 3.4 The Scheduled Model's Entries, Fitted at 0.50 m and $K_m = 0.028\,28\,\mathrm{m}/\%$, Against Two Lags Without the Parallel Path and the Hand Linearization

x_v (%)	τ_a (s)	τ_b (s)	σ	RMS (µm)	RMS_2 (µm)	$(1-x_v)\tau_2$ (s)
0	39.67	31.05	0	7.8	—	35.36
25	25.01	36.77	0.257	5.4	443	26.52
50	16.91	35.90	0.504	5.4	411	17.68
75	8.19	35.55	0.748	9.2	162	8.84
100	8.19[a]	35.38	1	17.5	—	0

RMS_2 is the residual of two lags in series without the parallel path. The 0 % entry is Equation (3.19). [a]Borrowed from the 75 % entry.

forced at 36 s of case 9 without noise changes the pump by 1.9 %, the size of the next PID moves, and the fresh plan that takes it back moves it by 49 %. A plan is already computed while a case loads, so an MPC run is logged from $t = 0$. The controller reads only the LT2 reading, the applied pump, the setpoint, the pump limits and, when scheduled, the splitter setting, which keeps the plant of Equations (3.1) and (3.2), the instrument of Equation (3.7) and the model of Equation (3.19) apart.

3.5.5 Model Scheduled on the Splitter Setting

The splitter is a hand valve that the operator sets and no loop moves, so the controller may be told its position, as a lineup entry would tell it, without reading a process state. With the scheduled model chosen, the plan and the filter use the entry of a five-model table nearest the told setting, a multiple-model predictive controller [20]. Each entry passes a share σ of the input around the first lag of Equation (3.19), as shown in Equation (3.22).

$$\begin{aligned} \tau_a \frac{\mathrm{d}z_1}{\mathrm{d}t} &= -z_1 + K_m(1-\sigma)\,v, \qquad v = u - \bar{u}_m + d \\ \tau_b \frac{\mathrm{d}z_2}{\mathrm{d}t} &= z_1 - z_2 + K_m \sigma\, v \end{aligned} \tag{3.22}$$

Its transfer function is Equation (3.11) with $\tau_z = \sigma\tau_a$, so the balances linearized at 0.50 m give $\sigma = x_v$, $\tau_a = (1-x_v)\tau_2$ and $\tau_b = \tau_2 = 35.36\,\mathrm{s}$, with K_m unchanged because the splitter moves the path and not the steady level. Each entry was fitted as Equation (3.19) was, with its splitter set before the run, as listed in Table 3.4. The fits return σ within 0.007 of x_v, $K_m = 0.028\,28\,\mathrm{m}/\%$ throughout and lag sums within 1.1 % of the hand values. At 100 % a pump step cannot see τ_a, so the 75 % entry's value stands in, and replacing it by 0.5 or 17.7 s moves the IAE of five runs that enter that entry by at most 0.09 m s.

Weights, horizon, blocks and filter variances are those of Section 3.5.4, each entry with its own filter gain. The update over the period just ended uses the entry then in force, the new entry takes over before the plan, and $\hat{z}$, $\hat{d}$ and the input record carry over, so neither the estimated level nor the pump moves at a change. The carried $\hat{z}_1$ changes meaning, because each entry counts its upper path from the outflow of its own split, a shift of about $K_m\bar{u}_m\Delta\sigma$ (0.50 m at the 0 to 50 % opening) that the filter books as load, $\hat{d}$ reaching 26.0 % 5 s after the opening of case 10 without noise and 2.3 % after 35 s. Re-mapping $\hat{z}_1$ by script cuts the IAE of case 10 from 1.53 to 1.35 m s but takes TK-1 to its rim when the splitter goes to 50, 100 and 0 %, so it was not adopted. Scheduled plans must be optimal, with finite moves, predictions and cost and every move inside 0 to 100 %, a rejection is followed by a three-decision return, and an MHE restarts its window on the new entry from the filter's prediction.

3.5.6 Moving Horizon Estimation

The state from which each plan starts may instead come from a moving horizon estimator (MHE), while the filter of Equation (3.20) keeps running and logging under either choice. The MHE works on the augmented state $\xi = [z_1\ z_2\ d]^{\mathsf{T}}$ of that filter and, at execution k, fits the readings $j = s, \ldots, k$ with $s = k - N_e$ through the bounded least squares of Equation (3.23), the constrained estimator analyzed

in [21].

$$\begin{aligned} \min \; & \|\xi_s - \bar{\xi}_s\|^2_{\bar{P}^{-1}} + \sum_{j=s}^{k-1} \omega_j^{\mathsf{T}} Q^{-1} \omega_j \\ & + \sum_{j=s}^{k} \frac{\left(\hat{h}_{2,j} - \bar{h}_m - z_{2,j}\right)^2}{r} \\ \text{s.t. } & \xi_{j+1} = \Phi_e \, \xi_j + \Gamma_e \, (u_j - \bar{u}_m) + \omega_j \\ & 0 \le \bar{h}_m + z_{i,j} \le h_{\max,0}, \qquad i = 1,2 \end{aligned} \tag{3.23}$$

The decisions are ξ_s and the noise ω_j on all three states, $\Phi_e = \left[\begin{smallmatrix} \Phi & \Gamma \\ 0 & 1 \end{smallmatrix}\right]$ and $\Gamma_e = [\Gamma^{\mathsf{T}} \; 0]^{\mathsf{T}}$ repeat the prediction of Equation (3.20), u_j is the pump applied after reading j, $Q = \mathrm{diag}(10^{-6}r, 10^{-6}r, q_d)$ and r are the variances that designed the filter, $\bar{\xi}_s$ is the estimate the MHE made at execution $s-1$ advanced one step, and the steady prior covariance $\bar{P}$ of the filter is held fixed as the arrival cost. With $N_e = 40$ the window spans 20 s and 41 readings, which gives the active-set solver of Section 3.5.4 123 variables and 164 bound rows. Both model levels stay between the floor and $h_{\max,0} = 1\,\mathrm{m}$, the declared rim rather than the one set for a run, and d is free because no declared value limits it. Bounding z_1 treats the slower fitted lag as the top tank, whereas the plant has two equal lags of 35.36 s. With no bound active the problem is the filter's own, and a first window starts from the filter's prediction. A window left without an optimal, finite answer inside the bounds gives way to a Kalman measurement update of the MHE's one-step prediction, clipped into the bounds, and the history marks the substitution.

3.5.7 Relay Autotuning

PI constants for the PID can also come from a limit cycle that the loop produces under the relay feedback test of [22], an action that no shipped case starts with. Taken from the PID in automatic, a relay sets the pump to $u_0 \pm \delta_r$, with u_0 the mean pump of the last 30 s and δ_r equal to 5 % of the span, or to 0.9 of the gap between u_0 and the nearer limit when that is smaller. It switches side whenever the LT2 reading passes the setpoint by more than $\varepsilon_r = 3\sigma_{n,0} = 0.012\,\mathrm{m}$, three times the declared noise $\sigma_{n,0}$, and a cycle with a half under four executions is discarded. Once the period and the swing $\pm\alpha$ of two successive accepted cycles agree to within 5 %, the first harmonic of the relay output, corrected for the hysteresis, gives the ultimate gain K_{cu}, and the Tyreus–Luyben PI constants follow from K_{cu} and the period P_u as in Equation (3.24).

$$K_{cu} = \frac{4\delta_r \sin(\pi D)}{\pi \sqrt{\alpha^2 - \varepsilon_r^2}}, \qquad K_c = \frac{K_{cu}}{3.2}, \qquad \tau_I = 2.2\,P_u \tag{3.24}$$

Here D is the fraction of each cycle that the relay spends on its high side, near one half only when u_0 is the pump that holds the setpoint. The constants are written to three figures with $\tau_D = 0$ and stay editable. A test stops unconverged after 30 cycles or 600 s or on a second request, and the PID, which follows the applied pump, resumes without a bump. The angle $\psi = \arcsin(\varepsilon_r/\alpha)$ separates the cycle from -180° because of the hysteresis, and from $\psi = 10^\circ$ upward the result is flagged as a cycle placed by the band and not by the tanks.

3.6 Results

The app ships eleven cases, and every one starts at steady operation with the bottom level at 0.50 m, so the first scripted event of each, at 10 s, acts on a settled plant. Case 1 holds the reference operating point in manual with the pump at 35.36 %, and case 2 steps the pump from 35.36 to 45.36 %. Case 3 steps the setpoint to 0.60 m under the shipped tuning, case 4 applies a draw-off of 0.006 m/s in automatic and case 5 the same draw-off in manual. Cases 6 and 7 step the setpoint to 0.58 m at $K_c = 120$ and 600 %/m. Case 8 steps it to 0.90 m on a smaller pump ($c_1 = 0.035\,\mathrm{m/s}$, 80.8 % at

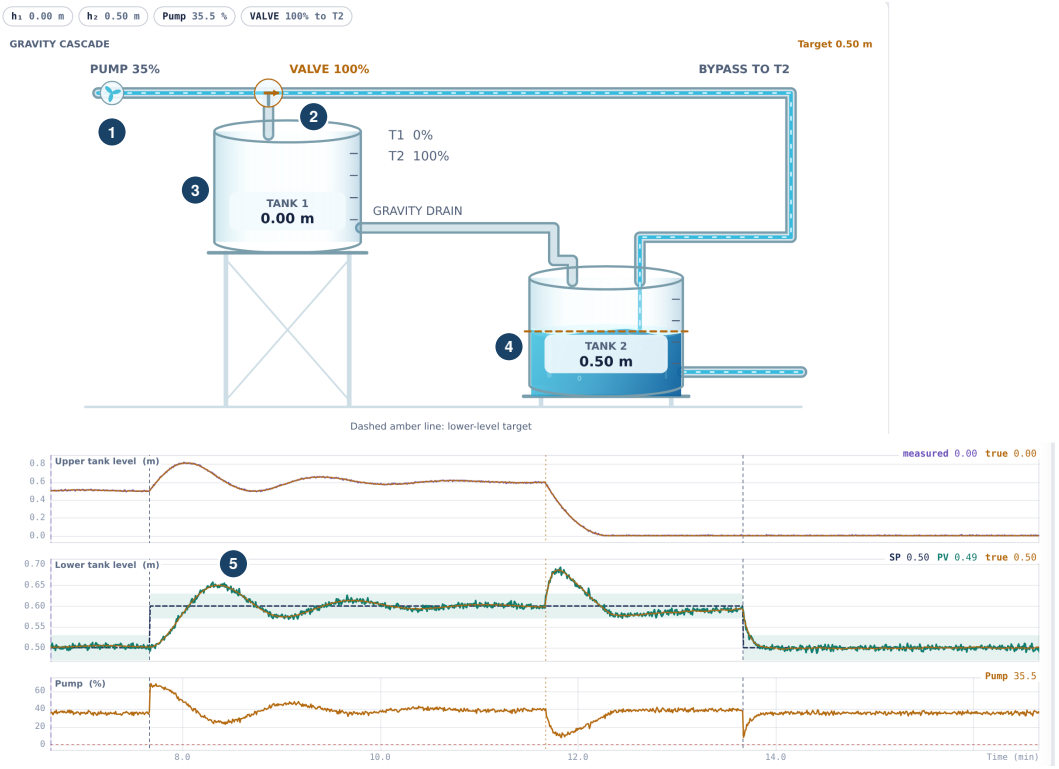


Figure 3.2 One controller meeting both plant structures in a single run built from case 1, shown over the last 10 min of 1000 s with the shipped $K_c = 300\,\%/\mathrm{m}$, $\tau_I = 60\,\mathrm{s}$ and $\tau_D = 0$ and both transmitter noises at their 0.004 m default (levels quoted are true levels). With the pump set to 35.4 % before the run starts, the bottom level rests at 0.501 m until automatic is engaged at 400 s. With the splitter shut the setpoint step from 0.500 to 0.600 m at 460 s rings through peaks of 0.650 and 0.613 m and troughs of 0.574 and 0.593 m and drives the top tank to 0.810 m. Opening the splitter to 100 % at 700 s drains the top tank and lifts the bottom level 88 mm in 7.95 s before the loop recovers it, and the step back to 0.500 m at 820 s on the first-order plant settles inside 0.005 m in 9.75 s, its lowest point 3.8 mm below the setpoint. The trends draw these levels without printing them, and the values are those of the same run. The picture is a composite of the application's window, the plant view above its three trend strips, with the header and the side panels cut. Callouts: (1) pump, (2) splitter valve, (3) top tank TK-1 and its gravity drain, (4) bottom tank TK-2 with the dashed target line, (5) the bottom level ringing after the setpoint step with the splitter shut.

0.50 m). Cases 9 and 10 run the predictive controller of Section 3.5.4, and case 11 is the graded challenge of Section 3.8. The experiments of this section and the next are built from case 1 with the settings chosen before Run, and Table 3.7 collects them.

The two extreme rows of Table 3.3 were converted into proportional-integral tunings with the internal model control correlation [23], with $\tau_c = \max(0.1\hat{\tau}_p, 0.8\hat{\theta}_p)$, $K_c = \hat{\tau}_p/[\hat{K}_p(\tau_c + \hat{\theta}_p)]$ and $\tau_I = \hat{\tau}_p$. The shut-splitter fit gives $\tau_c = 14.8\,\mathrm{s}$ and a slow tuning of $K_c = 66.6\,\%/\mathrm{m}$ and $\tau_I = 71.6\,\mathrm{s}$. The open-splitter fit has no dead time to protect against, so with the small negative fit taken as zero τ_c collapses to $0.1\hat{\tau}_p = 4.3\,\mathrm{s}$ and the fast tuning is $K_c = 309.4\,\%/\mathrm{m}$ and $\tau_I = 42.9\,\mathrm{s}$. The proportional gains differ by a factor of 4.6 and the integral gains K_c/τ_I by a factor of 7.7, and the process gain they were both derived from is identical. On the 3 %/m grid of the gain field the two tunings are entered as $K_c = 66$ with $\tau_I = 71.65\,\mathrm{s}$ and $K_c = 309$ with $\tau_I = 42.95\,\mathrm{s}$, and those are the values run below.

Table 3.5 The Same Setpoint Step From 0.500 to 0.600 m Under Two Tunings and Two Splitter Positions, $\sigma_n = 0$. Settling Is the Last Crossing of a ±0.005 m Band and Travel Is $\sum |u_k - u_{k-1}|$ Over the 400 s After the Step.

	slow tuning $K_c = 66$, $\tau_I = 71.65$		fast tuning $K_c = 309$, $\tau_I = 42.95$	
Splitter x_v	0 %	100 %	0 %	100 %
Peak h_2 (m)	0.6078	0.5996	0.6587	0.6000
Overshoot (%)	7.8	none	58.7	none
Peak h_1 (m)	0.6347	0	0.8354	0
IAE (m s)	4.255	3.580	4.950	0.419
Pump travel (%)	12.4	10.3	144.0	58.4
Settling (s)	117.9	159.7	210.8	12.3
Damped period (s)	—	—	79.8	—
Predicted ζ	0.532	real poles	0.199	real poles

Table 3.5 reports each tuning driving each structure through the same setpoint step from 0.500 to 0.600 m, with the transmitter noise set to zero so that the manipulated-variable travel measures control action rather than noise. Each run is case 3 with the tuning, the splitter and the noise changed before it starts, so the loop starts in automatic at rest at 0.500 m and the setpoint moves at 10 s. The predicted damping ratio is that of the dominant closed-loop pole pair of the linear model of Equation (3.11) in feedback with $K_c[1 + 1/(\tau_I s)]$.

The diagonal of Table 3.5 pairs each tuning with the structure it was identified on, and the off-diagonal entries quantify the cost of the mismatch. The fast tuning on the plant it was derived from settles the step inside 0.005 m in 12.3 s with an IAE of 0.419 m s and no overshoot, because the proportional-integral zero at −0.0233/s very nearly cancels the closed-loop pole at −0.0228/s and leaves a single closed-loop time constant of 3.96 s. Moving the same controller to the shut-splitter plant without changing a gain produces 58.7 % overshoot, five peaks above the setpoint inside the window, three of them outside the ±0.005 m band, on a damped period of 79.8 s, and 144.0 % of pump travel against the 58.4 % it spent on the fast plant. The linear model predicts that outcome closely, with a dominant pole pair at $\zeta = 0.199$ and a damped period of 76.6 s. The added lag has cost the loop essentially all of its phase margin, which is the result the effective-delay literature predicts for any controller of any structure once a dead time appears [4, 7].

The reverse mismatch produces no visible symptom and a larger performance loss. The slow tuning on the fast plant never overshoots and never rings, and it takes 159.7 s to settle against the 12.3 s the fast tuning needs, an IAE of 3.580 m s against 0.419 m s. A factor of 8.5 in integrated error is invisible on a trend unless the two runs are put side by side, which is why the slow tuning survives on a plant that could be controlled eight times more tightly. Figure 3.2 shows the first mismatch and the well-matched case produced by one controller in one run. Figure 3.3 draws four rows of Table 3.5 as bars.

One further measurement does not appear in the table. The fast tuning on the shut-splitter plant drives the top tank to 0.8354 m, which is 84 % of the rim, while moving the bottom tank by 0.100 m. The overshoot that a level trend shows is in the controlled vessel, but the inventory that comes closest to spilling is in the uncontrolled vessel.

The remedy that the structure itself suggests is to schedule the tuning on the splitter position, which is a measured variable and moves slowly. Volume 2, Chapter 6 schedules a PI loop on a gain that moves four orders of magnitude. Both entries in the IMC row of Table 3.3 come from the same correlation applied to the same plant at two settings of one valve, and interpolating them on x_v is the textbook shape of a scheduled controller. The scheduling literature is equally clear about what makes this case easy and the general case hard, because the scheduling variable here is an exogenous input

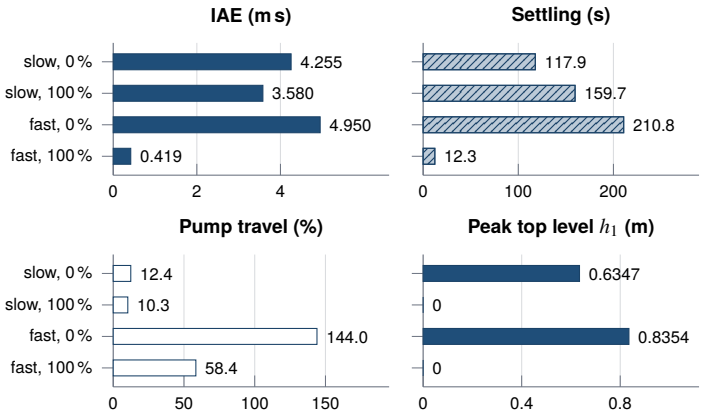


Figure 3.3 The setpoint step of Table 3.5 under the slow tuning (identified with the splitter shut) and the fast tuning (identified with it wide open), each run at splitter settings of 0 and 100 %: integrated absolute error and peak top-tank level (solid bars), settling time (hatched) and pump travel (open). Notice that each tuning settles fastest on the structure it was identified on, and that the fast tuning on the shut splitter spends the most pump travel and lifts the top tank highest, a level no loop controls.

Table 3.6 Load Disturbance Responses From a Settled 0.500 m With $K_c = 66\,\%/\text{m}$, $\sigma_n = 0$, Measured Over the 700 s After the Upset

Upset	x_v (%)	τ_I (s)	Minimum h_2 (m)	Final h_2 (m)	Final u (%)
$q_d = 0.010$	0	71.65	0.3345	0.5000	47.86
$q_d = 0.010$	0	off	0.3273	0.3796	43.30
$x_L = 100\%$	0	71.65	0.2824	0.4988	70.66
$q_d = 0.020$	0	71.65	0.1818	0.2500	100.0
$q_d = 0.020$	100	71.65	0.3020	0.4998	60.35

rather than a state, and the plant is linear in every respect except the square-root drains [8, 9]. The application schedules the model of its predictive controller on the splitter rather than the PID's tuning (Section 3.5.5), and Section 3.8 measures both against the shipped loop.

3.7 Load Disturbance Rejection

The other two disturbances behave in the way the splitter does not, because both of them move the steady state. Table 3.6 reports the responses, all with the splitter shut unless stated, the transmitter noise at zero and the slow tuning of Section 3.6 in automatic. Each run starts at rest at 0.500 m with automatic chosen before it starts, and the upset is applied at 10 s. The draw-off rows use a larger upset than case 4, whose 0.006 m/s keeps the top tank below its rim (Section 3.7.1).

3.7.1 Draw-Off Response and Proportional Offset

A draw-off of 0.010 m/s is 35 % of the 0.0283 m/s that the bottom drain passes at 0.5 m. Left uncontrolled with the pump held at the 35.36 % that held 0.500 m, Equation (3.3) puts the resting level at 0.2089 m, a fall of 0.291 m. Under proportional-only control with the bias left at that value, the loop would settle where the positional law and Equation (3.3) agree, which is the root of $c_1(\bar{u} + K_c e)/100 = c_2\sqrt{h_{2,\text{SP}} - e} + q_d$. At $K_c = 66\,\%/\text{m}$ that root is $e = 0.1204\,\text{m}$ at a pump position of 43.30 %, so proportional feedback alone would recover 59 % of the uncontrolled excursion and leave the rest as a standing offset, which is the classical result. The application runs that case directly, because the top of the integral-time range switches integral action off and a run started in

automatic keeps the 35.36 % bias of its starting steady state. The level falls to 0.3273 m and settles at 0.3796 m with the pump at 43.30 %, the predicted root to four decimal places, and the 0.120 m offset is unchanged 1500 s after the upset. The slow tuning's $\tau_I = 71.65$ s removes the offset entirely and takes the pump to 47.86 %, the value that Equation (3.3) requires.

What the successful rejection costs is not visible in the controlled vessel. Holding 0.500 m against that draw-off requires the top tank to sit at 0.9160 m by Equation (3.4), which is 92 % of the rim, and the transient carries it to 0.9319 m. With the shipped tuning, $K_c = 300\,\%/\text{m}$ and $\tau_I = 60$ s, the same draw-off carries TK-1 to the rim, where it spills for 31 s. The disturbance was applied to the bottom tank and the vessel it fills is the top one. Case 4 therefore applies 0.006 m/s, which needs TK-1 at 0.735 m, and the shipped tuning's transient peaks at 0.849 m (0.900 m even at $K_c = 1000\,\%/\text{m}$). The bottom level dips to 0.440 m at 28.5 s and is back within ±0.03 m from 44.8 s. Recovered by hand in case 5, the same draw-off is met by a pump of 42.9 % set at 10 s, and that setting completes the case at 123 s.

3.7.2 Leak Response

A 100 % leak doubles the discharge coefficient of the top tank and throws half of its drain away, so at a shut splitter the pump has to deliver twice as much to pass the same flow to TK-2. Equation (3.1) at steady state gives $u = 200 c_2 \sqrt{h_{2,\text{SP}}}/c_1 = 70.71\,\%$, exactly twice the undisturbed bias, and the loop reaches 70.66 % and 0.4988 m within 700 s while still converging. The leak also shortens τ_1 by the factor $1 + x_L$ through Equation (3.9), so a leak shortens the loop time constant while raising the pump position the same level requires. At a fully open splitter the top tank is empty and the leak has no effect at all, which is a compact demonstration that a disturbance path can be removed by a piping change rather than by a controller.

> **+ DID YOU KNOW? A Bypass That Lifts the Reachable Level**
>
> From the simulator: against a draw-off of 0.020 m/s, the shut splitter leaves the pump at 100.0 % and the bottom level at 0.2500 m. With the splitter wide open, the same tuning holds 0.4998 m with the pump at 60.35 % (Table 3.6).

3.7.3 Reachable Level Under Draw-Off

The last two rows of Table 3.6 are the same 0.020 m/s draw-off applied at the two ends of the splitter, and they do not differ in speed. They differ in whether the setpoint is reachable at all. With the splitter shut, all flow entering TK-2 must pass through TK-1, whose discharge is bounded by $c_2\sqrt{h_{\max}}$ once the top tank is brim-full. Substituting that bound into Equation (3.2) gives the highest bottom level a shut splitter can sustain against a draw-off, as shown in Equation (3.25).

$$h_2^{\max}\big|_{x_v=0} = \left[\frac{c_2\sqrt{h_{\max}} - q_d}{c_2}\right]^2 \tag{3.25}$$

At $q_d = 0.020\,\text{m/s}$ that ceiling is 0.2500 m, and the application settles at 0.2500 m with the pump parked on 100 % for 71 % of the 700 s and the top tank at the rim and overflowing for 93 % of it. The limit belongs to the plant rather than to the tuning. Opening the splitter to 100 % routes the whole pump flow around the bottleneck, and the identical disturbance is then rejected completely, to 0.4998 m after 700 s and 0.5000 m after 900 s, with the pump at 60.35 % against the 60.36 % of Equation (3.3) and the top tank empty. The IAE over the 900 s after the upset falls from 226.0 m s to 27.1 m s. The splitter therefore changes two things that are usually thought about separately, the order of the plant and the size of the load that the plant can reject, and it changes neither the process gain nor the nominal operating point. Volume 2, Chapter 3 meets a bound of the same kind in a composition that the upstream supply limits.

Table 3.7 Experiment Sequence Built From Case 1 and the Mechanism Each One Exposes

Experiment	Setup	What it exposes
Splitter bump	manual, pump fixed, x_v 0 to 50 %	zero steady gain of Equation (3.12), 0.176 m transient
Step test ×3	manual, 10 % pump step at $x_v = 0$, 50, 100 %	one $\hat{K}_p$, three $(\hat{\tau}_p, \hat{\theta}_p)$ pairs
Cross tuning	IMC from each end, run at both ends	58.7 % overshoot one way, 8.5× the IAE the other
Draw-off, P only	$q_d = 0.010$, integral off	the predicted 0.120 m standing offset
Draw-off, PI	same upset, $\tau_I = 71.65$ s	offset removed, TK-1 at 92 % of its rim
Leak	$x_L = 100$ % at $x_v = 0$, then at $x_v = 1$	a disturbance path removed by piping
Draw-off ceiling	$q_d = 0.020$ at both splitter ends	0.25 m ceiling and pump saturation
Fill and drain	from empty, pump 100 % then 0 %	overflow, dry limit and $2\sqrt{h}/c_2$

The pump sets a second limit that no tuning can move. In case 8 the smaller pump sits at 100 % from the execution that sees the step to 0.90 m, and the level climbs only to 0.763 m in 300 s, toward the full-pump level $(c_1/c_2)^2 = 0.766$ m of Equation (3.3).

3.7.4 Filling and Draining Against the Limits

Two manual runs bound the apparatus. A pump of 0 % chosen before the run starts puts both tanks at rest empty. With the pump raised to 100 % 10 s into the run and the splitter shut, TK-1 reaches the 1.0 m rim 19.3 s after the pump starts, matching the 19.3 s obtained by integrating Equation (3.1) in closed form, and it spills from then on. TK-2 rises toward the rim but does not reach it, holding 0.9979 m after 300 s, because with the splitter shut its only inflow is the discharge of a tank that is itself pinned at $h_{\max}$, so h_2 approaches h_1 asymptotically from below and can never exceed it. Any splitter opening at all removes that protection, because the bypass delivers $c_1 x_v \bar{u}/100$ into TK-2 on top of the TK-1 discharge. Shutting the pump from that state empties TK-1 in 49.8 s against the $2\sqrt{h_{\max}}/c_2 = 50.0$ s of the analytic drain expression, and empties TK-2 27.3 s later, because TK-2 keeps receiving what TK-1 is losing until TK-1 is dry.

3.8 Predictive Control on Both Plant Structures

Table 3.8 runs cases 3, 4, 8, 9 and 11 under each controller for 300 s, or 120 s for the challenge. Each run selects its controller before it starts, which gives both controllers identical events and an identical noise realization. The PID runs each compared case at its shipped tuning, $K_c = 300\,\%/\text{m}$ and $\tau_I = 60$ s, except the challenge, which ships $K_c = 60\,\%/\text{m}$ and suggests 300 %/m, while the MPC runs everywhere with the settings given in Section 3.5.4. The comparison is a second cross-mismatch experiment beside Table 3.5, because the shipped PID gain is close to the open-splitter IMC gain of Table 3.3 while the MPC model belongs to the shut splitter. Transmitter noise accounts for most of the pump travel. Over the quiet first 10 s of case 3 alone the PID moves the pump by 31.7 % and the MPC by 28.5 %.

With the splitter shut the model matches the plant and the MPC wins cases 3 and 4. At the setpoint step of case 3 it holds the pump at 100 % for 6 s, peaks at 0.618 m against 0.649 m and settles 101 s sooner. Under the draw-off of case 4 the level falls to 0.461 m against 0.440 m, and $\hat{d}$ averages −7.38 % over the last 100 s against the −7.50 % that Equation (3.3) implies. The improvement is paid for in the uncontrolled vessel, as with the fast PID tuning of Section 3.6. The model has no top tank, and its larger pump moves carry TK-1 to 0.898 and 0.983 m in the two cases against 0.807 and 0.849 m, the second only 17 mm below the rim. Case 8 is close to a tie, both controllers holding the pump at 100 % from 10.05 s and ending at 0.763 m. The IAE of the MPC is 1.0 % higher because of the first 10 s, where its plans answer the noisy readings with pump positions of 68.4 to 84.7 %

Table 3.8 Shipped PID and Predictive Controller on Five Cases With the Same Noise Sequence, and the Predictive Controller With Its Model Scheduled on the Splitter or Frozen at the Half-Open Entry

Case	Controller	IAE (m s)	IAE_h	t_s (s)	$h_{1,max}$ (m)	Travel (%)	Limit (%)
3, setpoint	PID	4.27	4.00	132.0	0.807	852	0
	MPC	2.18	1.65	31.0	0.898	750	2.0
	MPC^a_{50}	14.96	14.96	—	1.000	1164	58.7
4, draw-off	PID	3.28	2.86	114.0	0.849	825	0
	MPC	1.74	1.25	37.5	0.983	728	2.0
8, limit	PID	60.82	60.67	—	0.765	52	96.7
	MPC	61.44	61.30	—	0.765	48	96.7
9, splitter	PID	3.14	2.65	19.0	0.502	882	0
	MPC	8.68	8.68	—	0.506	5376	70.7
	MPC^c_s	1.53	1.00	5.5	0.506	574	0.7
	MPC^a_{50}	1.54	1.00	5.5	0.506	557	2.0
11, challenge	PID, $K_c = 60$	7.21	7.11	—	0.755	66	0
	PID, $K_c = 300^b$	2.59	2.51	—	0.849	331	0
	MPC	1.12	0.99	37.5	0.983	333	5.0

IAE uses the measured bottom level, as the scorecard does, and IAE_h the true one. t_s is the time after the last event from which the true bottom level stays within ±0.01 m of the setpoint, with a dash where it never does. $h_{1,max}$ is the highest true top level, Travel is $\sum|u_k - u_{k-1}|$ and Limit is the share of the run with the pump at 0 or 100 %, all on the 2 Hz log. [a]The scheduled model frozen by script at its 50 % entry, not a setting of the application. [b]Suggested, $\tau_I = 60$ s. [c]Scheduled model (case 10), which repeats the MPC rows exactly in the other four cases.

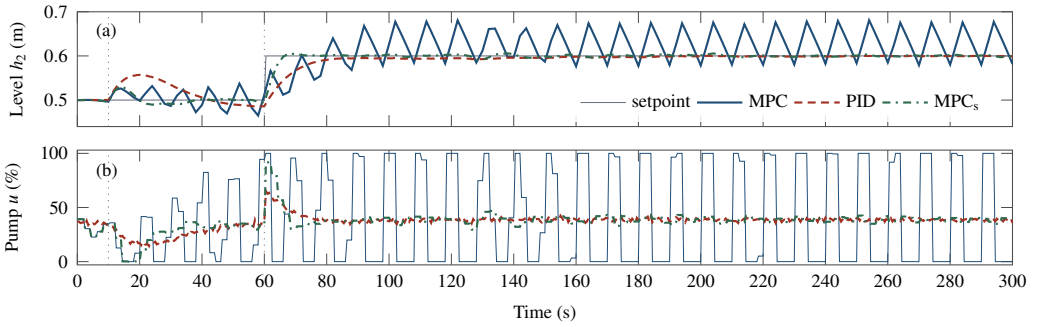


Figure 3.4 The predictive controller with its fixed model (MPC) and with its scheduled model (MPC_s, case 10) and the shipped PID on the events of case 9, each run with an identical noise realization. The splitter opens from 0 to 50 % at 10 s and the setpoint steps to 0.60 m at 60 s (dotted lines). The MPC is solid (thin in panel b), the PID dashed, MPC_s dash-dotted and the setpoint a thin gray line. (a) True bottom level and setpoint. (b) Pump position, logged at 2 Hz.

and leave TK-1 at 0.486 m against 0.500 m, so the climb at full pump starts from a smaller upstream inventory.

Case 9 opens the splitter to 50 % at 10 s, which leaves the operating point unchanged, and steps the setpoint to 0.60 m at 60 s, as shown in Figure 3.4. The MPC loses it. After the step its pump sits at 0 or 100 % for 81 % of the time, the level cycles between 0.512 and 0.681 m with a period near 10 s and never settles, and the true-level IAE is 8.68 m s, 3.3 times the 2.65 m s of the PID, whose pump stays between 14.3 and 65.2 %. The cause is the fast path of the bypass. By Equation (3.11) at $x_v = 0.5$ a 1 % pump step raises the level by 0.82 mm within one decision period, where the model predicts 0.044 mm, and by 4.34 against 0.95 mm after 10 s, though both end at 28.3 mm. The filter books the unexplained response as load, $\hat{d}$ rising from 1.5 to 28.0 % within 5 s of the opening and swinging from −52.5 to 58.5 % after the step, so each plan answers its own earlier moves with a full-range move and 106 of the 150 first moves sit on a limit. Over the first 60 s the cycle costs little, 0.78 against 1.26 m s, and the setpoint step turns it into the loss, 7.91 against 1.39 m s over the last 240 s.

TRY IT IN THE APP **9 · MPC When the Splitter Reshapes the Plant**

Load the case and run it to 300 s. The splitter opens to 50 % at 10 s and the setpoint steps to 0.60 m at 60 s. Watch the pump: the predictive controller, whose model knows only the shut splitter, throws it from limit to limit. Then run case 10, whose model follows the splitter, and the pump stays off its limits.

The loss belongs to the model, not to the tuning. With λ anywhere on its range of 0 to 3 and q_d set by script to 1, 10 or 100 times r/K_m^2, the true-level IAE of case 9 stays between 6.20 and 9.43 m s, and horizons of 40 to 600 s move it by under 1 %. Told the splitter setting, the same controller takes its 50 % entry when the splitter opens and scores 1.53 m s on the events of case 9 (case 10), against 8.68 fixed and 3.14 for the PID, as shown in Figure 3.4. After the step its pump stays between 29.6 and 91.7 %, never on a limit, and from 180 s the true level holds within 0.596 to 0.606 m. Over seeds 1001 to 1010 the scorecard IAE spans 1.48 to 1.65 m s scheduled, 8.48 to 9.13 fixed and 3.09 to 3.13 for the PID. With the splitter shut the two models give identical runs in cases 3, 4, 8 and 11.

Frozen by script on its shut entry, the scheduled controller repeats the fixed one pump for pump, so switching, carry-over and the stricter plan check add nothing. Frozen on its 50 % entry it wins case 9 (1.54 m s) but loses case 3 at 14.96 m s with TK-1 at its rim and fails the challenge (worst error 0.066 m). The gain is the model identified at the setting in force, kept in place by the schedule, as Table 3.5 found for the two IMC tunings. Volume 9, Chapter 3 makes the same point with two predictive controllers that differ only in the step test that filled their model. The fixed model at $\lambda = 3$ still scores 8.20 m s, and the PID scheduled by the IMC rule on first-order fits of the same step tests (K_c from 67 to 354 %/m) scores 4.20 m s and fails the challenge, the shipped 300 %/m being already near the open-splitter gain.

On the challenge the MPC, which does not read K_c, repeats the first 120 s of its case 4 run. Over the scored final third, from 79.2 to 120 s, its largest measured error is 0.012 m, against 0.072 m for the shipped K_c, which fails the ±0.05 m band, and 0.023 m for the suggested one. With the shipped gain the true level falls to 0.398 m at 47 s and is still 0.015 m short at 120 s. The suggested tuning, which raises K_c alone, stays inside the band with 0.027 m to spare. None of the 660 plans of the five cases was rejected. Only the three plans that answered the event in each of cases 3, 4 and 11 had a bound active, against 145 in case 8 and 112 in case 9, with up to 12 active-set iterations. A plan took 0.08 to 0.18 ms on average under Node.js and 211 and 217 µs in two Chrome runs of case 9, at most 2.7 ms, or 0.011 % of one processor core in real time, wall-clock values from one desktop computer. The MPC wins cases 3, 4 and 11, trails case 8 by 1.0 % and loses case 9 by a factor of 3.3 unless its model is scheduled.

The schedule omits the splitter's own transient, which still lifts the level to 0.532 m in case 10. With the splitter at 25 to 100 % before the step of case 3 it scores 1.30 to 1.44 m s, the PID 1.38 to 1.96 and the fixed model 3.54 to 16.24, though the PID leads on the true level from 60 % up (0.68 to 0.71 against 0.80 to 0.82 m s). Between table entries the schedule loses little. At splitter settings of 37 and 60 %, where it picks the 25 and 50 % entries, case 3 scores 1.43 and 1.31 m s scheduled, 1.68 and 1.46 under the PID, and 5.41 and 12.13 under the fixed model, whose pump sits at a limit for 36 and 92 % of the time after the step. An entry ten or twelve percentage points away therefore costs far less than the shut-splitter entry. Over 40 runs (every case plus extremes of splitter motion, load, noise and geometry) under the PID and both models with either estimator, the scheduled controller never overflows and is never longer outside ±0.03 m than the safest shipped structure, while the fixed model overflows TK-2 for 60.9 s with a 0.65 m rim. Its TK-1 peak exceeds the fixed model's in three runs, 0.907 against 0.825 m (PID 0.978 m) for a splitter taken to 50, 100 and 0 %, 0.891 against 0.874 m when it shuts at 150 s, and 0.470 against 0.443 m at 25 % throughout. A scheduled plan took a median of 85 µs under Node.js on a loaded host.

Table 3.9 Shipped PID, Relay-Written PI, Two Counterfactual PI Loops and the MPC on Cases 3 and 4 Over 900 s With One Noise Seed

Case	Controller	IAE (m s)	IAE_h (m s)	e_{pk} (m)	$h_{1,\max}$ (m)	Travel (%)	Travel^0 (%)
3, setpoint	PID, shipped	6.17	4.31	0.049	0.807	2424	114.5
	PI, relay	5.92	4.22	0.012	0.674	1084	30.9
	PI[a]	5.43	3.45	0.034	0.777	2428	95.1
	PI[b]	6.19	4.22	0.031	0.704	1082	35.9
	MPC	4.30	2.60	0.018	0.898	1973	216.1
4, draw-off	PID, shipped	5.17	3.10	0.060	0.849	2396	64.3
	PI, relay	9.40	8.17	0.082	0.752	1072	21.5
	PI[a]	5.48	3.81	0.061	0.811	2400	49.1
	PI[b]	6.67	4.67	0.080	0.800	1070	28.1
	MPC	3.82	2.12	0.039	0.983	1914	223.5

PID, shipped: $K_c = 300$, $\tau_I = 60$ s. PI, relay: $K_c = 134$, $\tau_I = 149$ s. [a] $K_c = 300$, $\tau_I = 149$ s. [b] $K_c = 134$, $\tau_I = 60$ s. Neither counterfactual is shipped or written. e_{pk} is the true overshoot of the new setpoint in case 3 and the largest true deviation after the draw-off in case 4. Travel^0 repeats Travel with $\sigma_n = 0$. Other columns as in Table 3.8. Either estimator gives the same MPC rows.

3.8.1 Estimator Choice and the Relay Test

Each run here lasts 900 s (three trend windows), the challenge 120 s, uses the shipped noise seed and has all of its settings entered before it starts. Over the 1799 executions of case 9 the two estimates of the bottom level differ by at most $2.35 \times 10^{-7}\sigma_n$, a residue of the filter's Riccati iteration stopped 3.3×10^{-9} short of convergence, and in every case, case 10 included, the two estimators give pump records within 1.15×10^{-5} % of one another with no bound active, no window rejected and no failed plan, the filter's own estimate coming no closer than 0.053 m to a bound, so no trend compares them.

Against the true bottom level, read by the measuring script alone, the two are equally accurate, with a root-mean-square error of 1.55 to 2.36 mm (under $0.6\sigma_n$) and a largest error of 5.6 to 8.8 mm in the nine cases with the splitter shut. The open bypass of case 9 raises them to 24.4 and 44.8 mm. The model level $\bar{h}_m + \hat{z}_1$ misses the true top level by up to 0.354 m in case 4, where $\hat{d}$ carries the draw-off, and by 0.753 m in case 9. On case 9 an estimate took a median of 1410 µs under Node.js and a plan 96 µs, about 0.3 % of each 2 s decision for four estimates and a plan, and in Chrome estimates and plans took up to 0.11 % of a core at 1× and 6.3 % at 10×, within a factor of two of the budget of a tenth of a core.

The relay test was started at 30 s of case 3 with its setpoint step removed, the bottom level at 0.50 m and u_0 at 35.34 %, against the 35.36 % that holds the level, so δ_r was the full 5 %. It converged after 8 cycles and 280 s with $\alpha = 0.0191$ m, $D = 0.503$, $P_u = 67.5$ s and $K_{cu} = 428$ %/m, and it wrote $K_c = 134$ %/m and $\tau_I = 149$ s in place of the shipped 300 and 60 s, labeled because $\psi = 38.9°$. All ten further seeds (1001 to 1010) converged as well, giving K_{cu} from 271 to 383 %/m, P_u from 63.9 to 73.0 s and ψ from 27.1° to 35.8°.

The measured point belongs to the experiment more than to the tanks. Two lags in series approach −180° without reaching it, and sampled with a zero-order hold at Δt_c the model of Equation (3.19) first crosses −180° at a period of 13.1 s, where the ultimate gain is 10 100 %/m. A relay with hysteresis settles where the imaginary part of the frequency response equals $-\pi\varepsilon_r/(4\delta_r)$, which on the sampled model predicts 410 %/m at 77.3 s, and the same relay on the noise-free model returns 407 %/m at 77.1 s. The measured cycle is 5.2 times slower than the ultimate one and its K_{cu} a twenty-fourth of the ultimate gain, although ψ is only 39°. With $\sigma_n = 0$ the simulator's test starts from 35.36 %, runs at $D = 0.498$ and returns 416 %/m at 77.0 s, 3 percent below the noisy gain at a period 14 percent longer.

Started at 300 s of case 3, settled at 0.60 m, the relay centers on 38.46 % against the 38.73 % that holds the level, spends $D = 0.506$ of each cycle high and converges after five cycles on $K_{cu} = 293$ %/m and $P_u = 74.4$ s, writing $K_c = 91.6$ %/m and $\tau_I = 164$ s. With $\sigma_n = 0$ it starts at 38.67 %, cycles at $D = 0.502$ and returns 389 %/m at 82.1 s. In both runs $\sin(\pi D)$ of Equation (3.24) departs from one

by less than 0.001, so the ratio 0.75 of the two gains comes from the wider noisy swing, 0.0248 against 0.0203 m.

The written constants win the setpoint step narrowly and lose the draw-off in Table 3.9. On the setpoint step the true-level IAE falls from 4.31 to 4.22 m s, and the loop reaches 0.60 m with an overshoot of 12 mm against 49 mm. On the draw-off it rises from 3.10 to 8.17 m s and the largest deviation from 0.060 to 0.082 m, because the integral gain K_c/τ_I falls from 5.00 to 0.899 %/(m s), 18 percent of the shipped value. The counterfactual rows show that neither constant alone accounts for that loss. The shipped K_c with the written τ_I scores 3.81 m s and the written K_c with the shipped τ_I scores 4.67 m s. The relay-tuned loop saves pump travel and top-tank inventory. Its travel is 1084 against 2424 % on case 3, a ratio of 0.447 equal to the gain ratio $134/300 = 0.447$, because transmitter noise passed through K_c makes up 95 to 98 percent of the travel of both loops, as Travel^0 shows, and TK-1 peaks at 0.674 m against 0.807 m.

3.9 Technology in Practice

The loop of Section 3.3 reads LT2 through white noise and moves a pump whose flow is strictly proportional to its signal. Figure 3.5 draws it with the devices a plant pair of tanks would carry in the place of the schematic's symbols.

What Measures It

A level in an open tank can be read from the pressure at its floor, from an echo timed from its roof, or from a float [24]. The magnetostrictive transmitter is a float instrument that times a pulse in a rod. It can carry two floats, one on the surface and one on an interface between two liquids, and with digital communication it reports both levels over one pair of wires [24]. No range, accuracy or response time is given here, because none was found in an open source that names no maker. The noise σ_n of Equation (3.7) stands for the moving surface of TK-2, and the model's transmitters have no lag, drift or fouling. LT1 gives an indication only, so no loop watches the vessel that comes closest to its rim.

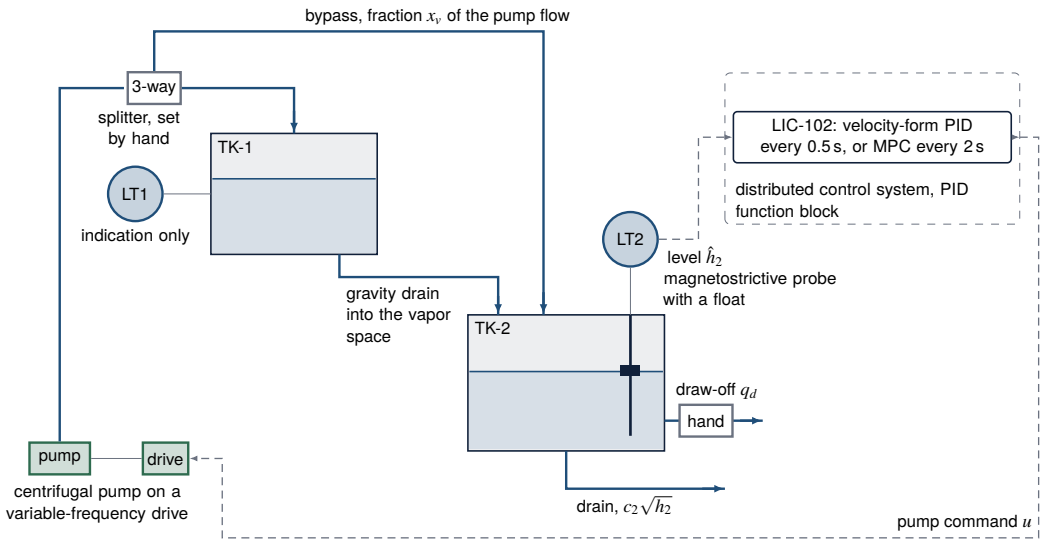


Figure 3.5 The level loop of the chapter with plant devices in place of the schematic's symbols. LT2 reads the bottom level with a float on a magnetostrictive probe, and LIC-102 on the control platform sets the speed of the pump through its drive. Notice that the splitter and the draw-off are hand valves the loop never moves, and that LT1 on the top tank only indicates, so no loop guards the vessel that the bypass can empty and a load can fill.

TECHNOLOGY **Sensor: A Magnetostrictive Level Transmitter**

A magnetostrictive transmitter hangs a metal rod, the waveguide, into the vessel. A light ring-shaped float with a magnet rides on the rod and follows the liquid surface. A current pulse sent along the rod meets the float's magnetic field and starts a twisting pulse that travels at the speed of sound to a sensing head at the top, and its travel time gives the distance to the float. Sound keeps a steady speed in a metal rod over wide ranges of temperature and pressure, which gives the method a stable calibration and fine resolution. Its weak point is the close fit of float on rod: solids or sludge can bind the float, and the reading then stops following the level [24].

What Moves It

The pump of Equations (3.1) and (3.2) has no head dependence, no minimum flow and no slew limit. A plant pump on a variable-frequency drive has a flow that falls as the head rises and a speed that changes at the rate its drive allows [24]. As a final element a pump also differs from a valve. A stopped pump does not necessarily stop the flow through it, and a drive offers fewer choices of a fail-safe state [24]. The splitter is a three-way valve, a body with three ports that divides one stream between two directions [24]. In the model it is set by hand and moves instantly, and its two paths have no resistance of their own.

TECHNOLOGY **Final Element: A Centrifugal Pump on a Variable-Frequency Drive**

The centrifugal pump is the most common pump in the process industries. On a variable-frequency drive, the line power is rectified and then rebuilt at the frequency the controller asks for, and the motor's speed follows that frequency. Running a pump only as fast as needed saves power and wear. Speed alone does not set the flow, because a centrifugal pump's discharge pressure falls as its flow rises. A pump also changes flow more slowly than a valve, since drives limit acceleration and deceleration, and a drive shuts down on overcurrent or overvoltage whatever the controller asks [24].

Where the Controller Runs

A distributed control system runs production at one site in refining, chemical manufacturing, water treatment and power generation. Its loops may be nested or cascaded, and they run with cycle times from milliseconds to minutes [25]. LIC-102 executes every 0.5 s and the predictive controller decides every 2 s, both inside that range. The velocity-form PID of Equation (3.17) is the PID function block of such a platform, with the bumpless transfer of Section 3.5. The predictive controller and its table of models would run in the supervisory level that oversees the loop controllers [25].

IN PRACTICE **Ask How Far the Level May Move**

An instrument handbook treats a change of tuning as seriously as the reconfiguration of a field instrument. Before any change, it advises asking the operators how far and how fast the process variable may move, and keeping clear of alarm and shutdown limits, because few loops in a plant stand alone. Its own example is three vessels in series, drained by gravity, with one transmitter on the last. That loop is hard to control, not least because the vessels in between may overflow while nothing measures them [24].

What Is Modernizing It

Three developments touch this loop. The first is variable-speed pumping, which saves the energy that a throttling valve dissipates, although a pump changes flow more slowly than a valve and cannot positively stop it [24]. The second is the digital level instrument that reports more than one level from one probe, such as the two-float transmitter above [24]. The third is the correction of a predictive controller's model from routine operating data, without a new step test, the subject of the open challenge in Section 3.11 [26].

3.10 Limitations

The simulator is a teaching instrument rather than a design tool. Two of its elements are stylized on purpose. A drain that empties a 1 m vessel in 50 s compresses the two-lag response into minutes, and no structural result depends on it. A draw-off reported as a level velocity keeps every term of Equation (3.2) in one unit and makes the ceiling of Equation (3.25) a one-line calculation. The non-interacting discharge law is imposed rather than checked, so the model does not detect the flooded-nozzle condition under which the same two vessels would become a genuinely coupled pair, and it must not be used to study that case. Both tanks share one area, one drain coefficient and one rim height, so mismatched vessels cannot be represented. The pump is ideal and its flow does not fall with discharge head, which removes the one physical mechanism that would couple the bottom level back to the pump. The splitter is a pure fraction with no dynamics and no pressure interaction between its branches, so the plant order changes exactly when the fraction moves and no real valve does that. Mass is not conserved during an overflow and the spill is not routed anywhere. Transmitter error is white noise and nothing more, which leaves out the slow drift and surface effects that dominate a real level reading. No flow controller sits under the pump, no second actuator exists and the PID's tuning is never scheduled, so the remedy that Section 3.6 names runs only by script. The noise generator is seeded, so even a noisy run repeats exactly at every speed setting. The square-root drains, the four hard limits, the triangular structure and the 0.5 s sample period are retained because each produces behavior the tuning has to accommodate, and Section 3.3.5 says where each one stops being true.

With the fixed model the predictive controller plans on two lags fitted at 0.50 m with the splitter shut. Without the bypass path its predicted response over one decision period is 18.6 times too small at the half-open splitter, and without the square-root drains the plant gain at 0.60 m is 9.5 % above its own. It has no top-tank level or rim, so nothing protects TK-1, which reached 0.983 m in case 4, and the splitter, the draw-off and the leak reach it only through $\hat{d}$. Its Kalman gains assume the default noise and are never recomputed, its q_d sits at the top of the tuning grid, and its comparison rests on one noise sequence. The scheduled model knows the splitter only as the setting it is told, so a valve moved without telling it leaves the plan on the wrong entry, and every entry is fitted at 0.50 m and none holds the splitter's own transient. The bounds of the moving horizon estimator act on the states of that fitted model, whose first state misses the true top level by up to 0.753 m in case 9, so they could not protect TK-1 even when active. The relay keeps its center fixed during a test and, on delay-free lags, measures its own band as much as the plant.

3.11 Discussion

The situation this application reproduces is ordinary. A surge drum is bypassed for maintenance, a spare vessel is lined up in parallel, or a plant runs a summer and a winter lineup, and in each case the piping between two units changes while the material balance does not, so every steady-state indication looks normal while every dynamic indication has moved. Inventory between units exists so that upstream variation need not pass downstream at once, and the structure deciding which flow regulates which holdup is settled before any loop is tuned [6]. Taking a vessel out of that chain removes a lag and a buffer together. The last two rows of Table 3.6 show both, because opening the splitter turned a load the loop could not reject into one it rejects to within 0.2 mm in 700 s. A gain and a time constant identified on one lineup belong to that lineup, as the scheduling surveys [8, 9] and the longer history of feedback design [27] both record, and the splitter separates what a single step test cannot. The gain is a property of the material balance and survives the change, while the phase is a property of the path and does not.

⚠ OPEN CHALLENGE Keeping a Controller's Model True to the Plant

A predictive controller is only as good as its model, and a plant drifts from the model identified at commissioning as feedstocks, equipment and operating conditions change. A new step test disturbs production and keeps operators watching, while routine data carry little excitation because setpoints may stay fixed for months. Progress would be a model corrected from routine data with no injected signal, as one simulation study has shown for a pilot distillation column [26].

Methods beyond the two implemented controllers can be built on the model of Section 3.3. A predictive controller built on Equations (3.1) and (3.2) rather than on fitted lags would hold the two rims and the two dry bottoms as output constraints and would recognize the ceiling of Equation (3.25) as unreachable before the pump saturates. Moving horizon estimation or Kalman filtering on both level readings would recover the splitter fraction and the leak as parameters entering through known structure, which the estimator of Section 3.5.6 does not attempt and which would free the schedule of Section 3.5.5 from what the operator reports. A multivariable formulation could move the pump and the splitter together, trading top-tank inventory against bottom-level regulation. The APMonitor framework implements this combination of nonlinear modeling, estimation and predictive control [28], and the Gekko Optimization Suite brings it to Python so that the exported runs feed the estimator directly [29].

3.12 Conclusion

Two gravity-drained tanks behind one pump and a three-way splitter valve form an interactive laboratory in which the operator changes the order of the plant rather than its operating point. Linearization gives a lower-triangular pair of non-interacting lags with an upper time constant of $(1-x_v)\tau_2$, a pole and a zero that cancel when the splitter is wide open, and a process gain $c_1\tau_2/100$ free of the splitter, and step tests in the application return 0.0323 m/% at every setting while the apparent dead time falls from 18.5 to −0.6 s. Internal model control tunings from the two ends differ by a factor of 4.6 in proportional gain, and each on the other structure gives 58.7 % overshoot in one direction and 159.7 s of settling against 12.3 s in the other. A linear predictive controller fitted with the splitter shut cuts the true-level IAE of the shipped PID from 4.00 to 1.65 m s on the setpoint step but cycles against the pump limits half open, 8.68 against 2.65 m s, while the same controller scheduled on the told splitter setting scores 1.00 m s there, a gain that freezing the schedule traces to the model. A bounded moving horizon estimate leaves every pump record within 1.2×10^{-5} % of the filter's, and a relay test writes $K_c = 134\,\%/\mathrm{m}$ and $\tau_I = 149\,\mathrm{s}$, which raise the true-level IAE of the draw-off from 3.10 to 8.17 m s. A 0.020 m/s draw-off that the shut-splitter plant cannot reject above 0.2500 m is rejected completely with the splitter open, so the valve bounds the load as well as the bandwidth. A fuller model would add a PID gain schedule on the splitter, a flooded-nozzle option that makes the pair interacting, a head-dependent pump curve and unequal tank areas. The application is an accessible browser instrument for level control experiments in which the flow path, not the operating point, sets the plant order.

KEY POINTS

- The splitter changes only the path the water takes, not where the bottom level settles.
- Step tests return 0.0323 m/% at every setting, while the apparent dead time falls from 18.5 to −0.6 s.
- Each end's tuning on the other structure overshoots 58.7 %, or settles in 159.7 s against 12.3 s.
- A predictive controller fitted with the splitter shut cycles against the pump limits half open unless its model follows the splitter.

Where else this applies. Volume 15, Chapter 7 strings five such lags in series.

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] S. Skogestad, "Simple analytic rules for model reduction and PID controller tuning," *Journal of Process Control*, vol. 13, no. 4, pp. 291–309, 2003.

[5] K. H. Johansson, "The quadruple-tank process: a multivariable laboratory process with an adjustable zero," *IEEE Transactions on Control Systems Technology*, vol. 8, no. 3, pp. 456–465, 2000.

[6] J. J. Downs and S. Skogestad, "An industrial and academic perspective on plantwide control," *Annual Reviews in Control*, vol. 35, no. 1, pp. 99–110, 2011.

[7] K. J. Åström, "Limitations on control system performance," *European Journal of Control*, vol. 6, no. 1, pp. 2–20, 2000.

[8] D. J. Leith and W. E. Leithead, "Survey of gain-scheduling analysis and design," *International Journal of Control*, vol. 73, no. 11, pp. 1001–1025, 2000.

[9] W. J. Rugh and J. S. Shamma, "Research on gain scheduling," *Automatica*, vol. 36, no. 10, pp. 1401–1425, 2000.

[10] E. P. Gatzke, E. S. Meadows, C. Wang, and F. J. Doyle, III, "Model based control of a four-tank system," *Computers & Chemical Engineering*, vol. 24, no. 2-7, pp. 1503–1509, 2000.

[11] H. Pan, H. Wong, V. Kapila, and M. S. de Queiroz, "Experimental validation of a nonlinear backstepping liquid level controller for a state coupled two tank system," *Control Engineering Practice*, vol. 13, no. 1, pp. 27–40, 2005.

[12] C. C. Ko, B. M. Chen, J. Chen, Y. Zhuang, and K. C. Tan, "Development of a web-based laboratory for control experiments on a coupled tank apparatus," *IEEE Transactions on Education*, vol. 44, no. 1, pp. 76–86, 2001.

[13] E. J. Ionascu, "Drainage time and shape: inequalities from Torricelli's law," arXiv preprint 2511.00023, 2025.

[14] O. Tramis, E. Merlin-Anglade, G. Paternoster, M. Rabaud, and N. M. Ribe, "Torricelli's curtain: morphology of horizontal laminar jets under gravity," *Physics of Fluids*, vol. 33, no. 6, p. 062116, 2021.

[15] K. J. Åström and T. Hägglund, *Advanced PID Control*. Research Triangle Park, NC: ISA – The Instrumentation, Systems, and Automation Society, 2006.

[16] 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.

[17] 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.

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

[19] 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.

[20] B. Aufderheide and B. W. Bequette, "Extension of dynamic matrix control to multiple models," *Computers & Chemical Engineering*, vol. 27, no. 8-9, pp. 1079–1096, 2003.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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, Gaithersburg, MD, Special Publication 800-82r3, 2023.

[26] Y. Kono and M. Koizumi, "Model life extension for continuous process: non-invasive correction of model-plant mismatch with regularization," in *2023 European Control Conference (ECC)*, 2023, pp. 1–8.

[27] K. J. Åström and P. R. Kumar, "Control: A perspective," *Automatica*, vol. 50, no. 1, pp. 3–43, 2014.

[28] 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.

[29] 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.

3.13 Exercises

3.1 Concept check · 2 min

Case 1 holds the bottom level at 0.50 m in manual with the pump at 35.36 % and the splitter shut. The splitter is opened to 50 % and the pump is left where it is. What does the bottom level do?

(a) It settles about 0.18 m higher, because half of the pump flow now reaches TK-2 without delay

(b) It falls at first, because the top tank drains away the flow it used to pass on to TK-2

(c) It stays at 0.50 m throughout, because the process gain does not depend on the splitter

(d) It rises for about half a minute and returns to 0.50 m, because the splitter acts through a zero at the origin

3.2 Concept check · 1 min

The same 10 % pump step is made from rest at 0.5013 m with the splitter at 0, 50 and 100 %, as in Table 3.3. Which result is the same in all three tests?

(a) The fitted process gain, because every unit of pump flow leaves through the bottom drain

(b) The apparent dead time, because the two tanks and their drains are the same in every test

(c) The fitted time constant, because the bottom tank's time constant does not move with the splitter

(d) The top-tank level, because the pump that holds the bottom level does not move with the splitter

3.3 Concept check · 2 min

The shut-splitter step test of Table 3.3 returns a fitted gain of 0.0323 m/%, while Equation (3.11) gives 0.0283 m/% at the starting point. Which explanation is right?

(a) The two-point fit reads part of the apparent dead time as extra gain

(b) The step lifts the level to 0.8245 m, and the local gain grows with level along the way

(c) Transmitter noise biases the final level that the fit reads upward

(d) The 0.05 s Euler substep accumulates error over the long response

3.4 Theory and calculation · 5 min

Two tanks have the default coefficients $c_1 = 0.08\,\mathrm{m/s}$ and $c_2 = 0.04\,\mathrm{m}^{1/2}/\mathrm{s}$, no leak and no draw-off, and hold the bottom level at 0.50 m.

(a) Find the pump position that holds this level and the resting top-tank level at splitter settings of 0 and 25 %. *Checkpoint:* 35.36 %, then 0.500 m and 0.281 m.

(b) At $x_v = 0.25$, find τ_1, τ_2, τ_z and K_p of Equation (3.11). *Checkpoint:* 26.52 s, 35.36 s, 6.63 s and 0.0283 m/%.

(c) Write the denominator of Equation (3.11) in the standard form of Equation (1.11) and find the natural period and the damping factor at $x_v = 0$ and at $x_v = 0.25$. *Checkpoint:* 35.36 s and 1.000, then 30.62 s and 1.010.

3.5 Theory and calculation · 8 min

The plant of the previous exercise holds 0.50 m with the splitter shut and $h_{\max} = 1$ m. A draw-off of $q_d = 0.010\,\mathrm{m/s}$ is applied to the bottom tank.

(a) The pump is held at 35.36 %. Find the level at which the bottom tank comes to rest. *Checkpoint:* 0.2089 m.

(b) A PI loop restores 0.50 m. Find the pump it settles at and the top-tank level that this pump requires. *Checkpoint:* 47.86 % and 0.916 m.

(c) Find the largest draw-off that the shut-splitter plant can reject at a bottom level of 0.50 m, and the highest level it can hold against 0.020 m/s. *Checkpoint:* 0.0117 m/s, then 0.250 m.

3.6 Theory and calculation · 5 min

The PI tunings of Section 3.6 come from the internal model control rule $\tau_c = \max(0.1\hat{\tau}_p, 0.8\hat{\theta}_p)$, $K_c = \hat{\tau}_p/[\hat{K}_p(\tau_c + \hat{\theta}_p)]$ and $\tau_I = \hat{\tau}_p$, applied to the rows of Table 3.3.

(a) Compute K_c and τ_I from the shut-splitter row and from the half-open row. *Checkpoint:* 66.6 %/m and 71.6 s, then 174.0 %/m and 51.3 s.

(b) The open-splitter row gives $K_c = 309.4\,\%/\mathrm{m}$ and $\tau_I = 42.95$ s. Find the ratios of the proportional gains and of the integral gains K_c/τ_I between the open and the shut tunings. *Checkpoint:* 4.6 and 7.7.

(c) The splitter position is not reported to the control room. Using Table 3.5, say which of the two end tunings to install. *Checkpoint:* The shut-splitter tuning.

3.7 App challenge · 9 min

Load case 2, Open-loop actuator step. The pump steps from 35.36 to 45.36 % at 10 s. Run to 160 s at 5× speed with the default seed and export the CSV file. Press the reload button, set the Splitter bypass to 100 % before Run, and repeat.

(a) From Equation (3.3), find the level at which the new pump settles and the gain of the step. *Checkpoint:* 0.823 m and 0.0323 m/%.

(b) In each file find the times after the step at which the true bottom level, column `true_pv`, has covered 28.3 % and 63.2 % of the change toward that level. Interpolate between rows. *Checkpoint:* Shut 42.3 and 90.0 s, open 13.7 and 42.3 s, each within 0.5 s.

(c) Fit $\hat{\tau}_p$ and $\hat{\theta}_p$ to both tests by the two-point method of Section 3.4. *Checkpoint:* Shut 71.6 s and 18.5 s, open 42.9 s and −0.6 s, each within 1 s.

3.8 App challenge · 8 min

Load case 3, Closed-loop setpoint step. Before Run set $K_c = 309$ and $\tau_I = 42.95$ s, the open-splitter tuning of Section 3.6, set the Lower level sensor noise to zero in Show Model and the Splitter bypass to 100 %. Run to 410 s at 10× speed with the default seed. Then set the splitter to 0 %, press Reset and run to 410 s again, exporting the CSV file of the second run.

(a) Read the scorecard IAE and overshoot of the first run. *Checkpoint:* 0.419 m s within 0.005 m s, no overshoot.

(b) Read the same tiles for the second run, and from column `true_pv` the period between the first two peaks. *Checkpoint:* 4.950 m s within 0.01 m s, 58.7 %, then 79.5 s within 1 s.

(c) Find the highest true top-tank level of the second run in column `level1`, and compare it with the rim. *Checkpoint:* 0.835 m, 84 % of the rim.

3.9 App challenge · 8 min

Load case 4, Disturbance rejection, which applies a draw-off of 0.006 m/s at 10 s under $K_c = 300\,\%/\mathrm{m}$. Before Run set the integral time to 500 s, which switches integral action off, and the Lower level sensor noise to zero. Run to 300 s at 5× speed with the default seed and export the CSV file.

(a) The loop now acts about the fixed bias of 35.36 %. Solve $c_1(\bar{u} + K_c e)/100 = c_2\sqrt{0.5 - e} + q_d$ for the steady offset e and the pump. *Checkpoint:* $e = 0.0223\,\mathrm{m}$ and 42.06 %.

(b) Read `true_pv` and `mv` at 300 s. *Checkpoint:* 0.4777 m within 0.001 m and 42.06 % within 0.05 %.

(c) Find the level at which the draw-off would leave the bottom tank with the pump held at 35.36 %, and the share of that fall the proportional loop recovers. *Checkpoint:* 0.3104 m and 88 %.

3.10 Application area · 3 min

On a pair of plant tanks the outlet nozzle of TK-1 enters TK-2 at 0.40 m above its floor, and the loop holds TK-2 at 0.50 m.

(a) Does the discharge law of the model hold at this operating point? Name the entry of the state matrix of Equation (3.8) that it sets to zero and say what replaces it. *Checkpoint:* No. The upper-right entry is no longer zero.

(b) Does the steady bottom level still depend on the pump alone, whatever the splitter setting? Answer from the balances. *Checkpoint:* Yes.

3.11 Application area · 4 min

A surge drum between two units is bypassed for two weeks of maintenance, the way the splitter bypasses TK-1.

(a) Which results of a step test made before the bypass still hold, and which must be measured again? *Checkpoint:* The gain holds. The time constant and dead time do not.

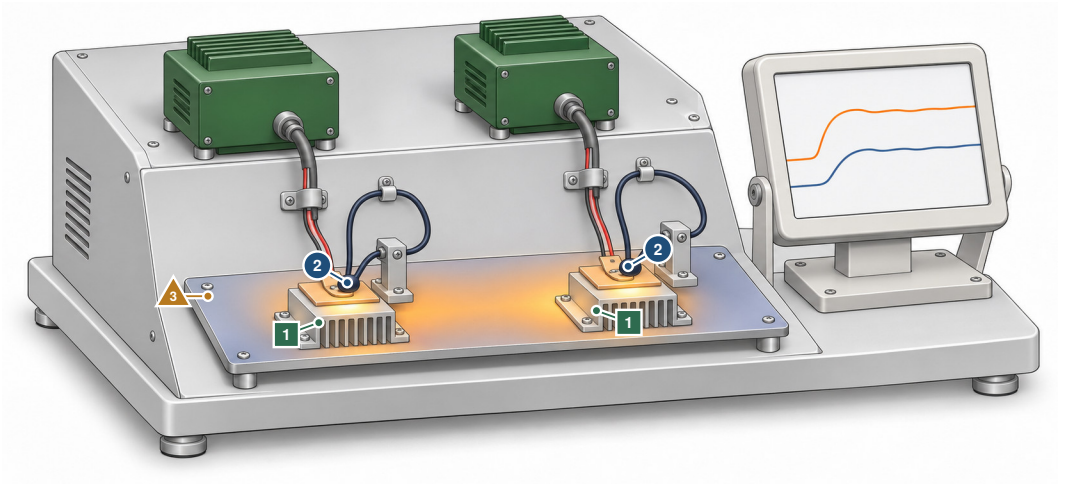
(b) Using the 0.020 m/s draw-off of Table 3.6, name one gain and one cost of the bypass. *Checkpoint:* A higher load ceiling and the lost buffer.

3.12 Review the agent · 5 min

An AI assistant reviews a run of case 9 and writes: “The process gain is 0.0283 m/% at every splitter setting, so the MPC’s model is still right after the splitter opens. The cycling between the pump limits after the setpoint step is a tuning problem: a move suppression of 0.3 is too small. Raise it to 3, the top of its range, and the MPC will beat the PID’s 3.14 m s.” Find the flaw and say how you would check the claim in the app.

CHAPTER 4

A Two-Heater Digital Twin That Runs Beside the Device It Models



The controller moves 1 the power to each of the two heaters

The controller reads 2 the temperature sensor beside each heater

What disturbs it 3 the room air around the board

Notice the warmth spreading across the shared plate, so each sensor reads its own heater and some of its neighbor, on a real board far smaller than drawn and with the power switching on the board itself rather than in separate boxes. Illustration, not to scale.

IN PLAIN TERMS **Two Heaters That Warm Each Other**

The main idea. A small board holds two heaters and two temperature sensors, and a controller sets the power of each heater to hold its sensor at a target. A computer model of the board runs beside it, so real and predicted temperatures share one screen.

The challenge. Each sensor warms more slowly than its heater, so the controller acts on old news. The heaters also warm each other through the board, so a change on one side upsets the other.

Why it matters. Large industrial heaters have the same slow sensors and shared heat, and a test on a model costs far less than a test on the plant.

THREE WAYS TO READ THIS CHAPTER

⊘	**Explore**	11 min	the first page, the history of take-home kits (Section 4.2), the fact on thermistor wiring (Section 4.3), the try-it on the lag (Section 4.4), the try-it on the PID types and the fact on heater travel (Section 4.7) and the key points
🕮	**Study**	60 min	the whole chapter with its three pause points and the twelve exercises at its end
⛰	**Practice**	20 min	the first page, the results (Section 4.7), Technology in Practice (Section 4.8), Limitations (Section 4.9) and the open challenge

The control question Two heaters share one small board, and each thermistor lags its heater by more than two minutes. What does a step test reveal about the instrument and the coupling, what do the three velocity forms of PID do differently, and how far does a twin that is its controller's own model speak for the device beside it?

Learning objectives

- Derive the steady-state gain matrix and the relative gain of the two-heater balance, and the heater powers that hold a pair of setpoints.
- Fit a first-order plus dead-time model to a heater step and explain why its time constant belongs to the thermistor and its dead time to the heater.
- Predict the first-cycle heater move of the three velocity forms on a setpoint step, and explain why they coincide on a load step.
- Compute the heater travel that sensor noise causes through the proportional and derivative terms.
- Explain how a predictive controller with the conduction path in its model holds the second loop where two PID loops cannot, and why the result is a best case.

Prerequisites Sections 1.4.2 (energy balance), 1.4.4 and 1.4.6 (lags in series and apparent dead time), 1.4.8 (two-point fit) and 1.6.2 (velocity form). Section 1.6.3 is needed only for the predictive controller and its estimators.

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

See also Volume 2, Chapter 4 for step tests and tuning on an industrial thermal loop, Volume 2, Chapter 2 for the velocity-form PID in detail and Volume 2, Chapter 3 for a relay test placed by its switching band. Volume 5, Chapter 11 for a temperature that no instrument reaches and Volume 5, Chapter 2 for a temperature loop through a fuel delay.

4.1 Introduction

The Temperature Control Lab (TCLab) is a shield for an Arduino board carrying two transistors used as heaters and two thermistors used as temperature sensors, and it was introduced as a hardware benchmark for modeling and control methods precisely because a physical device imposes constraints that a simulation study does not, including a cycle time that must be met, discrete sampling, communication overhead and model mismatch [1]. It has since carried undergraduate proportional-integral-derivative (PID) experiments and remote laboratory sections in which every student holds the same board [2]. The device is inexpensive and slow enough that a full step response fits inside a class period.

The physical fundamentals are the ones a first course in process dynamics establishes, and Chapter 1 develops them on a heated tank (Section 1.4.2). Each heater is an energy balance on a lumped capacitance driven by electrical dissipation and losing heat to the surroundings and to its neighbor, and each thermistor is a second capacitance connected to its heater through a thermal resistance, so the measured temperature is separated from the state that generates it by a first-order lag [3]. That is the structure of the heater path of the running example behind its thermowell, Equation (1.14), without the transport delay. The experiments that follow from that structure are step testing and first-order plus dead-time identification, steady-state gain and interaction between two loops, the difference between a state and a measurement, sampled control with a finite cycle time, actuator saturation and anti-windup, derivative action on error against derivative action on measurement, and the amplification of measurement noise by a derivative term [4, 5]. Thermal loops with this structure, a slow plant behind a lagging measurement and an actuator that saturates at both ends of its range, are common throughout process plants, so what is learned on a two-transistor board transfers to equipment that is orders of magnitude larger.

Simulation and hardware are usually kept apart when the TCLab is taught. A student first runs a simulation to choose gains, then connects the board and runs the same gains on hardware, and the two runs are carried out in separate scripts and often in separate weeks, so the comparison between prediction and measurement is reconstructed after the fact. However, the model and the device can

be driven from the same controller instance in the same process, so that a single setpoint change produces two trends on one time axis and the mismatch is read directly.

This chapter works with a four-state TCLab simulation and a physical TCLab under one velocity-form PID implementation inside a single browser page. It measures what the three selectable arrangements of that implementation do to a setpoint change and to a load disturbance. Lumped thermal modeling and PID structure are both long settled. The chapter writes down the twin, the serial protocol layer, the latched cutoff and the three difference equations in enough detail to be reproduced elsewhere, and reports closed-loop numbers from the running application that separate setpoint tracking from disturbance rejection across the three arrangements. The experiments run from a manual step test to hardware running beside its own prediction.

Section 4.2 traces laboratory practice from shared rigs to boards a student keeps. Section 4.3 states every constant needed to rebuild the twin, with its limits, numerics and assumptions. Section 4.4 derives the steady-state gain matrix in closed form, checks it against the running application and fits a first-order plus dead-time model to the step response. Section 4.5 gives the three PID difference equations and three options beside them: a model predictive controller (Section 4.5.2), a moving horizon estimator for it (Section 4.5.3) and a relay test that writes the PID gains (Section 4.5.4). Section 4.6 covers the serial protocol and the safety latch. Section 4.7 measures the three PID arrangements on a setpoint step, an ambient load step and sensor noise. It then compares the predictive controller with PID on the twin, which at the default parameters is that controller's own model, so the comparison is a matched-model best case with no measurement from a physical board (Section 4.7.5), and measures the estimator and the relay on the twin alone (Section 4.7.6). Section 4.8 sets the board's devices beside their industrial counterparts, and Sections 4.9 and 4.10 state where the twin stops holding and what transfers to plant practice.

4.2 Background

Laboratory instruction in control has moved along a path that hardware cost determined at every stage. Physical rigs were expensive, so they were shared, scheduled and supervised, and the response was first to instrument them for remote access and then to replace them with simulations. Comparative reviews of the three arrangements, hands-on, simulated and remote, found that they are not substitutes for one another because they serve different objectives, with the hands-on arrangement retaining the design and diagnosis skills that the other two shed [6]. A bibliometric survey of the field counted the growth of virtual and remote laboratories across two decades and mapped the communities that built them [7], and an earlier review of control learning set out the interactive tools that the discipline had accumulated by then [8].

The economics reversed once microcontrollers became cheap. A portable platform assembled from low-cost components was shown to support hands-on work in control and systems theory at a scale that shared rigs never could, with each student keeping a board [9]. The TCLab belongs to this generation of devices. It is a thermal plant rather than an electromechanical one, with time constants of the order of minutes and no stored mechanical energy, so an unstable tuning is observed rather than repaired. The history box below follows kits of this kind to 2018, and the laboratory's role in engineering education has been argued from the objectives it serves rather than from the equipment it owns [10].

> ⧖ HISTORY **Laboratories Students Take Home, 2013 to 2018**
>
> Laboratory hardware for each student was long out of reach because of its cost. A survey of control education published in 2018 put a single piece of laboratory equipment at 5,000 to 10,000 pounds sterling. Hardware once driven from an expensive input and output board inside a computer could by then run from its USB port through a simple microcontroller board. By 2013 a university had built a take-home kit, a small helicopter on three axes driven by two fans, for under 300 pounds of parts each, enough to lend one to every student in the course. The survey expected such kits to grow significantly [11].

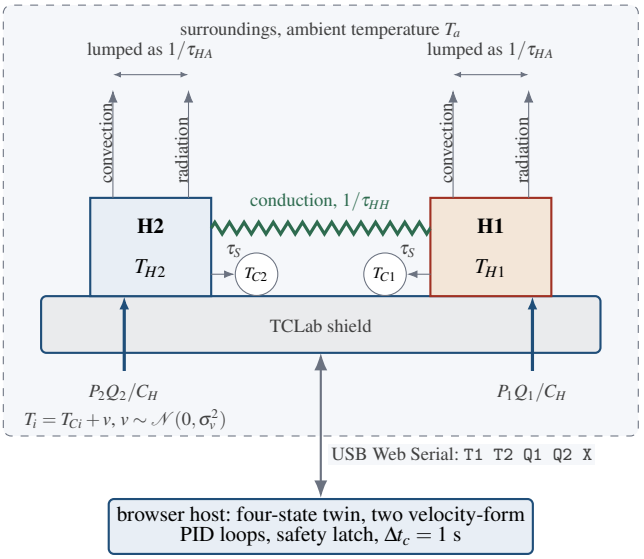


Figure 4.1 Two-heater Temperature Control Lab and the four states of its browser twin. Heater powers Q_1 and Q_2 are commanded from 0 to 100%, the two heater temperatures exchange heat through the board, both lose heat to the surroundings, and each thermistor temperature follows its heater through the lag τ_S.

The browser is the part of this history that is recent. Web Serial gives a page direct access to a USB serial device without a driver, a helper application or an installed runtime, so the same page that carries the simulation can also carry the hardware session. The application of this chapter uses that capability to put the twin and the device under one controller.

4.3 Process Model

The physical TCLab holds two heaters and two sensors on one small board. Heater 1 and its thermistor sit at one end, heater 2 and its thermistor at the other, and the board itself conducts heat between them. Both heaters lose heat to the room by convection from their packages and by radiation from their surfaces. Each thermistor is mounted beside its transistor rather than on it, so the measured temperature follows the heater temperature through a thermal resistance and lags it. Figure 4.1 draws that arrangement and carries every symbol used below.

4.3.1 Signal Classification

The two controlled variables are the reported sensor temperatures. The two manipulated variables are the heater duty cycles, which the firmware accepts as percentages and applies by pulse-width modulation. The single disturbance available in the twin is the ambient temperature, which is also the reset condition. An adjustable room temperature is deliberately artificial, since no room obeys a slider, but it supplies the only load disturbance that leaves both setpoints untouched. The heater temperatures are states that no instrument measures, on the twin or on the board. Table 4.1 lists the classification and the ranges the application enforces.

🕮 **Pause.** Try Exercise 4.2 before reading on.

4.3.2 Energy Balances

Each heater is treated as a lumped mass at a single temperature. Its energy balance, in the form of Equation (1.1), carries three terms, which are the electrical power dissipated in the transistor, the loss to the surroundings and the conduction to the other heater. Equations (4.1) and (4.2) give the pair.

Table 4.1 Signal Classification and Ranges

Symbol	Role	Units	Range
T_1, T_2	controlled, reported	°C	unbounded in the twin
Q_1, Q_2	manipulated	%	0 to 100
T_a	disturbance and reset	°C	0 to 45
T_{H1}, T_{H2}	state, unmeasured	°C	unbounded
T_{C1}, T_{C2}	state, measured with noise	°C	unbounded
SP_1, SP_2	setpoint	°C	20 to the cutoff
K_p	controller gain	%/°C	0 to 50
K_i	controller gain	%/(°C s)	0 to 5
K_d	controller gain	% s/°C	0 to 50

$$\frac{\mathrm{d}T_{H1}}{\mathrm{d}t} = \frac{P_1 Q_1}{C_H} + \frac{T_a - T_{H1}}{\tau_{HA}} - \frac{T_{H1} - T_{H2}}{\tau_{HH}} \tag{4.1}$$

$$\frac{\mathrm{d}T_{H2}}{\mathrm{d}t} = \frac{P_2 Q_2}{C_H} + \frac{T_a - T_{H2}}{\tau_{HA}} + \frac{T_{H1} - T_{H2}}{\tau_{HH}} \tag{4.2}$$

where T_{H1} and T_{H2} are the heater temperatures, Q_1 and Q_2 are the duty cycles in percent, P_1 and P_2 are the dimensionless heater capacities, C_H is a fixed capacity constant, τ_{HA} is the heater-to-ambient time constant and τ_{HH} is the heater-to-heater time constant. The conduction term enters the two balances with opposite signs, so the total energy exchanged between the heaters is conserved. The group $P_1 Q_1 / C_H$ has units of °C/s, which fixes the units of C_H at 5720 % s/°C for a dimensionless capacity.

The published physics-based TCLab model separates a convective loss proportional to $T_a - T_H$ from a radiative loss proportional to $T_a^4 - T_H^4$, and the radiative contribution is what makes the measured steady-state slope fall below unity at elevated temperatures [1]. The twin implemented here folds both mechanisms into the single linear coefficient $1/\tau_{HA}$, which is the form used by the reference simulation shipped with the TCLab software package for offline work. The consequence is that the twin is linear in the states and exactly reproducible, and that it will not reproduce the loss of steady-state gain that a real board shows above roughly 55 °C. The default τ_{HA} of 20 s is therefore best read as a fit over the operating band rather than as a convection coefficient.

4.3.3 Sensor States and the Reported Measurement

Each thermistor is a separate state that relaxes toward its own heater temperature with a single time constant. Equations (4.3) and (4.4) are the two sensor relations.

$$\frac{\mathrm{d}T_{C1}}{\mathrm{d}t} = \frac{T_{H1} - T_{C1}}{\tau_S} \tag{4.3}$$

$$\frac{\mathrm{d}T_{C2}}{\mathrm{d}t} = \frac{T_{H2} - T_{C2}}{\tau_S} \tag{4.4}$$

where τ_S is the sensor time constant, taken as identical for both channels. What the interface reports is the sensor state plus zero-mean Gaussian noise, given by Equation (4.5).

$$T_i = T_{Ci} + \sigma_v z, \qquad z \sim \mathcal{N}(0,1), \qquad i = 1,2 \tag{4.5}$$

where σ_v is the noise standard deviation and z is generated by the Box-Muller transform from two uniform variates. The noise is drawn fresh on every read and is not part of the state, so it does not accumulate. Nothing in the twin models thermistor quantization, self-heating or the calibration offset that separates two physical boards.

Table 4.2 Twin Parameters, Controller Defaults and Limits

Symbol	Meaning	Default	Range, visibility
T_a	ambient temperature	21 °C	0 to 45, adjustable
P_1	heater 1 capacity	200	50 to 255, adjustable
P_2	heater 2 capacity	100	25 to 255, adjustable
τ_{HA}	heater to ambient	20 s	5 to 100, adjustable
τ_{HH}	heater to heater	100 s	20 to 300, adjustable
τ_S	sensor lag	140 s	20 to 300, adjustable
σ_v	sensor noise	0.2 °C	0 to 1, adjustable
C_H	heater capacity constant	5720	fixed, hidden
h	integration step	0.1 s	fixed, hidden
Δt_c	nominal control interval	1 s	fixed
K_p	proportional gain	2	0 to 50, adjustable
K_i	integral gain	0.1	0 to 5, adjustable
K_d	derivative gain	1	0 to 50, adjustable
SP_1	loop 1 setpoint	45 °C	20 to cutoff
SP_2	loop 2 setpoint	40 °C	20 to cutoff
T_{cut}	hardware cutoff	70 °C	50 to 90, adjustable
s	simulation speed	1×	1 to 10, adjustable

✦ **DID YOU KNOW? Wires That a Thermistor Ignores**

A thermistor's resistance runs to tens or even hundreds of thousands of ohms. Two ohms of lead wire are 1.96 % of the circuit of a 100 Ω platinum resistance thermometer but only 0.004 % of a 50 kΩ thermistor, so the thermistor's own resistance swamps its wiring [12].

4.3.4 Parameters, Limits and Initial Conditions

Parameters are listed in Table 4.2. Seven of them are operator adjustable and take effect on the running simulation, which makes the twin a small parameter-estimation problem in its own right. Two are hidden. The capacity constant C_H is compiled into the model and is never displayed, and the integration step is fixed.

A reset sets all four states to T_a and both heater outputs to zero, so the twin always starts isothermal with the room. Heater commands are clamped to the interval 0 to 100% before they reach the model, and the twin has no upper temperature bound of its own. With $P_1 = 200$ and $Q_1 = 100\%$ the twin settles at 80.9 °C, which is above both the default hardware cutoff and the ceiling of a real board, so a heater left at full power in simulation reaches a temperature that the device it models cannot.

4.3.5 Numerical Integration and Timing

All four derivatives are evaluated at the old state and advanced with an explicit Euler step of $h = 0.1\,\mathrm{s}$, as shown in Equation (4.6).

$$x_{n+1} = x_n + h\,f(x_n, Q_1, Q_2), \qquad h = 0.1\,\mathrm{s} \tag{4.6}$$

where $x = [T_{H1}, T_{H2}, T_{C1}, T_{C2}]$. The step is two orders of magnitude below the smallest eigenvalue time constant of the model, so the Euler error is not visible at the resolution of the charts. The simulation speed control from 1× to 10× multiplies the number of 0.1 s substeps taken per timer tick rather than enlarging h, which preserves the integration step at every speed. A speed of 10× compresses a 45 min experiment into about five minutes, which no physical board allows. Unlike the framework apps of Section 1.6.1, where the speed never changes what a run computes, this standalone app does not guarantee identical closed-loop histories when control timing or noise reads differ, and its noise is not seeded.

Two clocks run in the application and they are deliberately different. The physics loop advances simulated time in fixed substeps. The control loop measures wall-clock time between executions, multiplies it by the speed factor and clamps the result to the interval 0.05 s to 3 s, and that clamped value is the Δt used in the integral and derivative terms. The nominal control interval is 1 s of simulated time at any speed. When a physical board is connected the speed factor is forced to $1\times$, because the hardware cannot be accelerated.

4.3.6 Assumptions

The twin rests on six assumptions, each of which breaks down somewhere.

Each heater is isothermal. The transistor package, its leads and the copper around it are treated as one temperature. This fails during the first seconds after a large step, when the junction is far hotter than the package.

The loss to the surroundings is linear in temperature difference. Radiation is not linear, and the error grows with the fourth power of absolute temperature, so the twin overpredicts the steady-state rise at high heater settings [1].

The ambient temperature is uniform and known. On a real board the air near the heaters is warmer than the room, and the ambient reading itself drifts, which is one reported source of run-to-run variation in TCLab benchmarks [1].

The two sensors share one time constant and the two heaters share one loss coefficient. Physical boards differ from one another and the two channels on a single board differ, which is why the twin cannot be expected to overlay a particular device without re-estimating the parameters.

The heater capacities are constant with duty cycle. Pulse-width modulation into a transistor whose gain depends on temperature is not exactly linear in duty cycle.

Measurement noise is white and Gaussian. Thermistor readings quantized by a ten-bit converter are neither, although the resulting difference is small at $\sigma_v = 0.2\,^{\circ}\mathrm{C}$.

4.4 Steady-State Gains and Open-Loop Dynamics

The steady state of Equations (4.1) and (4.2) is a two-by-two linear solve, and the sensor states equal the heater states once the derivatives vanish, so the steady-state gain matrix follows in closed form. The gains are collected in Equation (4.7), evaluated at the default parameters.

$$K = \begin{bmatrix} \partial T_1/\partial Q_1 & \partial T_1/\partial Q_2 \\ \partial T_2/\partial Q_1 & \partial T_2/\partial Q_2 \end{bmatrix} = \begin{bmatrix} 0.599 & 0.050 \\ 0.100 & 0.300 \end{bmatrix} \ ^{\circ}\mathrm{C}/\% \tag{4.7}$$

The diagonal gains differ by a factor of two because the heater capacity $P_2 = 100$ is half of $P_1 = 200$, which reproduces the asymmetry of the physical shield, and the off-diagonal gains inherit the same ratio through the conduction path. The relative gain array of K has $\lambda_{11} = 1.029$, so the diagonal pairing of T_1 with Q_1 is correct and the steady-state interaction is weak [4]. Interaction is nevertheless present in the transient, and Table 4.3 shows how much.

A manual step of Q_1 from 0 to 50% with Q_2 held at zero, taken from the application with the noise disabled and the speed set to $10\times$, gives $T_1 = 50.96\,^{\circ}\mathrm{C}$ and $T_2 = 25.99\,^{\circ}\mathrm{C}$ after 1140 s. The corresponding gains of 0.599 °C/% and 0.100 °C/% agree with Equation (4.7) to three decimal places, which is the check that the running twin implements the equations stated above. A second manual step of Q_2 gives the other column of K.

Table 4.3 Manual Step Tests from the Running Application, $\sigma_v = 0$

Response	K (°C/%)	t_{28} (s)	t_{63} (s)	τ_p, θ_p (s)
T_1 to Q_1	0.599	65	159	141, 18
T_2 to Q_1	0.100	82	177	142.5, 34.5
T_1 to Q_2	0.050	82	177	142.5, 34.5
T_2 to Q_2	0.300	65	159	141, 18

TRY IT IN THE APP 1 · Watch the Lag

Set the sensor noise to 0 in Settings and the speed to 10×. In Manual, type 50 for Q_1 with Q_2 at 0 and watch the two temperature trends. T_1 starts slowly, then climbs, and stands at 50.96 °C about 1140 s after the step. T_2, warmed only through the board, rises to 25.99 °C.

The dynamics are dominated by an instrument, not by the plant. The heater subsystem alone has eigenvalue time constants of 20 s and 14.3 s, obtained from the two-by-two state matrix of Equations (4.1) and (4.2), while each sensor contributes a pole at $\tau_S = 140$ s. A two-point fit of the measured T_1 response, taking $\tau_p = 1.5(t_{63} - t_{28})$ and $\theta_p = t_{63} - \tau_p$ as in Equation (1.18) [4], returns a first-order plus dead-time (FOPDT) approximation whose constants are $\tau_p = 141$ s and $\theta_p = 18$ s, so the identified time constant approximates the thermistor lag and the identified dead time is the heater dynamics collapsed into a delay. The model has no transport delay at all, so this dead time is the effective dead time of Section 1.4.6, the value a method reads from lags in series, as for the two tanks of Chapter 3. Off-diagonal responses fit the same τ_p with roughly twice the apparent dead time, 34.5 s, because heat has to cross the board before it reaches the second sensor. A student who tunes from the step test alone is therefore tuning against the sensor, and speeding up the loop past that lag is not possible with heater power as the only handle. The FOPDT constants are what a tuning correlation consumes, and the ratio $\theta_p/\tau_p = 0.13$ places this loop in the lag-dominant regime where the achievable closed-loop speed is set by the model error rather than by the delay [13].

Pause. Try Exercise 4.1 before reading on.

4.5 Controller Design

Both loops use the velocity form of Section 1.6.2 in code of their own, which keeps an independent integral contribution for anti-windup [14, 15]. The three increments are given by Equations (4.8) to (4.10), with $e_k = \mathrm{SP} - y_k$ and y_k the reported temperature at cycle k.

$$\Delta u_{P,k} = \begin{cases} K_p\,(e_k - e_{k-1}) & \text{Types A and B} \\ -K_p\,(y_k - y_{k-1}) & \text{Type C} \end{cases} \tag{4.8}$$

$$\Delta u_{I,k} = K_i\, e_k\, \Delta t \tag{4.9}$$

$$\Delta u_{D,k} = \begin{cases} \dfrac{K_d\,(e_k - 2e_{k-1} + e_{k-2})}{\Delta t} & \text{Type A} \\ -\dfrac{K_d\,(y_k - 2y_{k-1} + y_{k-2})}{\Delta t} & \text{Types B and C} \end{cases} \tag{4.10}$$

where K_p, K_i and K_d are independent parallel gains rather than a controller gain with reset and rate times, and Δt is the measured control interval. Type A places error on all three terms and is the textbook velocity algorithm. Type B places derivative on the measurement, which is the default in industrial controllers because it removes the derivative spike that a setpoint step would otherwise produce [16]. Type C places both proportional and derivative on the measurement, so only the integral term sees a setpoint change directly. The application defaults to Type B.

The three increments are combined and the demand is clipped to the actuator span, as shown in Equation (4.11).

$$\begin{aligned} s_k &= s_{k-1} + \Delta u_{P,k} + \Delta u_{D,k} \\ c_k &= s_k + I_{k-1} + \Delta u_{I,k} \\ u_k &= \min\big(100,\ \max(0,\ c_k)\big) \\ I_k &= I_{k-1} + \Delta u_{I,k} + (u_k - c_k) \end{aligned} \tag{4.11}$$

where s_k accumulates the proportional and derivative increments without limit, c_k is the heater demand the algorithm computes and I_k carries the integral contribution and absorbs the whole of the clipping through the term $(u_k - c_k)$, which is applied only when $K_i \neq 0$. In the symbols of Equations (1.25) to (1.27), s_k is S_k, c_k plays the part of v_k with no bias and with the present integral increment included, and the back-calculation acts in a single execution, as a tracking time equal to Δt would. Substituting the third line into the fourth gives $c_k = u_{k-1} + \Delta u_{P,k} + \Delta u_{I,k} + \Delta u_{D,k}$, so whenever $K_i \neq 0$ the recursion is algebraically identical to clipping the running sum and every result of this chapter is unchanged by writing it this way. The two forms differ only at $K_i = 0$, where clipping the running sum would turn symmetric measurement noise into a drift away from the limit (Section 1.6.2).

The correction $(u_k - c_k)$ is the entire anti-windup mechanism [14, 17]. It removes from the integral exactly the part of the demand the limit refused, so an increment computed against a limit is absorbed rather than stored and the heater comes off the limit on the first increment that points back into range. There is no separate tracking time constant, no positional reset state and no rate limit on u. The 1 s nominal cycle limits update frequency but does not impose a bound on heater-command slew rate.

Both loops are reverse acting in the ISA sense, meaning that the controller output falls as the measurement rises, which is the required action because the process gain from heater power to temperature is positive. There is no bias term and no manual reset. Transfer into automatic is handled by seeding at the switch, as shown in Equation (4.12), where the framework apps of Section 1.6.2 instead keep the controller executing in Manual.

$$\begin{aligned} s_{k-1} &\leftarrow \mathrm{clip}(u_{\mathrm{manual}}), \qquad I_{k-1} \leftarrow 0 \\ y_{k-1} = y_{k-2} &\leftarrow y \\ e_{k-1} = e_{k-2} &\leftarrow \mathrm{SP} - y \end{aligned} \tag{4.12}$$

Pre-loading two samples of history with the present measurement and the present error leaves only the integral increment in the first automatic cycle, so the transfer adds no proportional or derivative step. The same seeding is applied whenever the PID type is changed, so switching between Types A, B and C mid-run does not bump the output. Four PID controller objects exist at all times, one for each loop of the twin and one for each loop of the device. They share the gains, the setpoints and the type, and they keep independent histories of s, I, y and e.

4.5.1 Equivalence of the Three Forms at a Fixed Setpoint

With the setpoint constant, $e_k - e_{k-1} = -(y_k - y_{k-1})$ and $e_k - 2e_{k-1} + e_{k-2} = -(y_k - 2y_{k-1} + y_{k-2})$, so Equations (4.8) and (4.10) collapse onto a single expression and the three types produce identical increments. The distinction between them exists only while a setpoint is moving. On the cycle at which the setpoint steps by $\Delta\mathrm{SP}$, the immediate change in output is given by Equation (4.13).

$$\Delta u_k = \begin{cases} K_p\,\Delta\mathrm{SP} + \dfrac{K_d\,\Delta\mathrm{SP}}{\Delta t} + K_i e_k \Delta t & \text{Type A} \\ K_p\,\Delta\mathrm{SP} + K_i e_k \Delta t & \text{Type B} \\ K_i e_k \Delta t & \text{Type C} \end{cases} \tag{4.13}$$

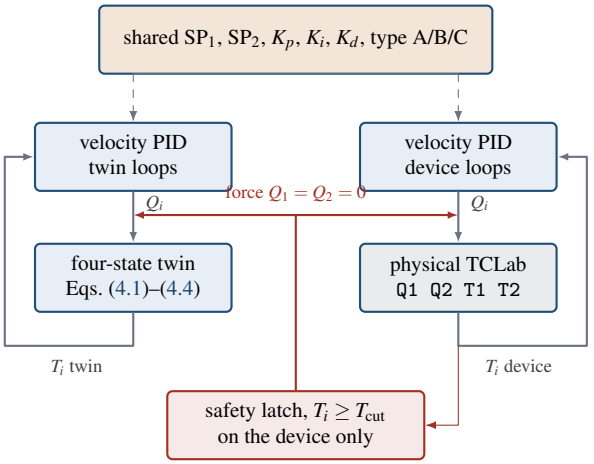


Figure 4.2 One set of gains driving two plants. The twin and the device each keep their own controller history of output, measurement and error, and the latched cutoff is armed by the physical measurements alone.

Type A adds a further $-K_d\,\Delta\mathrm{SP}/\Delta t$ on the following cycle, so the derivative kick is a one-cycle pulse whose net contribution to the output is zero. Type C moves the heater by the integral increment alone, which at the default gains is about 1% of full scale. Figure 4.2 shows the two parallel loops and the point at which the safety latch intervenes.

4.5.2 Model Predictive Control

A model predictive controller is the third entry of the app's mode switch, after Manual and PID auto. It is an option beside the PID loops, and PID auto remains the default closed loop. Where each loop reads one thermistor and moves one heater, the predictive controller moves both heaters from one plan. Every $T_s = 1\,\mathrm{s}$ a quadratic program chooses both duty cycles over $N = 90$ steps and applies the first move, the receding-horizon scheme of Section 1.6.3 [18, 19]. A connected board receives its own instance of the same controller.

The controller's model is Equations (4.1) to (4.4) at the declared defaults of Table 4.2, with $x = [T_{H1}, T_{H2}, T_{C1}, T_{C2}] - T_a$ in deviation from $T_a = 21\,^{\circ}\mathrm{C}$ and $u = [Q_1, Q_2]$. The balances are linear in temperature and power, so every operating point gives the same matrices, and holding u over each step gives Equation (4.14) exactly.

$$x_{k+1} = A_d\,x_k + B_d\,u_k, \qquad \hat{y}_k = C\,x_k + d_k, \qquad C = [0\ \ I_2] \tag{4.14}$$

Its steady-state gains are 0.5994, 0.04995, 0.0999 and 0.2997 °C/%, the matrix K of Equation (4.7) to four figures. No step test was fitted, because the twin's balance is the model. An independent re-derivation of A_d and B_d from the balances matches every entry exactly, and the one residual on the default twin is the 0.1 s Euler step, which departs from the exact hold by at most 0.0045 °C over 900 s at $Q_1 = 50\%$.

The model runs open loop on the duty cycles actually applied, whether the PID loops, the plan, manual operation or the latch set them, and d_k is an output bias filtered by Equation (4.15).

$$d_k = d_{k-1} + \alpha\left(y_k - C\,x_k - d_{k-1}\right), \qquad \alpha = 0.1 \tag{4.15}$$

Here y_k is the pair of reported temperatures less T_a, and the filter's time constant is about 10 s. Held over the horizon, the bias makes the prediction offset free for a constant disturbance such as an

ambient step or a wrong parameter [20]. No preview enters the plan, which minimizes Equation (4.16).

$$J = \sum_{k=1}^{N} q \, \|\hat{y}_k - r\|^2 + S \sum_{m=1}^{12} \|\Delta u_m\|^2 + \sum_{i=1}^{2} \left(\rho_2 \, \sigma_i^2 + \rho_1 \, \sigma_i \right) \tag{4.16}$$

In it r is the setpoint pair less T_a, $q = 1$ per °C^2, and $S = 0.3$ per %^{2} weights the change Δu_m at each of twelve move blocks of $1, 1, 1, 1, 2, 2, 3, 4, 6, 9, 15, 45$ steps, with Δu_1 taken from the duty cycles applied now. The 24 decisions are the block inputs less those applied duty cycles. The constraints are Equation (4.17).

$$\begin{aligned} 0 \le u_{m,i} \le 100\%, \qquad & \sigma_i \ge 0 \\ \hat{y}_{k,i} \le T_{\text{cut}} - T_a - 1\,^\circ\text{C} + \sigma_i & \end{aligned} \tag{4.17}$$

The duty limits are hard, the same limits the PID loops clip to. The ceiling is soft, with one slack per temperature priced at $\rho_2 = 10^5$ per °C^2 and $\rho_1 = 10^4$ per °C so that it outranks any tracking error. It reads the configured cutoff as the latch does and sits 1 °C below it, because a plan held at the cutoff would leave a reading one noise sample from a trip, and the latch remains the protection. S is the largest of seven trial values whose S3 and noise-driven travel stayed within Type B's and whose IAE stayed within 10% of the best, and α the smallest of five whose S4 IAE stayed below Type B's on both loops, a rule set after that sweep had been seen.

A dual active-set method solves each plan [21] and needs a ridge ε on the Hessian diagonal. The shared solver's default, 10^{-9} times one plus the largest diagonal entry, would take the slack's $2\rho_2$ and give $\varepsilon = 2 \times 10^{-4}$, an unintended move weight. The controller passes $\varepsilon = 10^{-9}\,(1 + 2(qNg^2 + S)) = 6.63 \times 10^{-8}$ instead, with $g = 0.5994$ the largest gain. On an unconstrained test, the model lined out at 50 and 35 °C and SP_2 raised by 2 °C, the first move then agrees with a closed-form solution to 3×10^{-7} percentage points on Q_1, a move of 0.081%, while the default ridge shifts it by 9.5×10^{-4} points, 1.2% of the move.

While a plan acts, each PID loop tracks the applied duty, its output set to that value and its histories advanced with an empty integral, so its next execution is its ordinary velocity step. With the noise off at 45 and 40 °C, a switch from MPC to PID auto moved both heaters by 0.0000%, and one from PID auto to MPC moved Q_1 by -0.035%. The first move into the plan is its own decision, not a transfer, so after two minutes of manual operation at $Q_1 = 100$% during the S3 step it was -27.9% against -0.77% under PID. A switch to manual keeps the plan's last request exactly. Whenever the solver reports a status other than optimal, or its first move is undefined or outside 0 to 100%, the plan is discarded, the tracked loops set the heaters for that cycle, the record is marked and a notice posted, and the next valid plan takes over again. A rejection forced 30 s after the S3 step handed over the PID's own step of 0.163%, and the run rejoined the unforced one within 0.0085 °C.

The safety latch of Section 4.6.1 sits above the plan. On a trip both heaters go to zero whatever the plan requested, the interface returns to manual, and no plan is solved while the alarm is latched. The three models of Section 1.4.9 are kept apart. The plant is the twin at whatever parameters are set, the instrument adds the noise of Equation (4.5), and the controller's model is Equation (4.14) at the declared defaults. The controller reads only the reported temperatures, the setpoints, the cutoff and the applied duty cycles, so an edited twin and every physical board are model mismatch, and each of ten deliberately altered copies that reads the twin's true state or its operator-set parameters fails the application's tests. A plan took 326 and 329 µs on average over 1000 plans in two Node measurements, at one-minute load averages of 3.50 and 1.95. In headless Chrome on the built application the controller used 0.19 and 0.20% of one core at 1× and 0.47 and 0.63% at 10×, at load averages of 3.45 and 2.77.

4.5.3 Moving Horizon Estimation

The output bias of Equation (4.15) shifts the prediction but keeps the declared gains and speed, so a weaker heater or a faster loss is planned for with the declared plant. A moving horizon estimator

(MHE) may therefore feed the predictive controller instead, fitting an augmented balance to the latest readings with its parameters held inside bounds [22]. The Estimator switch under MPC selects it. The output bias remains the default and keeps running under either choice. The estimated heater temperatures are states no instrument reaches, the same problem as the brick core of Volume 5, Chapter 11 on a far smaller scale. The augmented balance replaces Equations (4.1) and (4.2) with Equation (4.18).

$$\begin{aligned} \frac{\mathrm{d}T_{H1}}{\mathrm{d}t} &= g_1 \frac{P_1 Q_1}{C_H} + \ell\,(T_a + b - T_{H1}) - \frac{T_{H1} - T_{H2}}{\tau_{HH}} \\ \frac{\mathrm{d}T_{H2}}{\mathrm{d}t} &= g_2 \frac{P_2 Q_2}{C_H} + \ell\,(T_a + b - T_{H2}) + \frac{T_{H1} - T_{H2}}{\tau_{HH}} \end{aligned} \tag{4.18}$$

Here g_1 and g_2 are heater gains as multiples of the declared capacities, ℓ is the loss coefficient $1/\tau_{HA}$ and b is the offset of the room from the declared $T_a = 21\,^{\circ}\mathrm{C}$, while every other constant and Equations (4.3) and (4.4) keep the defaults of Table 4.2. At $g_i = 1$, $\ell = 0.05/\mathrm{s}$ and $b = 0$ the balance is the twin's. The parameters $\theta = [g_1, g_2, \ell, b]$ are random walks joined to the four temperatures less T_a in the state z. The product $\ell\,(T_a + b - T_H)$ is linearized once per step, at the estimate held when the step enters the window, and held exactly over the measured Δt, which gives A_j and c_j, with an error of at most $0.012\,^{\circ}\mathrm{C}$ per step when ℓ and b are off by 0.01/s and $1\,^{\circ}\mathrm{C}$. Every second the estimator solves Equation (4.19) over the last 20 steps, $s = k - 20$.

$$\begin{aligned} &\min \ \|z_s - \bar{z}_s\|^2_{\bar{P}_s^{-1}} + \sum_{j=s}^{k-1} \|w_j\|^2_{W^{-1}} + \sum_{j=s}^{k} \|y_j - C z_j\|^2_{R^{-1}} \\ &\text{s.t. } z_{j+1} = A_j z_j + c_j + G w_j, \qquad \theta_{\min} \le \theta_j \le \theta_{\max} \end{aligned} \tag{4.19}$$

In it C selects T_{C1} and T_{C2} and G drives the two heater states and the four parameters. R is $(0.2\,^{\circ}\mathrm{C})^2$ per channel, the declared σ_v, and W holds standard deviations per step of $0.05\,^{\circ}\mathrm{C}$ on each heater, 0.01 on each gain, 5×10^{-4}/s on ℓ and $0.05\,^{\circ}\mathrm{C}$ on b. The arrival cost carries the previous estimate forward as $\bar{z}_s$ with $\bar{P}_s$ from a Riccati recursion along the stored models. The bounds are the adjustable ranges of Table 4.2, never the values set, so g_1 lies in 0.25 to 1.275, g_2 in 0.25 to 2.55, ℓ in 0.01 to 0.2/s and b in -21 to $24\,^{\circ}\mathrm{C}$. The problem has 128 decisions and 168 bound rows for the solver of the plan [21], and 20 steps was the longest window of 10, 15, 20 and 30 whose estimate and plan stayed under 5 ms at the 99th percentile, a rule fixed before any run.

The plan of Equations (4.16) and (4.17) then starts from the estimated temperatures with Equation (4.18) frozen at $\hat{g}_1$, $\hat{g}_2$, $\hat{\ell}$ and $\hat{b}$, which is linear and is discretized exactly, and with every setting of the default path except the output bias, whose role $\hat{b}$ takes. The controller under this estimator is thus adaptive. Its 90 s prediction matches the twin at the estimated parameters to $1.3 \times 10^{-6}\,^{\circ}\mathrm{C}$. The estimator reads only the reported temperatures, the applied duty cycles and the declared constants, and eleven altered copies that pass the twin's state or the operator's settings to the estimator, the plan or the relay, seven of them by routes no search of the source finds, each fail the application's tests. A rejected window is replaced by a clipped Kalman update and a rejected plan by the tracked loops, and no measured run rejected either. On an unchanged plant a change of estimator moved the first duty by at most 0.019% from the unswitched run, but after a plant change the MHE restarts its parameters at their declared values, and 600 s after P_1 fell to 150 the duties then departed by up to 4.2 and 9.0% over 10 min with the noise off. In two Node.js measurements an estimate took a median of 1.79 and 1.80 ms and a 99th percentile of 2.80 and 2.87 ms, and in headless Chrome at 10× the estimate and plan together reached 7.2 and 7.3 ms at the 99th percentile, about 7% of the 100 ms that one second of twin time lasts, on a loaded computer.

4.5.4 Relay Autotuning

The PID gains can also be measured by the relay experiment of [23], started by the Relay autotune button from wherever the loop stands under PID auto, and never a standing mode. Heater 1 switches

between $u_0 \pm d$, with u_0 the mean heater 1 duty over the last 60 s and $d = \min\big(20,\ 0.9\min(u_0, 100 - u_0)\big)\%$, each time T_1 crosses SP_1 by more than $\varepsilon = 3\sigma_v = 0.6\,^{\circ}\mathrm{C}$, while heater 2 holds its mean over that window and both loops track. Once two cycles of amplitude a agree within 5%, the describing function and the Tyreus-Luyben rule write the editable gains of Equation (4.20), where P_u is the measured period and D the fraction of it spent at $u_0 + d$.

$$K_u = \frac{4d\sin(\pi D)}{\pi\sqrt{a^2 - \varepsilon^2}}, \quad K_p = \frac{K_u}{3.2}, \quad K_i = \frac{K_p}{2.2\,P_u}, \quad K_d = 0 \tag{4.20}$$

The derivative stays at zero because it supplies about 45% of the noise-driven travel (Section 4.7.3). On this loop the experiment measures the band as much as the plant, as the relay on the blending tank of Volume 2, Chapter 3 also does. The chain of lags has no dead time, so the sampled loop reaches -180° only through the 1 s hold, at $K_u = 543.5\,\%/^{\circ}\mathrm{C}$ and $P_u = 17.2\,\mathrm{s}$, while a relay with hysteresis cycles where the phase falls short of -180° by $\psi = \arcsin(\varepsilon/a)$, predicted at 38.06 %/°C and 85.1 s. From the 45 and 40 °C steady state with the noise off it converges in 184 s at 37.97 %/°C and 84.1 s with $\psi = 41.8^{\circ}$, and a band of 0.3 or 1.2 °C moves K_u to 53.41 or 26.20. A change of mode or of SP_1, a trip or a reading 1 °C below the cutoff cancels it with the gains unchanged. Near the cutoff heater 1 then receives u_0, a step up by d from the low level, and in 30 noisy runs at setpoints of 68 to 69 °C 8 readings in 3 runs reached 70 °C afterward, where a board's latch acts as under PID.

4.6 Hardware Interface

Connecting hardware begins with a port request, which the browser grants only in response to a direct operator action. The transport then tries 115 200 baud and falls back to 9600 baud, opening the port, waiting 1500 ms for the board to finish its reset, and writing `VER`. A reply line matching the firmware identity pattern confirms that the device is a TCLab and not some other serial peripheral. On success the transport immediately writes `Q1 0`, `Q2 0` and `X`, so the board is known to be cold-started before any control runs.

The steady-state protocol is four request and reply exchanges per control cycle. The controller writes `T1` and `T2` and parses the two temperatures, computes the two heater commands from them, then writes `Q1 <value>` and `Q2 <value>` with two decimal places and reads back the value the firmware actually applied. Every command waits for one newline-terminated reply with a 1800 ms timeout, and all operations are serialized through a promise chain so that two cycles can never interleave on the wire. A timeout, a closed stream or a removed cable is treated as a disconnect, which zeroes both plants and reports the cause.

The ordering of that exchange is what makes the comparison fair. Reading before computing gives the device loop the same within-cycle measurement the twin loop uses, so neither loop carries a sample of dead time the other does not. The remaining asymmetry is the transport itself, because the four round trips and the firmware's own pulse-width modulation period sit between the measurement and the applied power on the hardware side and are absent on the simulated side. Against a 141 s process time constant the residual is small but is not zero, and it is one of the reasons the two trends do not overlay exactly.

4.6.1 The Safety Latch

The latch is the one part of the application that treats hardware and simulation differently. Its input is the physical measurement only. Whenever either reported device temperature reaches the configured cutoff, which defaults to 70 °C and is adjustable from 50 °C to 90 °C, the alarm latches and five things happen in the same cycle. Both simulated heater outputs are set to zero, the four controllers are reset, the interface is forced back to manual with zero demand, the device is sent `Q1 0`, `Q2 0` and `X`, and the recorded sample is written with zero heater columns.

TCLab firmware 2.1.0 also supplies its own device latch, tripping at 90 °C and permitting release at or below 85 °C. The application honors that status independently of its configurable cutoff and sends

the firmware acknowledgment command `A`. The connection bar also provides a firmware-loading action.

Clearing the latch requires the hazard to have passed. Acknowledgment is accepted only when both device temperatures are strictly below the cutoff, so the alarm cannot be dismissed while the board is still hot, and a board that is already above the cutoff when it is first connected latches immediately rather than accepting commands. Acknowledgment restores the mode the trip interrupted, with both controllers reseeded at zero output, so a loop that was in automatic returns to automatic without a proportional or derivative step. Integral action can change the zero output on the next execution. The cutoff itself cannot be edited while a device is connected or while an alarm is latched, and setpoints are clamped to the cutoff, so no target can be entered that would trip the protection it enforces.

Six further paths lead to the same shutdown, which are USB removal, an unexpected stream close, a command timeout, the page being hidden, the page being unloaded and the stop control. Each of them writes the three shutdown commands on a best-effort basis before releasing the port. The firmware's own command timeout remains the last independent layer, and no browser guarantee substitutes for it.

The simulated plant has no cutoff of its own. A twin driven to 100% heater power settles at 80.9 °C and stays there, which is intentional because the latch protects a transistor and there is no transistor in the twin, but it does mean that a simulation-only session can be run into a region that the hardware session would refuse.

4.7 Results

The experiments of this section are the five defined in Table 4.4. Reported reference trajectories use the shipped physics and controller functions with a fixed 1 s control interval and ten 0.1 s Euler steps between executions. Controller output is held between executions, and its input measurement is the value stored with that output. This protocol removes browser scheduling jitter from comparisons of PID types without changing the application law. It is a controlled timing reference, not a guarantee about an arbitrary browser's control interval. The Type B step is also checked by operating the built application with its browser timers advanced at their scheduled cadence and exporting the recorded data. Its first 890 post-step samples agree with the reference to CSV rounding (0.001 degrees for temperature and 0.01% for heater duty). The browser peak is 60.342 °C.

Closed-loop experiments begin at ambient with zero output and then run for 1800 s at the default setpoints before the disturbance or comparison window. At that instant both noise-free temperatures are within 0.01 °C of their targets. Step and load results use the next 890 s, and noise results use 600 s. IAE sums absolute error over that window, and heater travel sums successive command changes, including the initial setpoint-step change. Noise is zero except in the declared noise experiment. The physical hardware was not attached during these checks.

The application itself has no scenario selector. Its progress panel lists fourteen scenarios that the reader sets up by hand and that it judges from the recorded history, as Section 1.6.4 describes. Scenarios 1 to 3, Watch the lag, See the interaction and Find the asymmetry, hold one heater near 50% in Manual with the other off, for at least 300 s or, in scenario 2, until T_2 has risen at least 1 °C. Scenario 4, Step test, sets the sensor noise to zero and records a manual hold of Q_1 at 20% with Q_2 off, followed by 40%. Each hold lasts at least 600 s and must end settled, with the least-squares trend of T_1 over its last 120 s below 0.1 °C per minute. Scenario 5, Close the loops, runs the default gains and targets under PID for at least 600 s with both final errors below 1 °C. Scenario 6, Compare the PID types, steps the loop 1 target by at least 4.5 °C under each type and follows each step for 60 s. Scenario 7, One tuning, two loops, runs at least 300 s under PID with $P_2 = 100$ and again with $P_2 = 200$. Scenarios 8 to 11 need a board: a connection, 300 s of PID on the board with a target change, a physical T_1 of 60 °C in Manual, and a latched alarm at a cutoff no higher than 55 °C. Scenario 12, Make the interaction worse, lowers the coupling time to 25 s or less and follows a PID target change for 120 s. Scenario 13, PID against MPC, makes the same loop 1 step under each law

Table 4.4 Experiment Definitions

Label	Mode	Action
S1	manual	Q_1 0 to 50%, Q_2 held at 0
S2	manual	Q_2 0 to 50%, Q_1 held at 0
S3	PID A, B, C	settle at SP = 45/40, then SP_1 to 55 °C
S4	PID A, B, C	settle at SP = 45/40, then T_a 21 to 26 °C
S5	PID B	hold 45/40 with $\sigma_v = 0$ and with $\sigma_v = 0.2$ °C

Table 4.5 Scenario S3, SP_1 Stepped from 45 to 55 °C

Metric	Type A	Type B	Type C
First-cycle change in Q_1, %	31.0	21.0	1.0
Peak Q_1, %	89.8	90.0	86.6
Time for T_1 to first reach SP, s	84	84	105
Peak T_1, °C	60.32	60.34	59.86
Overshoot, % of the step	53.2	53.4	48.6
IAE over 890 s, °C s	1564	1570	1638
Q_1 travel, %	189.4	172.1	157.5
Peak T_2 deviation, °C	1.67	1.68	1.57

and follows each for 120 s. Scenario 14, Estimate the plant, tune by relay, runs at least 300 s under the MPC with the MHE after P_1, τ_{HA} or T_a has moved from its default, and follows a converged relay experiment for 120 s under PID. Eight scenarios are required. The four hardware scenarios, 13 and 14 are optional. Some predicates combine present settings with earlier history: scenario 7 does not require a fresh 300 s after P_2 changes, and scenario 12 can reuse an earlier target change. Scenarios 5, 6, 7, 9 and 12 judge automatic mode rather than the control law, so a run under the MPC also completes them, and scenario 6 then credits a Type A step that the plan, not the PID, drove.

4.7.1 Setpoint Step

Scenario S3 raises SP_1 by 10 °C at the default gains $K_p = 2$, $K_i = 0.1$ and $K_d = 1$, with loop 2 holding 40 °C throughout. Results are given in Table 4.5.

The first-cycle changes agree with Equation (4.13). Type A moves 31.0% of full scale, comprising 20% proportional action, a 10% derivative pulse and 1.0% integral action at the fixed 1 s interval. Type B omits the pulse and moves 21.0%, while Type C moves 1.0% through the integral term alone. On the following cycle Type A reverses 9.0% after its derivative pulse. Its total travel is 189.4% against 172.1% for Type B, a difference of 17.3 points. Type C requires 157.5% of travel because it has no proportional move at the step.

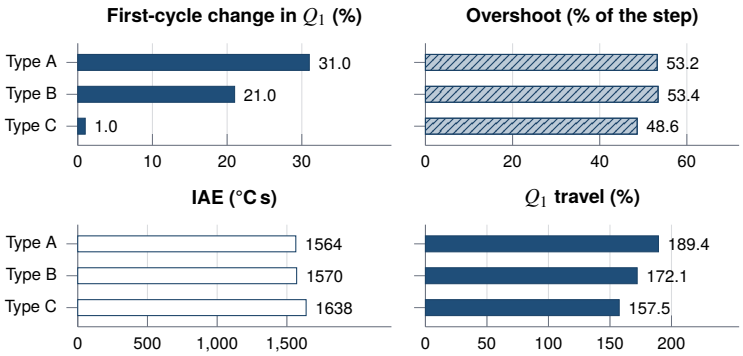


Figure 4.3 Scenario S3 of Table 4.5 under the three velocity forms of PID: the first-cycle heater move and the heater travel (solid bars), the overshoot (hatched) and the integrated absolute error (open). Notice that the first-cycle move falls from 31.0 to 1.0% while the overshoot stays between 48.6 and 53.4% of the step, and the IAE of Type C is only 4.3% above that of Type B.

Table 4.6 Scenario S4, Ambient Stepped from 21 to 26 °C

Metric, loop 1	Type A	Type B	Type C
Peak T_1 deviation, °C	1.178	1.178	1.178
IAE over 890 s, °C s	273.0	273.0	273.0
Q_1 travel, %	29.2	29.2	29.2
Final Q_1, %	28.78	28.78	28.78

TRY IT IN THE APP 6 · Compare the PID Types

With the sensor noise at 0, run PID auto at the default targets until both loops settle. Raise the loop 1 target from 45 to 55 °C under Type A, then repeat under Types B and C. Watch heater 1 on the first cycle after each step: it jumps 31.0, 21.0 and 1.0%, yet T_1 overshoots by about half the step every time (Table 4.5).

Type C first reaches the new target at 105 s, against 84 s for Type B, and increases IAE from 1570 to 1638 °C s, or 4.3%. The peak overshoots span 48.6 to 53.4% of the step, so removing the proportional setpoint response delays the rise without eliminating overshoot. Loop 2 departs from its target by up to 1.68 °C, illustrating transient interaction that the steady-state relative gain array alone does not describe. Figure 4.3 draws the table as bars.

4.7.2 Ambient Load Disturbance

Scenario S4 holds both setpoints fixed and steps the ambient temperature from 21 °C to 26 °C, which is a load change on both loops at once. Results are given in Table 4.6.

The three columns coincide to the last stored digit under this common sampling protocol. Their final heater position of 28.78% approaches the 28.60% steady requirement at $T_a = 26\,°C$ with $T_1 = 45\,°C$ and $T_2 = 40\,°C$. Residual transient error remains at 890 s. Types A, B and C have identical increments at a fixed setpoint and with matching initialized histories. Differences caused by changing type during a run or by different timer histories are initialization or sampling effects, rather than different load rejection laws.

Pause. Try Exercise 4.3 before reading on.

4.7.3 Derivative Action Against Sensor Noise

Scenario S5 holds the default setpoints for 600 s after the 1800 s warm-up. Noise-free heater travel is 0.24% on loop 1 and 0.14% on loop 2. Four declared pseudorandom streams at the default $\sigma_v = 0.2\,°C$

Table 4.7 Scenario S3 Under PID and Model Predictive Control on the Twin

Metric	Type A	Type B	Type C	MPC
IAE T_1, °C s	1563.8	1570.0	1637.9	384.6
IAE T_2, °C s	520.9	522.2	494.8	13.8
Peak T_1, °C	60.32	60.34	59.86	55.08
Overshoot, % of the step	53.2	53.4	48.6	0.8
First reaches SP_1, s	85	85	106	91
Within 0.5 °C for good, s	769	770	671	74
Q_1 travel, %	189.4	172.0	157.5	115.6
Q_2 travel, %	54.9	55.0	52.1	22.7
Peak T_2 deviation, °C	1.67	1.68	1.57	0.20

Noise off, matched model. Under MPC, 2690 plans, none rejected, 2.50 solver iterations on average and 26 at most. Windows end at 2690 s and times count from 1800 s, one sample earlier than in Table 4.5, so Type B reads 172.0 rather than 172.1% of travel and 85 rather than 84 s, while IAE agrees at the precision of Table 4.5.

produce 484 to 525% travel on loop 1 and 485 to 526% on loop 2. The standard deviation of loop 1's recorded error is 0.200 to 0.205 °C, against 0.005 °C in the noise-free reference. Exact noisy histories vary between visits because the browser application uses unseeded randomness. The current CSV records the same noisy value that the controller acted on. Equations (4.8) and (4.10) account for the size of it. With derivative and error taken as in Type B, the noise part of the per-cycle increment is $-(K_p + K_d/\Delta t)\, v_k + (K_p + 2K_d/\Delta t)\, v_{k-1} - (K_d/\Delta t)\, v_{k-2}$, whose standard deviation at the default gains and $\Delta t = 1$ s is $\sigma_v\sqrt{26} = 1.02$ %. The expected movement is $\sqrt{2/\pi}$ times that standard deviation, or 0.81 % per cycle and 488 % over 600 cycles, against the measured 484 to 525%. Setting K_d to zero leaves $\sigma_v\sqrt{8} = 0.57$ % and a predicted 271 %, so the derivative term supplies roughly 45% of the noise-driven travel and the proportional term supplies most of the remainder. The four noise streams with $K_d = 0$ give 272 to 296% travel on loop 1. This confirms the direction and approximate size of the derivative contribution. The application implements no derivative filter, which is the usual industrial remedy [14], so lowering K_d is an available tuning handle. The nominal control interval is fixed by the application, so changing it requires a controller implementation change. The omission is deliberate, because it makes the cost of an unfiltered derivative measurable rather than asserted.

4.7.4 Operating Scenario in the Application

Figure 4.4 shows the current built application during the noise-free Type B step. The default targets are held for 1800 s, then loop 1's target rises from 45 to 55 °C. The capture is taken 300 s after that change. The twin and a connected board share tuning and targets while retaining separate controller histories. Exported samples permit direct comparison of their measurements and applied heater commands.

4.7.5 Comparison With Model Predictive Control

The predictive controller was run against the PID arrangements on the twin under the reference protocol above, with the same program, duration and noise for both, PID at the default gains, the plan at its fixed settings and no retuning for any case. With the default parameters the twin is the controller's own model apart from its Euler step, so every number here is a matched-model best case. A physical board will differ from that model in its gains, its losses and, above about 55 °C, its radiation, and no measurement of the predictive controller on a board is reported. Table 4.7 gives scenario S3.

The plan reaches the new target with 0.8% overshoot against 53.4% for Type B and cuts the loop 1 IAE from 1570.0 to 384.6 °C s with less travel, 115.6 against 172.0% on Q_1. It holds Q_1 at 100% from 1808 to 1829 s and takes it off well before T_1 arrives, because it carries the 140 s thermistor lag that the PID loops meet only as error. The larger difference is on loop 2, as shown in Figure 4.5.

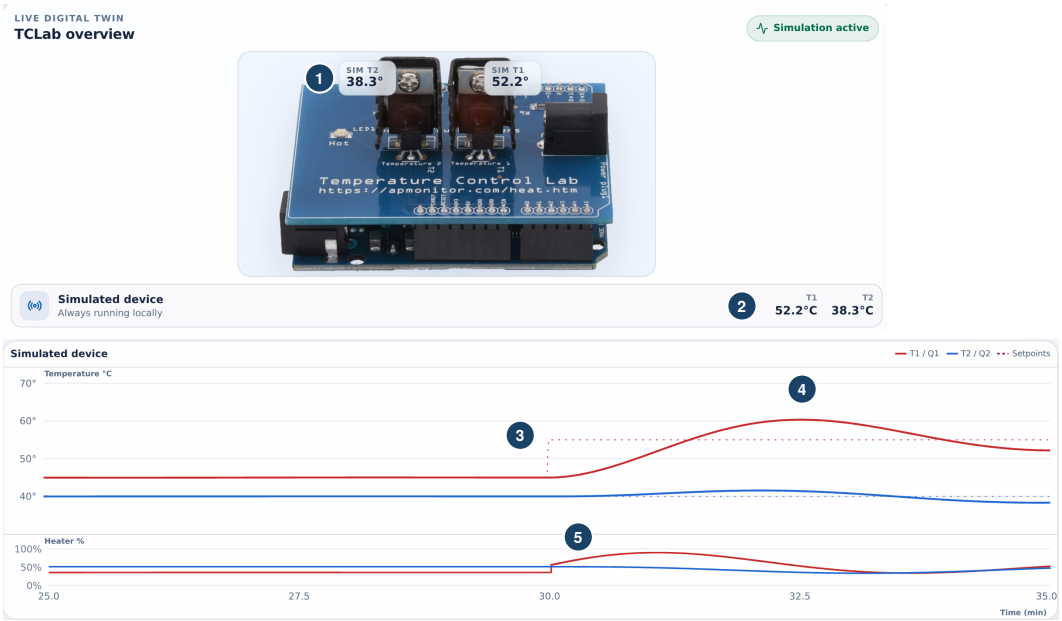


Figure 4.4 Current TCLab Control Studio during the noise-free Type B setpoint step, cropped and composited from one capture: the live twin panel above its trends. Loop 1's target changes from 45 to 55 °C after 1800 s at the default targets, and loop 2 stays at 40 °C. The capture follows 300 s later. Callouts: (1) the board with the twin's readings on its two heaters, (2) the readout row, (3) the loop 1 target step, (4) the overshoot of T_1, (5) the heater trends, with heater 1 stepping up at the target change. The trends draw the response without printing its values. The reference Type B experiment reaches 60.34 °C at its peak and moves heater 1 by 21.0% on the first execution after the target change (Table 4.5).

The plan has lowered Q_2 by 1% at 1802 s, before T_2 moves, because its model carries the conduction from heater 1. Under PID, loop 2 cannot act until that heat reaches its own thermistor, so T_2 is 0.1 °C off at 1829 s and Q_2 is first 1% below its value before the step at 1844 s. The largest T_2 deviation is 1.68 °C at 2093 s under PID against 0.20 °C at 1876 s under the plan, and the loop 2 IAE falls from 522.2 to 13.8 °C s, 2.6% of the PID value, with 22.7 against 55.0% of Q_2 travel. On the reverse step, SP_2 from 40 to 45 °C, T_1 moves 0.42 °C under PID and 0.13 °C under the plan.

Table 4.8 collects the other cases. On the ambient step of S4, which the model keeps at 21 °C, the plan wins by less, because a bias filtered at $\alpha = 0.1$ learns a load only as fast as that filter, and its final Q_1 of 28.60% is the steady requirement. Under the default noise, with four seeded streams of the comparison's own generator rather than those of S5, the plan moved Q_1 28 to 32% in 600 s against 485 to 551% under PID, while the standard deviation of the loop 1 error stayed between 0.196 and 0.213 °C under both. Noise reaches the plan only through the bias filter and the move weight, whereas the unfiltered PID terms act on every sample. With the twin's parameters changed and the model kept, every run under the plan still ends on both setpoints but its lead shrinks. A coupling time of 20 s, within the range of scenario 12, raises its loop 1 IAE from 384.6 to 652.3 °C s, and T_2 then peaks 2.21 °C off against 4.58 °C under PID.

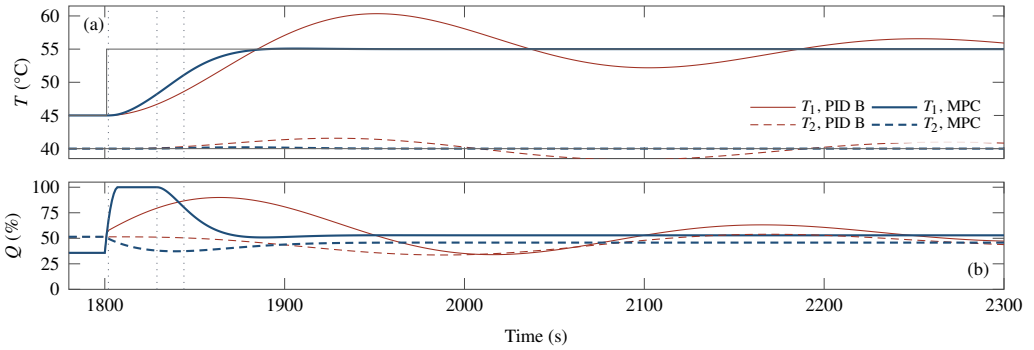


Figure 4.5 Scenario S3 on the twin under PID Type B and under the predictive controller, noise off, with SP_1 stepped from 45 to 55 °C after 1800 s. Thin lines are PID Type B and thick lines the predictive controller. (a) Reported T_1 (solid) and T_2 (dashed) with the setpoints as thin gray lines. (b) Heater duty cycles Q_1 (solid) and Q_2 (dashed). The dotted verticals mark 1802 s, when the plan has lowered Q_2 by 1%, 1829 s, when T_2 under PID is 0.1 °C off its setpoint, and 1844 s, when PID has lowered Q_2 by 1%.

Table 4.8 Further Cases on the Twin, PID Type B Against MPC

Case	Measure	PID B	MPC
S4, T_a 21 to 26 °C	IAE T_1, T_2	273.0, 296.2	202.9, 260.2
	travel Q_1, Q_2, %	29.2, 30.5	7.7, 14.3
SP_2 40 to 45 °C	IAE T_1, T_2	130.6, 814.0	9.1, 245.5
$\sigma_v = 0.2$ °C	travel Q_1, %	485–551	28–32
	travel Q_2, %	475–518	29–37
S3, $\tau_{HH} = 20$ s	IAE T_1, T_2	1301.5, 938.8	652.3, 328.7
S3, $P_2 = 200$	IAE T_1, T_2	1432.7, 558.7	387.4, 36.8
S3, $P_1 = 150$, $\tau_S = 180$ s	IAE T_1, T_2	1485.3, 551.7	760.8, 16.1
SP_1 to 68 °C	peak T_1, IAE T_1	71.45, 2319.2	68.04, 1664.8
SP_1 to 70 °C	peak T_1, IAE T_1	72.94, 2514.9	69.00, 2698.5

IAE in °C s over the 890 s after the change, noise off except the σ_v rows (600 s, four seeds, a range). The last three S3 rows change the twin's parameters while the controller keeps its model. The twin has no latch, so the 70 °C PID row exists only on the twin.

✦ DID YOU KNOW? A Quieter Heater Under the Plan

From the simulator: at the default sensor noise of 0.2 °C, PID Type B moved heater 1 by 485 to 551% in 600 s, while the predictive controller moved it by 28 to 32%, with about the same spread of the loop 1 error under both (Table 4.8).

The cutoff case needs its context. With SP_1 at 70 °C, the highest setpoint the default cutoff admits, PID Type B peaks at 72.94 °C and first reaches 70 °C 171 s after the step, while the plan holds 69.00 °C, its ceiling, with zero slack. The plan's larger IAE, 2698.5 against 2514.9 °C s, exists only because the twin has no latch. On a board with the same cutoff the PID reading at 171 s would trip the latch and end the run in manual at zero duty, leaving no IAE to compare, so the difference is a property of the twin and not a result for the device. With SP_1 at 68 °C the plan peaks at 68.04 °C against 71.45 °C and has the smaller IAE, while PID passes 70 °C 174 s after the step.

Table 4.9 Estimators on Changed Twins, IAE of T_1 and T_2

Law	matched	fan	room +5	$P_1 = 150$
PID B	1612/538	1591/824	1871/829	1631/571
MPC, bias	430/41	1442/721	600/287	907/86
MPC, MHE	469/76	859/157	471/111	660/96
frozen, one offset	437/48	2200/1721	469/117	3054/2807
frozen, two offsets	445/57	1321/267	487/138	745/61
RMS error, bias/MHE	.046/.055	.170/.062	.077/.057	.086/.057

IAE in °C s over 601 to 2400 s, $\sigma_v = 0.2$ °C, mean of three seeds. The frozen rows plan with the MHE's arithmetic and the parameters held at their declared values. RMS error is the estimated against the true thermistor temperature, in °C. No plan or window was rejected.

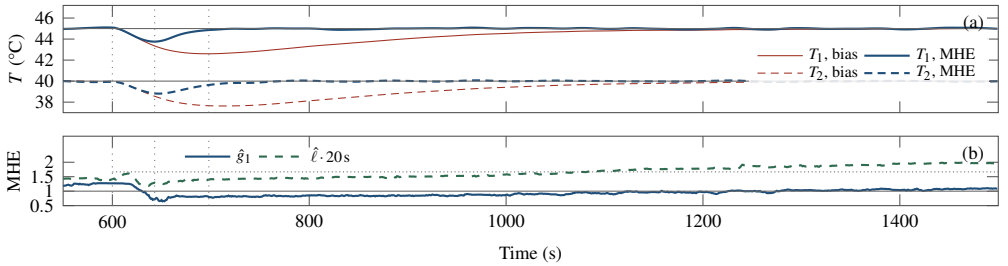


Figure 4.6 Fan case on the twin under the predictive controller with each estimator, first seed, $\sigma_v = 0.2$ °C, with τ_{HA} lowered from 20 to 12 s at 600 s. Thin lines are the output bias and thick lines the MHE. (a) Noise-free T_1 (solid) and T_2 (dashed) with the setpoints as thin gray lines. (b) The MHE's $\hat{g}_1$ (solid) and $\hat{\ell}$ scaled by the declared 20 s (dashed), whose true values are 1 (thin gray) and 1.67 (thin gray dotted). The dotted verticals mark the change and the largest T_1 sag under the MHE (643 s) and under the bias (698 s).

4.7.6 Estimator and Relay Comparisons

The estimators are compared on the twin from the 45 and 40 °C steady state, with a plant change at 600 s, SP_1 stepped to 55 °C at 1500 s and the end at 2400 s. A fan shortens τ_{HA} from 20 to 12 s, the room warms by 5 °C or P_1 falls from 200 to 150, while every controller keeps its declared model, and a fourth case leaves the twin matched. Entries are means over three seeded streams at the default noise, with IAE from 601 to 2400 s on the noise-free thermistor temperatures, which only the measuring script reads. Two frozen rows ask what the parameters earn, running the same window and plan with the parameters at their declared values and either the room offset b or one offset per heater (the offset-free form for two outputs) left free, as listed in Table 4.9.

The parameters earn the fan case, where the MHE cuts the bias path's IAE from 1442 and 721 to 859 and 157 °C s and the frozen forms reach only 2200 and 1721 or 1321 and 267. Most of that gain is in rejecting the change, as shown in Figure 4.6. In the first seed the largest T_1 sag is 1.24 °C at 643 s under the MHE against 2.40 °C at 698 s under the bias, and the T_1 IAE to 1500 s is 103 against 646, while after the step it is 764 against 808 and the T_2 IAE 63 against 37. The loss is taken up first as a weaker heater ($\hat{g}_1$ reaches 0.65 at 651 s), then as a larger $\hat{\ell}$. On the warmer room the frozen form with one offset matches the MHE, 469 and 117 against 471 and 111, so that gain belongs to estimating the room as a room. On the matched twin the wander of the parameters costs 9% on T_1 and 85% on T_2, while the frozen form stays near the bias at 437 and 48. With P_1 lowered the parameters earn T_1, 660 against 745, and cost T_2, 96 against 61.

Under the default noise the estimator does not identify the gain and the loss apart. One operating point fixes each heater's steady effect on its own temperature, which depends on g and ℓ together,

and the posterior correlation of $\hat{g}_1$ with $\hat{\ell}$ is 0.82 to 0.86. On the matched twin with the first seed, $\hat{g}_1$ ends at 1.000, 1.046, 1.147 and 1.265 for σ_v of 0, 0.05, 0.1 and 0.2 °C, and at 0.2 °C it rests on its bound of 1.275 in 14% of the seconds. Every seed ends between 1.25 and 1.27, a bias whose cause is not established, and the wrong pair gives T_1 a steady gain from Q_1 of 0.56 to 0.57 °C/% against 0.599, but T_2 0.070 to 0.073 against 0.100 and a heater time constant of 14.3 to 14.6 s against 20, so T_1 benefits and T_2 pays. With the noise off the two estimators tie on the matched twin at a T_1 IAE of 395, and the MHE wins the fan, room and heater cases at 794, 425 and 593 against 1434, 578 and 895 °C s. Overshoot stayed at or below 0.17 °C under both estimators, and the bias filter holds no parameter, so no case shows a bound that it violates and the MHE respects. Over 13 cases near and away from the cutoff, the MHE path came at most 0.23 °C closer to it than the bias path, and both kept at least 1.00 °C, where PID Type B passed the cutoff in three.

Over 40 noise seeds every relay experiment converged, in 95 to 962 s, at K_u of 23.5 to 43.9 %/°C (median 30.0) and P_u of 60 to 84 s, and wrote K_p of 7.4 to 13.7. On S3 with the noise off, the gains from the noise-free experiment, $K_p = 11.87$ and $K_i = 0.0641$, give IAE of 1126 and 87 °C s against 1570 and 522 for the default tuning, and all 40 noisy gain sets also won the step on T_1 (650 to 989 °C s). The same gains lose the warm-up from ambient, a T_1 IAE over 1800 s of 5002 against 3689 °C s, as did 39 of the 40 others, because the smaller K_i approaches slowly. Under the default noise the written gains move the heaters 2.0 to 3.7 times as far as the default tuning.

4.8 Technology in Practice

The twin of Section 4.3 reads two thermistor states through white noise and moves two heaters the instant they are commanded. Figure 4.7 draws the two loops with the devices of the board.

What Measures It

Each channel reads a thermistor, a resistance element whose high resistance makes the wires to it negligible [12]. The ten-bit converter of the Assumptions turns that resistance into a reading, and Equation (4.5) stands for its noise. A plant loop would read a platinum resistance thermometer or a thermocouple, usually inside a thermowell. A thermowell adds a first-order lag through its own mass and heat capacity, and an air gap between element and well adds a second lag [12]. The thermistor mounted beside its transistor is that arrangement in miniature: the sensor lag τ_S of Equation (4.3) is seven times the heater's own τ_{HA} and sets the fitted time constant of Table 4.3. No range, accuracy or response time is given here, because none was found in an open source that names no maker.

> **TECHNOLOGY** **Sensor: A Thermistor**
>
> A thermistor is a small piece of metal oxide whose resistance rises or falls with temperature. It is highly sensitive and nonlinear, so it serves where high accuracy is unimportant and the range is narrow, as in many consumer devices and in building heating and cooling. Plants with wide temperature ranges use platinum resistance thermometers instead. Reading a resistance means passing a current through it, and that current heats the element and adds a positive error. Pulsing the current and sampling only while it flows reduces the heating, at the price of a slow sample rate, which most temperature measurements can tolerate [12].

What Moves It

Each heater is a transistor driven by pulse-width modulation, with the duty cycle setting the share of full power [12]. An industrial electric heater whose element can only be fully on or fully off is driven the same way, by a PID function block feeding a time-proportioned output, as in a handbook example for an electric oven [12]. The model's heater is linear in duty cycle and has no switching period, which Section 4.6 counts among the delays the twin leaves out.

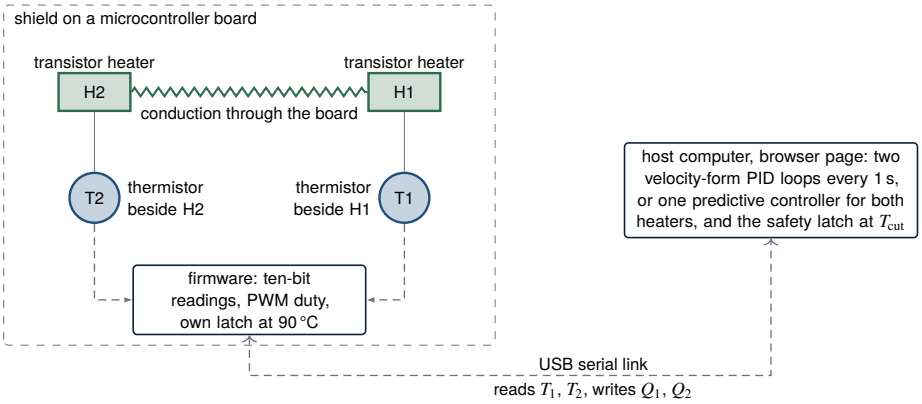


Figure 4.7 The two loops of the chapter on the board itself. Each thermistor sits beside its transistor heater, the firmware converts the readings and applies the duty cycles, and the control law runs on the host behind a serial link. Notice that heat from either heater reaches both thermistors through the board, so the two loops share one plant although each reads one sensor and moves one heater.

> **TECHNOLOGY** **Final Element: A Transistor Switched On and Off**
>
> Electronic power control switches a transistor fully on and fully off in rapid succession, a method called pulse-width modulation. The duty cycle, the share of each period spent on, sets the fraction of full power delivered, and a drive that accepts a control signal acts like any other final control element. In large motor drives the square-edged pulses radiate electrical noise, so power conductors are routed away from signal wiring [12]. On this board the transistor is itself the heater: the firmware turns the commanded percentage into a duty cycle, and the power the transistor dissipates warms the package that the thermistor reads.

Where the Controller Runs

The control law runs in a browser page on the host computer, not on the board. The firmware reads the thermistors, applies the duty cycles and keeps its own latch, and each cycle the host reads both temperatures, computes both commands and writes them over the serial link (Section 4.6). That sequence is the cycle time of a standards glossary: sensor signals read into memory, the algorithm executed and the actuator signals sent [24]. Plant loops run with cycle times from milliseconds to minutes [24], and the 1 s nominal control interval sits inside that range, short against the 141 s fitted time constant. In a plant the velocity-form PID of Equation (4.11) would be the PID function block of a single-loop controller or a distributed control system.

> **IN PRACTICE** **Judge Derivative Action on the Heater Trend**
>
> An instrument handbook sums up the three actions for the person tuning a loop. Proportional action reproduces process noise on the output, derivative action amplifies it dramatically, and integral action is able to ignore it. Derivative action helps most on processes dominated by large lags, where a small amount may allow more aggressive proportional and integral settings. The open-loop tuning rules overlook noise, and a small apparent dead time makes them predict gains too large to be practical. The handbook's advice is that no algorithmic tuning method replaces the informed judgment of the person tuning the loop [12].

What Is Modernizing It

Three developments reach this loop. The first is the learned model with physics built in: a neural network constrained to keep the rising response of temperature to heater power was trained on 250 samples from a two-heater kit, with heater inputs between 10 and 50 %, and then served a predictive

controller allowed up to 65 %, beyond its training range [25]. The second is the digital twin in the process industry, which joins sensor data to physics-based or data-based models. About 40 % of the 31 studies in one review combined the two kinds of model [26]. The third is the learning-based controller tested on a twin before a board, the subject of the open challenge at the end of Section 4.10.

4.9 Limitations

The loss to the surroundings is linear, so the twin will not reproduce the falling steady-state gain that a physical board shows above roughly 55 °C, and a parameter set fitted at 40 °C will overpredict at 65 °C.

The twin is not calibrated to any particular device. Board-to-board variation in TCLab hardware is large enough that benchmark comparisons are recommended on one device under similar ambient conditions [1], and the application's estimator does not separate heater gain from heat loss under the default noise (Section 4.7.6), so matching a specific board remains manual work. Chapter 11 fits nested energy balances, one of them with a radiative term, to recorded TCLab runs.

The physical loop reads its temperatures at the start of the same cycle in which it acts, so it carries no extra sample of dead time, but the serial round trips and the firmware pulse-width modulation period separate the measurement from the applied power in a way the simulated loop does not reproduce.

The PID controller has no derivative filter, no bias term, no feedforward path and no rate limit on the heater command, so the only PID structures available are the three velocity forms, beside manual operation and the one predictive controller.

The predictive controller's model is the twin's balance at the declared defaults, so it leaves out radiation and the fall in gain above about 55 °C, every parameter an operator sets and every difference between a particular board and the defaults. With the default estimator its output bias absorbs a constant difference slowly, at $\alpha = 0.1$, and nothing in it corrects a wrong gain. Its horizon and weights are fixed, and so is its model unless the MHE is chosen. Every result for it here is from the twin, matched or deliberately changed, none is from a board, and its soft ceiling keeps a plan below the cutoff only when the model is right, so on a board the latch remains the protection.

No board comparison of the estimator or the relay was run. Under the default noise the estimator does not separate heater gain from loss, its pair underrates the coupling, and τ_{HH} and τ_S stay declared. The relay's band sets its K_u as much as the loop does.

The safety latch is armed by the physical measurements alone, so the twin and the device do not execute the same logic once the cutoff is reached. That asymmetry is intended, because the latch protects a transistor and the twin has none, but it is a caveat on the claim that one set of code drives both plants. A simulation-only session can be driven to 80.9 °C, and no browser mechanism is a substitute for the firmware's own command timeout or for watching the board.

Web Serial is available only in Chromium-family browsers on desktop operating systems and only from a secure context, so the hardware half of the application is unavailable on iOS and on pages served over plain HTTP. The simulator remains fully functional in that case, which is the reason the two halves were built to run independently.

4.10 Discussion

Temperature loops of exactly this shape, a slow first-order plant behind a lagged sensor with a saturating actuator, regulate reboilers, extruder barrels, bioreactor jackets and battery packs, and the steam-heated exchanger of Volume 2, Chapter 4 is tuned from step tests in the same way. A two-transistor board is worth studying because the structure transfers while the cost does not. A department can issue several hundred TCLab kits for the price of one pilot-scale rig, and each student then owns the apparatus for the semester rather than sharing it for three hours. Reported experience with the device in online and remote sections indicates that this arrangement survives the loss of a physical laboratory [2].

Industrial digital twins are justified by the tests that are too expensive or too dangerous to run on the plant. The same argument holds at the scale of a classroom board, where running the loop unstable, forcing the actuator onto its limit and tripping the protection are all instructive and all cheaper in simulation. Placing the twin beside the device rather than in front of it makes the comparison the object of study, which is the habit that transfers to a plant where the model is the only thing available before startup.

OPEN CHALLENGE **Trusting a Learned Controller With Real Heaters**

In one study, controllers built on large language models set the power of two heaters on kits of this kind while a fan cooled one. Trials on the physical boards met varying initial conditions, varying response times and poor reproducibility under changing room conditions, so the study built a digital twin for repeatable tests. Use in a plant would need deterministic fallback policies, formal verification of the form of each output and the standards IEC 61511 and NIST SP 800 for functional safety and cybersecurity [27]. Progress would be a learned controller whose behavior on the board is bounded before it is given the heaters.

4.11 Conclusion

This chapter studied an application that runs a four-state Temperature Control Lab simulation and a physical TCLab from one set of controller code, with the two responses plotted on a common time axis. The twin integrates two heater temperatures and two thermistor temperatures with explicit Euler steps of 0.1 s, and its steady-state gain matrix, measured from the running application as 0.599 and 0.050 °C/% on the first row and 0.100 and 0.300 °C/% on the second, agrees with the closed-form solution of the energy balances. A two-point fit of the same response returns $\tau_p = 141$ s against $\theta_p = 18$ s, so the identified time constant belongs to the thermistor rather than to the heater.

Three velocity forms of PID are implemented and measured, and a 10 °C setpoint step separates them by a first-cycle heater movement of 31.0%, 21.0% and 1.0%, while an ambient load step of 5 °C leaves them indistinguishable because the three difference equations coincide when the setpoint is constant. Sensor noise of 0.2 °C gives 484 to 525% heater travel over 600 cycles after settling, compared with 0.24% without noise, and the derivative term accounts for roughly 45% of that travel, which is the case for the derivative filter that the application deliberately omits.

A linear predictive controller on the twin's own balance, offered beside the PID loops, holds loop 2 within 0.20 °C of its setpoint on the 10 °C step where Type B departs 1.68 °C, a matched-model result on the twin with no device measurement behind it. On a moving horizon estimate it cuts the T_1 IAE after a faster heat loss from 1442 to 859 °C s, with gain and loss unidentified under noise, and relay-tuned PI gains win the step and lose the warm-up.

A fuller study would extend the moving horizon estimator to the coupling and sensor time constants, and add a filtered derivative, feedforward from the second heater and a comparison across several physical boards. What the application offers throughout is the same controller on a model and on the device that model was built for.

KEY POINTS

- The board answers a heater step slowly because each sensor lags the heater beside it.
- A two-point fit returns $\tau_p = 141$ s and $\theta_p = 18$ s, so the time constant belongs to the thermistor.
- The three velocity forms of PID differ only while the setpoint moves: a 10 °C step moves the heater 31.0, 21.0 and 1.0% on the first cycle, and an ambient step leaves them identical.
- Sensor noise of 0.2 °C drives 484 to 525% of heater travel over 600 cycles, and the unfiltered derivative supplies roughly 45% of it.

Where else this applies. Volume 14, Chapter 5 couples one hundred heater zones on one plate through the same conduction, and Volume 2, Chapter 4 tunes a plant thermal loop from step tests.

References

[1] 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.

[2] B. Yu, "Teach online controls laboratory using a low-cost temperature control lab hardware," in *2022 ASEE Annual Conference & Exposition Proceedings*. ASEE Conferences, 2022.

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

[4] 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.

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

[6] J. Ma and J. V. Nickerson, "Hands-on, simulated, and remote laboratories: A comparative literature review," *ACM Computing Surveys*, vol. 38, no. 3, p. 7, 2006.

[7] R. Heradio, L. de la Torre, D. Galan, F. J. Cabrerizo, E. Herrera-Viedma, and S. Dormido, "Virtual and remote labs in education: A bibliometric analysis," *Computers & Education*, vol. 98, pp. 14–38, 2016.

[8] S. Dormido Bencomo, "Control learning: present and future," *Annual Reviews in Control*, vol. 28, no. 1, pp. 115–136, 2004.

[9] B. Taylor, P. Eastwood, and B. L. Jones, "Development of a low-cost, portable hardware platform to support hands-on learning in the teaching of control and systems theory," *Engineering Education*, vol. 9, no. 1, pp. 62–73, 2014.

[10] L. D. Feisel and A. J. Rosa, "The role of the laboratory in undergraduate engineering education," *Journal of Engineering Education*, vol. 94, no. 1, pp. 121–130, 2005.

[11] J. A. Rossiter, B. Pasik-Duncan, S. Dormido, L. Vlacic, B. Jones, and R. Murray, "A survey of good practice in control education," *European Journal of Engineering Education*, vol. 43, no. 6, pp. 801–823, 2018.

[12] 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.

[13] S. Skogestad, "Simple analytic rules for model reduction and PID controller tuning," *Journal of Process Control*, vol. 13, no. 4, pp. 291–309, 2003.

[14] K. J. Åström and T. Hägglund, *Advanced PID Control*. Research Triangle Park, NC: ISA – The Instrumentation, Systems, and Automation Society, 2006.

[15] A. Visioli, *Practical PID Control*, ser. Advances in Industrial Control. Springer London, 2006.

[16] C. C. Hang, K. J. Åström, and W. K. Ho, "Refinements of the Ziegler-Nichols tuning formula," *IEE Proceedings D Control Theory and Applications*, vol. 138, no. 2, pp. 111–118, 1991.

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

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

[21] 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.

[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] 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.

[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, Gaithersburg, MD, Special Publication 800-82r3, 2023.

[25] N. T. Nguyen and J. C. Tique, "Physics-constrained Taylor neural networks for learning and control of dynamical systems," arXiv preprint 2410.02258, 2024.

[26] M. Mayr, G. C. Chasparis, and J. Küng, "Learning paradigms and modelling methodologies for digital twins in process industry," arXiv preprint 2407.02275, 2024.

[27] J. Vyas and M. Mercangoz, "Autonomous control leveraging LLMs: An agentic framework for next-generation industrial automation," arXiv preprint 2507.07115, 2025.

4.12 Exercises

4.1 Concept check · 2 min

A two-point fit of the twin's response of T_1 to a step in Q_1 returns $\tau_p = 141\,\mathrm{s}$ and $\theta_p = 18\,\mathrm{s}$. Which element of the twin does the fitted time constant mostly describe?

(a) The heater-to-ambient loss, since it sets how fast heater 1 warms

(b) The conduction between the heaters, since it is the longest path on the board

(c) The thermistor lag, since the measurement is the slowest state in the loop

(d) The control interval, since the fit reads one sample per second

4.2 Concept check · 2 min

Which of these signals of the twin is a state that no instrument measures, on the twin or on a physical board?

(a) The reported temperature T_1 that loop 1 controls

(b) The heater duty cycle Q_1 that loop 1 manipulates

(c) The ambient temperature T_a that sets the reset condition

(d) The heater temperature T_{H1} that drives the thermistor

4.3 Concept check · 2 min

Scenario S4 steps the ambient temperature from 21 to 26 °C with both setpoints fixed. Table 4.6 gives identical results under PID Types A, B and C. Why?

(a) With a constant setpoint the error increments are the negative measurement increments, so the three laws coincide

(b) The derivative term is zero during a load change, so only the proportional and integral terms act

(c) The anti-windup correction absorbs the difference between the three forms at every cycle

(d) A change of ambient reaches the controller only through the integral term, which all three share

4.4 Theory and calculation · 7 min

Take the twin at its defaults: $P_1 = 200$, $P_2 = 100$, $C_H = 5720\,\%\,\mathrm{s/°C}$, $\tau_{HA} = 20\,\mathrm{s}$ and $\tau_{HH} = 100\,\mathrm{s}$. Write $a = 1/\tau_{HA}$ and $b = 1/\tau_{HH}$.

(a) Set the derivatives of Equations (4.1) and (4.2) to zero and solve for the four entries of the steady-state gain matrix K. *Checkpoint:* $K_{11} = 0.599$, $K_{12} = 0.0500$, $K_{21} = 0.0999$ and $K_{22} = 0.300$, all in °C/%.

(b) Compute the relative gain λ_{11} and say which pairing it recommends. *Checkpoint:* $\lambda_{11} = 1.029$, the diagonal pairing.

(c) Find the duty cycles that hold $T_1 = 45\,°\mathrm{C}$ and $T_2 = 40\,°\mathrm{C}$ at $T_a = 21\,°\mathrm{C}$, and again at $T_a = 26\,°\mathrm{C}$. *Checkpoint:* $Q_1 = 35.75$ and $Q_2 = 51.48\,\%$, then $Q_1 = 28.60$ and $Q_2 = 37.18\,\%$.

4.5 Theory and calculation · 6 min

The step tests of Table 4.3 start at $t = 0$ with the noise off.

(a) Find the two time constants of the heater subsystem alone, from the state matrix of Equations (4.1) and (4.2) with $a = 0.05/\mathrm{s}$ and $b = 0.01/\mathrm{s}$. *Checkpoint:* 20 s and 14.3 s.

(b) Apply the two-point rule of Equation (1.18) to the direct response ($t_{28} = 65\,\mathrm{s}$, $t_{63} = 159\,\mathrm{s}$) and to the cross response ($t_{28} = 82\,\mathrm{s}$, $t_{63} = 177\,\mathrm{s}$), and give θ_p/τ_p for the direct one. *Checkpoint:* $\tau_p = 141\,\mathrm{s}$, $\theta_p = 18\,\mathrm{s}$ direct, $\tau_p = 142.5\,\mathrm{s}$, $\theta_p = 34.5\,\mathrm{s}$ across, ratio 0.13.

Explain where the apparent dead time comes from, given that the model has no transport delay.

4.6 Theory and calculation · 7 min

Loop 1 runs Type B at the default gains $K_p = 2$, $K_i = 0.1$ and $K_d = 1$ with $\Delta t = 1\,\mathrm{s}$, held at a fixed setpoint. The reported temperature carries white Gaussian noise v_k with $\sigma_v = 0.2\,^\circ\mathrm{C}$.

(a) From Equations (4.8) and (4.10), write the noise part of the per-cycle heater increment as a combination of v_k, v_{k-1} and v_{k-2}, and find its standard deviation. *Checkpoint:* Coefficients -3, $+4$ and -1, standard deviation 1.02 %.

(b) The mean absolute value of a zero-mean Gaussian variable is $\sqrt{2/\pi}$ times its standard deviation. Predict the heater travel over 600 cycles. *Checkpoint:* 0.81 % per cycle, about 488 %.

(c) Repeat with $K_d = 0$, and find the share of the travel that the derivative term supplies. *Checkpoint:* 0.57 % per cycle standard deviation, 271 %, a share of about 45 %.

4.7 App challenge · 7 min

Use scenario 1, Watch the lag. In Settings set the Sensor noise to 0 and apply it, press Reset, and set the Sim speed to 10×. Keep the app's tab in front during the run. In Manual, type 50 in the Q_1 field with Q_2 at 0. Let the session clock run about 20 min (two minutes of real time) and export the CSV file. Its time column is in minutes. Take the step time as the first row in which `sim_q1_pct` reads 50.00.

(a) Read `sim_t1_c` and `sim_t2_c` 1140 s after the step, and compute the gains of T_1 and T_2 to Q_1. *Checkpoint:* 50.96 °C and 25.99 °C, each within 0.03 °C. Gains 0.599 and 0.100 °C/%.

(b) Using the final rise of $0.599 \times 50 = 29.97\,^\circ\mathrm{C}$, find the times after the step at which T_1 first passes 28.3 and 63.2% of it, and fit τ_p and θ_p. *Checkpoint:* $t_{28} = 65\,\mathrm{s}$ and $t_{63} = 159\,\mathrm{s}$, each within 3 s. $\tau_p = 141\,\mathrm{s}$, $\theta_p = 18\,\mathrm{s}$.

Say why T_2 rises at all, and why its rise is so much smaller.

4.8 App challenge · 9 min

Use scenario 6, Compare the PID types, with the default gains, Type B and setpoints of 45 and 40 °C. In Settings set the Sensor noise to 0, press Reset, set the Sim speed to 10× and select PID auto. Run until the session clock passes 30 min. Then raise the loop 1 setpoint to 55 °C (type 55 in the field and press Enter, which commits it), run 15 min more and export the CSV file. The step row is the first row in which `setpoint_t1_c` reads 55.00.

(a) Find the change in `sim_q1_pct` from the row before the step row to the step row, and predict from Equation (4.13) what Types A and C would give. *Checkpoint:* 21.0 %, within 0.3 %. Type A 31.0 % and Type C 1.0 %.

(b) Find the peak of T_1, its time after the step row, and the overshoot as a percentage of the step. *Checkpoint:* 60.34 °C within 0.05 °C, about 150 s after the step, 53.4 %.

(c) Find the largest departure of T_2 from 40 °C after the step, and when it occurs. *Checkpoint:* 1.68 °C within 0.03 °C, about 290 s after the step.

Reconcile (c) with $\lambda_{11} = 1.029$.

4.9 App challenge · 9 min

Use scenario 13, PID against MPC. Repeat the run of the previous exercise with the mode switch at MPC instead of PID auto and the Estimator at its default, Bias: noise 0, Reset, 10×, 30 min, then the loop 1 setpoint from 45 to 55 °C with Enter, 15 min more, and export the CSV file. The step row is the first row in which `setpoint_t1_c` reads 55.00.

(a) Find the peak of T_1 and the overshoot. *Checkpoint:* 55.08 °C within 0.05 °C, 0.8 %.

(b) Find the largest departure of T_2 from 40 °C after the step. *Checkpoint:* 0.20 °C within 0.03 °C, about 75 s after the step.

(c) Find how many seconds after the step row `sim_q2_pct` first stands 1 % below its value in the row before the step. *Checkpoint:* About 1 s, against about 43 s under PID Type B.

Explain why the plan lowers Q_2 before T_2 moves, and why this run says little about a physical board.

4.10 Application area · 3 min

An engineer has fitted the twin's parameters to a board at about 40 °C and now wants to hold T_1 at 65 °C with Q_2 off.

(a) What steady Q_1 does the default twin predict for 65 °C at $T_a = 21\,^{\circ}\mathrm{C}$? *Checkpoint:* 73.4 %.

(b) Will the board, held at that duty, settle above or below 65 °C, and how could the bend be located? *Checkpoint:* Below.

4.11 Application area · 3 min

The predictive controller keeps both predicted temperatures under a soft ceiling 1 °C below the hardware cutoff, and the latch trips on any single board reading at or above the cutoff.

(a) With Gaussian noise of $\sigma_v = 0.2\,^{\circ}\mathrm{C}$, about what fraction of readings reach the cutoff when the true temperature sits on the cutoff, and when it sits 1 °C below it? *Checkpoint:* One half, and about 3×10^{-7}.

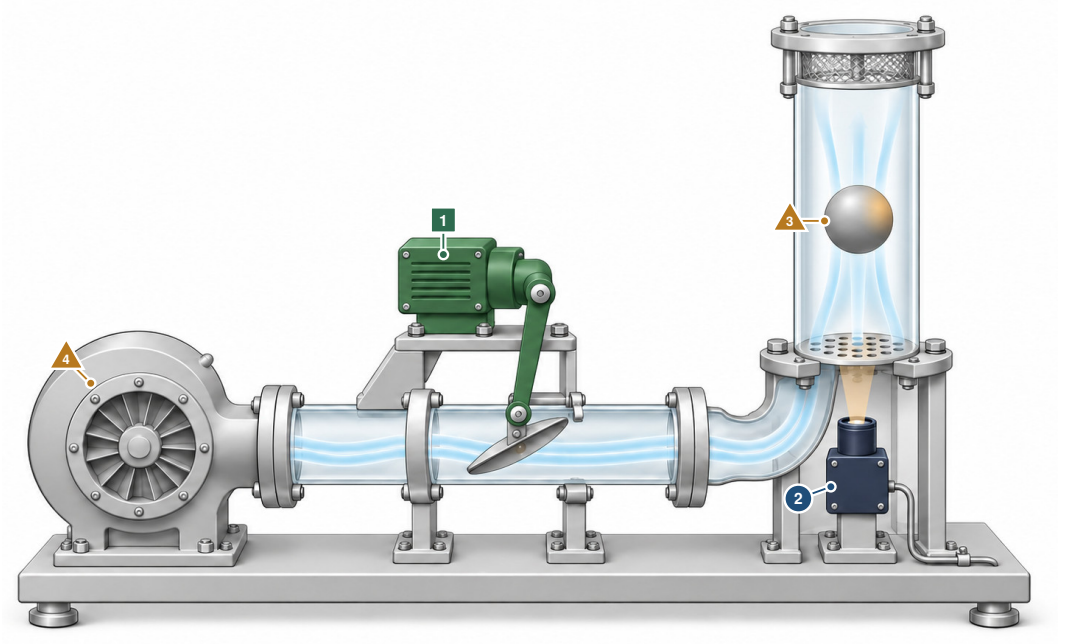
(b) Why may the alarm be acknowledged only once both board temperatures are below the cutoff, and why does the twin have no latch? *Checkpoint:* The hazard must have passed. The twin has no transistor.

4.12 Review the agent · 5 min

An AI assistant reads a relay experiment on loop 1 of the twin, run with the noise off from the 45 and 40 °C steady state, which returned $K_u = 37.97\,\%/^{\circ}\mathrm{C}$ and $P_u = 84.1\,\mathrm{s}$. It writes: "The ultimate gain of loop 1 is 38 %/°C, so a proportional controller above that gain will oscillate without end. The default $K_p = 2$ has a gain margin of 19, and the Tyreus-Luyben gains $K_p = 11.87$ and $K_i = 0.0641$ use a third of the available margin." Find the flaw and say how it could be checked.

CHAPTER 5

Air-Levitated Ball in a Tube Whose Height No Valve Angle Holds



The controller moves 1 the flap, turned by a servo, that sets how much of the blower's air reaches the ball

The controller reads 2 the light ranger below the seat that reads the ball's height

What disturbs it 3 a ball heavier or lighter than declared 4 a brief sag in the blower's supply

Notice that the blower runs at full power and the flap, not the blower's speed, sets the air that lifts the ball, which on the rig nearly fills a tube far longer and narrower than drawn. Illustration, not to scale.

IN PLAIN TERMS A Ball That Floats on a Stream of Air

The main idea. A light ball floats inside a clear upright tube on air blown from below. A controller reads the ball's height and turns a small flap that lets more or less air through.

The challenge. The air moves as fast at the top of the tube as at the bottom, so no fixed flap setting holds the ball still. Left alone, the ball drifts up to the top or down to the bottom, so the controller must keep correcting.

Why it matters. Tank levels and gas pipelines in plants drift in the same way, and the same lessons about valves and sensors apply to them.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the fact on a rig's cost (Section 5.3), the history of a tube built for remote students (Section 5.2), the try-it on hovering by hand, the fact on a slower sensor and the try-it on the valve with no series restriction (Section 5.6) 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 results (Section 5.6), Technology in Practice (Section 5.8), Limitations (Section 5.9) and the open challenge

The control question A ball that nearly fills a vertical tube floats on air throttled by a servo flap, and because the air is as fast at the top of the tube as at the bottom, its height integrates any error in the valve. What do the momentum balance, a drop test and two step tests say about that integrator, and how does an orifice in series with the valve change what the loop can do?

Learning objectives

- Derive the hover speed, the ball's lag and the integrating gain from the momentum balance and the static map of the valve.
- Compute the pressure split and the hover angle of the flap with and without a series restriction, and explain why the restriction widens the usable band.
- Identify the integrating gain from a drop test and step tests in both directions, and tell a chord of the valve curve from its tangent.
- Predict the offset of a proportional loop under a load and the tuning that the dead time of a slower sensor requires.
- Separate a tuning fault from an equipment fault such as a servo dead band.

Prerequisites Sections 1.4.2 and 1.4.3 (balances and linearization), 1.4.4 (integrating processes), 1.4.6 (dead time and the half rule) and 1.4.8 (step tests and least squares). Sections 1.6.2 and 1.6.4 are needed for the controller and the cases, and Chapter 4 for a twin run beside its device.

The app https://apmonitor.com/apps/ball-levitate/

See also Volume 2, Chapter 5 for tuning an integrating loop from an identified slope, Volume 2, Chapter 6 for a gain that moves with the operating point and Volume 2, Chapter 4 for a sticking valve. Volume 4, Chapter 3 for a rate limit on an actuator command and integrators that track it at a limit.

5.1 Introduction

A ball floating on a column of air is one of the few control plants whose controlled variable can be read from across a lecture hall, and laboratories have built it from a beach ball over a blower, from a light ball in a fan-driven tube and as a shield for a microcontroller board [1–3]. The version simulated here departs from them in two ways. A 22.9 mm styrofoam ball rides in a 25.4 mm bore, a diameter ratio of 0.902, so it behaves less like a body in a stream than like a leaky piston, floating on 0.38 L/s squeezed through a 1.25 mm gap on the nominal table. And the blower runs at full power while an SG90 hobby servo throttles it with a flap, so the actuator is a valve with a seat, a curve and a dead band rather than a fan's spin-up.

The plant belongs to the class of integrating processes (Section 1.4.4). Linearized about hover, the ball's velocity follows the air's with a lag of 38 ms and its height integrates that velocity, so a valve 0.5° away from the hover angle drifts the ball at 8.4 mm/s and carries it from 200 mm to the retainer in 20 s. The level of a tank with a pump on its outlet, the pressure in a gas header and the inventory in a surge drum have the same shape, and Chapter 2 meets the same integrator when it closes the outlet valve of its tank. Integrator-plus-dead-time plants have tuning rules written for them [4, 5]. The misconception the rig corrects is that a valve set exactly right will hold the ball. Switched to Manual at the controller's last command, the twin's ball reaches its seat or its retainer within 60 s in 323 of 364 trials (case 1 below), because a command that jitters with the sensor's noise is never left at the hover angle and nothing in a uniform tube pulls a drifting ball back.

The rig also teaches a lesson of valve sizing. On the nominal table a blower that makes 295 Pa at the hover flow against a ball that needs 7.15 Pa leaves the flap to drop 97.5 % of the pressure, and the flap's whole usable band is then 1.19° wide. A fixed orifice in series takes 71 % of the pressure instead, which moves the hover angle to 23.96° and widens the usable band to 22.34°. This is the valve-authority argument of process plants, measured on a desktop. Stiction and backlash in valves are a familiar cause of oscillating loops, and data-driven models of them have been proposed [6], so the servo's dead band joins the lesson as its own case. Volume 2, Chapter 4 includes a sticking steam valve among its cases.

Each surveyed air-levitation rig whose actuator is described drives the speed or voltage of its fan or blower, none floats a ball as close to its bore, and none runs a simulation and its hardware under one controller in one page. This chapter works with a twin of a close-fitting ball behind a throttling valve. It measures on the twin the integrating behavior, the valve's usable band with and without a series restriction, and the limits that the sensor's sample period and the servo's dead band set, and it states what the physical rig should show if the twin's nominal parameters hold. The integrator, the drag law, the shared velocity-form PID and the SIMC rules are standard. The central result is the series restriction: a band that widens 18.8 times and an open-valve air speed that falls from 7.08 to 1.17 times the hover speed, measured on the twin and awaiting the bench. Section 5.2 reviews levitated-ball laboratories. Section 5.3 writes out the model with each equation and constant, and Section 5.4 linearizes it about hover, compares the valve with and without the restriction and tunes the loop. Section 5.5 describes a controller that reads only the sensor, the bench tests that identify the rig, and a hardware session whose firmware, protocol and protections were checked without hardware. Section 5.6 measures ten cases and a graded challenge, and Section 5.7 states the predictions for the hardware. Section 5.8 sets the rig's devices beside their industrial counterparts, and Sections 5.9 and 5.10 give the limitations and the conclusions.

5.2 Floating Balls in Teaching Laboratories

The early air-levitation rigs were free jets. A beach ball above a dc blower, its height tracking a reference through the blower voltage, was built as a remote laboratory for a distance course, with MATLAB/Simulink and xPC Target as the controller platform and video conferencing as the link to students at a second campus [1]. In a survey the distance students rated the remote experience as favorably as the local students, and the course grades of the two groups were similar [1]. A jet holds its own height, because the air slows as it spreads, so a ball that rises meets slower air and falls back. The comparative literature on hands-on, simulated and remote laboratories had by then argued that the three serve different objectives [7], which is the case for running a twin beside a rig rather than choosing between them.

The tube came next, in a design for teaching on the bench and over the internet that the history segment below describes [2]. Fan-driven tubes followed as low-cost builds whose sensor and fan were chosen for the purpose [8], as a ball-and-pipe setup controlled through an Arduino and a real-time Simulink toolbox with an infrared ranger [9], and as a blower-fed plant with an ultrasonic ranger on which a modified PID structure was compared with the classic one [10].

> ⧖ HISTORY **A Ball in a Tube Goes Online, 2017**
>
> A design published in 2017 for teaching on the bench and over the internet floated a light object in a methacrylate tube above a computer fan. An infrared ranger, which updates about every 40 ms, read its position, and a single-board computer closed the loop. The model was a quadratic drag on the air's speed past the ball. Because it holds a pure integrator, the plant is unstable in open loop, and a step input is likely to drop the ball or send it to the top. The designers therefore identified it in closed loop under a proportional controller with a 100 ms period and fitted a fourth-order model with an 85 % fit. A servo-driven flap under the fan served only to disturb the flow, the role this chapter's flap reverses [2].

Open-source shields then packaged the plant. FloatShield is an air-levitation device built as an Arduino shield from off-the-shelf parts and 3D-printed mounts, programmable from the Arduino IDE as well as from MATLAB and Simulink [3]. Its hardware documentation lists a 30 mm cork ball in a 35.5 mm tube, a diameter ratio of 0.85, over a 40 mm axial fan, read by a VL53L0X time-of-flight ranger, the sensor family this rig uses. It belongs to an initiative of take-home shields with a shared programming interface [11]. The Temperature Control Lab had carried the board-sized plant into heat transfer earlier, as a benchmark in which the board's cycle time, its sampling, the serial link's

overhead and the gap between model and device are deliberate features [12], and Chapter 4 runs that device beside its twin.

Control on these rigs has grown past the single PID. A laboratory plant with its own board and firmware was modeled parametrically, fitted by nonlinear least squares and controlled by optimal state feedback with a Kalman filter [13]. A disturbance-rejection design with an extended state observer, needing only the sign of the plant's control coefficient, was tested on a ball driven by a blower voltage [14], and a fan-driven rig was given a honeycomb upstream of its fan to promote laminar flow and reduce the Magnus effect on the ball [15].

The physics behind the lift is older. A quadratic drag law holds for Reynolds numbers from about 10^3 to 2×10^5 and a linear one below 1 [16], and the wall effects on a sphere moving along a tube are collected in the particle literature [17]. A ball that nearly plugs its tube sits at the edge of both. In free air the 0.30 g ball would fall at 5.38 m/s with a drag coefficient of 0.47, a ball Reynolds number near 7100, while at hover the air squeezes through the gap at 4.0 m/s, a Reynolds number of about 290 on the gap's width. The rig differs from the surveyed ones in four respects. It throttles a constant blower rather than modulating a fan. Its ball is at a diameter ratio of 0.90, against 0.85 for the shield whose dimensions are published. Its sensor looks up through the inlet from below the seat. And its twin and its hardware run under one controller object in a browser page, the TCLab pattern carried to a mechanical plant.

5.3 The Rig as Modeled

The rig is a clear polycarbonate tube 16 in tall with a 1 in bore, a 0.9 in styrofoam ball, a small blower at full power on a 5 V, 2 A supply blowing along the base into an elbow, a flap on an SG90 servo, an orifice plate, a VL53L0X time-of-flight sensor under the seat and an Arduino Leonardo, as shown in Figure 5.1. The ball's height z is the height of its underside above the seat, positive upward. Every parameter in Table 5.2 is a nominal rig awaiting bench identification. No rig has yet been weighed, calipered or flown, and the Model dialog and the Instructions page of the application both carry that label. Table 5.1 classifies the signals, and the application's Rig view draws the same parts with the ball's true and read heights, as shown in Figure 5.2.

> ✦ DID YOU KNOW? **A Levitation Rig for Under 100 Euros**
>
> The designers of an open air-levitation rig published in 2017 estimated 200 hours for its software, 60 hours for the design of its structural parts and 10 hours with under 100 euros of parts for its assembly. Their times for replication were the mean of five people given the design files [2].

5.3.1 The Ball and Its Lift

One translational degree of freedom carries the ball, and a momentum balance on it in the form of Equation (1.1) gives Equation (5.1).

$$\begin{aligned} m\dot{v} &= F_{lift}(u-v) - mg - F_{wall}(v), \qquad \dot{z} = v \\ F_{lift}(w) &= \tfrac{1}{2}\rho\, C_{eff} A_b\, w\,|w|^{n-1}, \qquad u = Q_f/A_t \end{aligned} \tag{5.1}$$

where $w = u - v$ is the air's speed past the ball, $A_b = \pi d^2/4 = 411.9\,\text{mm}^2$, $A_t = \pi D^2/4 = 506.7\,\text{mm}^2$, ρ the air density, n the lift exponent (2 by default) and C_{eff} a lift coefficient that stands for the drag coefficient divided by the square of a wall factor. With the mass change, $m = m_0(1+\Delta m/100)$. Nothing on the right depends on z. The hover speed follows from $F_{lift} = mg$ as shown in Equation (5.2).

$$u_{eq} = \left(\frac{2mg}{\rho\, C_{eff} A_b}\right)^{1/n}, \qquad \tau_v = \frac{u_{eq}}{n\,g} \tag{5.2}$$

For the 0.30 g ball that is 0.753 m/s, or 0.382 L/s, with $\tau_v = 38.4$ ms and a pressure under the ball of $mg/A_b = 7.15$ Pa. The coefficient $C_{eff} = 24$ is bracketed by two design estimates. A Newton-regime

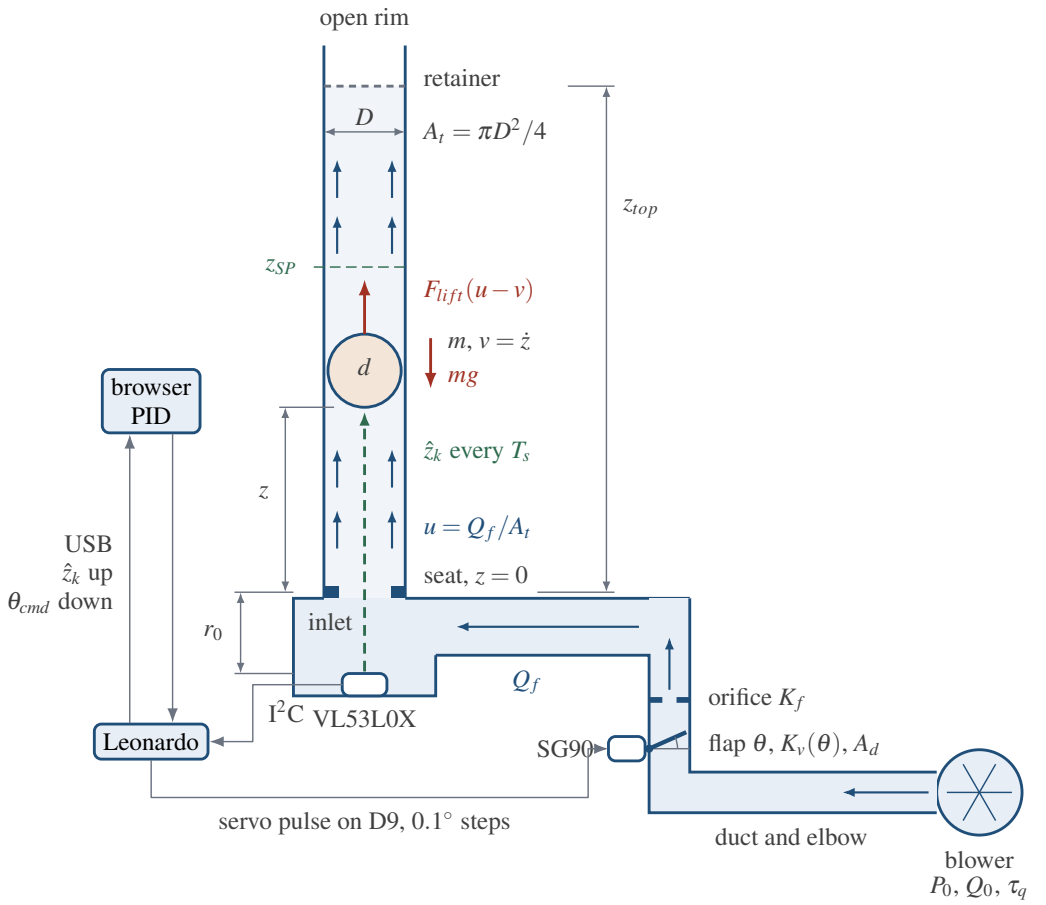


Figure 5.1 The rig as modeled, not to scale. The ball of diameter d and mass m rides in a tube of bore D with its underside at z above the seat, between the seat and a retainer that stops it at z_{top}, 10 mm below the open rim. The blower (P_0, Q_0, lag τ_q) blows along the duct into the elbow, through the flap at angle θ (resistance $K_v(\theta)$ in a duct of area A_d, seated at θ_0, open at θ_1) on the SG90's horn, then through the orifice plate K_f into the inlet below the seat. The flow Q_f rises at $u = Q_f/A_t$ and lifts the ball with $F_{lift}(u-v)$ against its weight. The VL53L0X sits r_0 below the seat and delivers $\hat{z}_k$ every T_s. The Leonardo passes readings to the PID in the browser and its valve requests to the servo.

wall factor $1-\lambda^{1.5}$ at $\lambda = d/D = 0.902$ scales the free-air terminal speed to 0.775 m/s, and the annulus treated as an orifice with a discharge coefficient of 0.65 passing 7.15 Pa gives 0.449 m/s. The correlation is used only as a bracket, so C_{eff} is a fitted parameter, and the drop test of Section 5.5 fits it. Wall friction is off by default and is the regularized Coulomb law $F_{wall} = F_c \tanh(v/v_\varepsilon) + c_v v$ with $v_\varepsilon = 10\,\text{mm/s}$ and $c_v = 1\,\text{mN s/m}$, which keeps the fourth-order Runge–Kutta stage stable at 1 ms up to 2 mN on the lightest ball.

5.3.2 Blower, Flap, Orifice and Tube as One Network

The air below the ball holds a few pascals in about a tenth of a liter, which relaxes in well under a millisecond, so the flow network is algebraic. A parabolic fan curve $P = P_0(1-(Q/Q_0)^2)$ meets the series resistances and the pressure the ball holds up, and the flow that results reaches the ball through

Table 5.1 Signals by Role

Symbol	Role	Nominal	Range
z	CV and state, true height	200 mm	0 to 370 mm
v	state, ball velocity	0	—
Q_f	state, air flow past the ball	0.382 L/s	0 to Q_0
θ	state, flap angle	24.0°	0 to 90°
$\hat{z}$	measured, delivered reading	z	1 mm steps
θ_{cmd}	MV, valve command	holds z	θ_0 to 90°
z_{SP}	setpoint	200 mm	50 to 350 mm
Δm	DV, mass change	0	−20 to 60 %
σ_n	DV, sensor noise at 33 ms	3 mm	0 to 8 mm
T_s	DV, timing budget	33 ms	20, 33, 66, 200
δ	DV, servo dead band	0	0 to 3°
F_c	DV, wall friction	0	0 to 2 mN
—	events: tap, add mass, sag, dropout	—	—
u, u_{eq}, K_i	outputs, air speed, hover speed, gain	—	—

one lumped lag, as shown in Equation (5.3).

$$\begin{aligned} Q &= Q_0\sqrt{\frac{1-\Delta P_b/P_0}{1+Q_0^2\,(K_v(\theta)+K_f+K_t)/P_0}} \\ \tau_q\dot{Q}_f &= Q-Q_f, \qquad K_v(\theta)=\frac{\rho}{2\,(C_cA_d\,a(\theta))^2} \\ a(\theta) &= a_{min}+(1-a_{min})\,\mathrm{clamp}\Big(\frac{\theta-\theta_0}{\theta_1-\theta_0},0,1\Big)^p \end{aligned} \tag{5.3}$$

where $\Delta P_b = F_{lift}/A_b$ is the ball's pressure drop (limited to $0.9P_0$ in the code), K_f the orifice, K_t the tube, C_c a discharge coefficient, A_d the duct area at the flap and $a(\theta)$ the opening fraction. The flap seats at $\theta_0 = 8°$ with 1 % leakage and opens fully at 90° as $(\Delta\theta)^{1.5}$, its inherent characteristic. A supply sag multiplies P_0 and Q_0 by 0.85 for 0.3 s. The resistances are in Pa s^2/m^6, and $K_f = 1.44\times10^9$ is a 6.1 mm hole at $C_c = 0.65$.

5.3.3 Servo

The command is rounded to the firmware's 0.1° and handed to the servo only when it moves more than the dead band from the angle the servo was last given, as shown in Equation (5.4).

$$\begin{aligned} \theta_{req} &= 0.1\,\mathrm{rnd}\big(10\,\mathrm{clamp}(\theta_{cmd},0,90)\big) \\ \theta_{held} &\leftarrow \theta_{req} \text{ only if } |\theta_{req}-\theta_{held}|>\delta \\ \dot{\theta} &= \mathrm{clamp}\big((\theta_{held}-\theta)/\tau_a,\ \pm\omega_{max}\big) \end{aligned} \tag{5.4}$$

with $\tau_a = 25$ ms and $\omega_{max} = 600°/\mathrm{s}$. The dead band δ stands for the SG90's pulse dead band and gear backlash. The firmware maps 0 to 180° onto 544 to 2400 µs, so 0.1° is about 1.03 µs of pulse and a dead band of 10 µs is about 1°.

🕮 **Pause.** Try Exercise 5.3 before reading on.

5.3.4 Sensor

Every T_s a ranging measurement starts, and its reading arrives one T_s later, as shown in Equation (5.5).

$$\begin{aligned} r_k &= \max\big(0,\ 10^{-3}\,\mathrm{rnd}(10^3(z(t_k)+r_0+b+\sigma\,n_k))\big) \\ \hat{z}_k &= r_k - r_0 \text{ delivered at } t_k+T_s \\ n_k &= \mathrm{e}^{-1}n_{k-1}+\sqrt{1-\mathrm{e}^{-2}}\,\xi_k \end{aligned} \tag{5.5}$$

Table 5.2 Nominal Rig Awaiting Bench Identification. V Marks a Model-Dialog Slider and H a Hidden Constant. The Provenance Column Names the Bench Measurement That Replaces Each Value

Sym.	Meaning	Value	Range	Provenance
m_0	ball mass	0.30 g	0.10–0.80	V weigh
d	ball diameter	22.9 mm	20–24	V given
D	tube bore	25.4 mm	24–32	V caliper
z_{top}	travel to retainer	370 mm	300–380	V measure
ρ	air density, Provo	1.05 kg/m^3	0.9–1.3	V site
C_{eff}	lift coefficient	24	5–60	V drop test
n	lift exponent	2	1–2	V step tests
P_0	dead-head pressure	300 Pa	50–1000	V manometer
Q_0	free delivery	3.0 L/s	0.5–10	V anemometer
τ_q	flow lag	30 ms	0–150	V step tests
A_d	duct area at flap	507 mm^2	50–600	V measure
θ_0, θ_1	seat, full open	8°, 90°	—	V hover map
p	opening exponent	1.5	1–3	V hover map
a_{min}	seated leakage	0.01	0–0.05	V drop test
K_f	orifice	1.44×10^9	0–1×10^{10}	V orifice size
C_c	discharge coefficient	0.65	0.5–0.9	V fixed
K_t	tube friction	2×10^6	—	H —
τ_a	servo lag	25 ms	5–100	V bench step
ω_{max}	servo rate	600 °/s	100–1000	V bench step
r_0	sensor below seat	40 mm	0–80	V calibration
b	residual bias	0	±10	V calibration
g	gravity	9.81 m/s^2	fixed	H —
u_{eq}, τ_v	hover speed, lag	0.753 m/s, 38.4 ms	derived	
θ_{hov}, K_i	hover angle, gain	23.96°, 16.88	derived	

where heights are in meters, $r_0 = 40$ mm is the sensor's distance below the seat, b a residual bias (0 by default), ξ_k a standard normal draw from the sensor's own random stream, seeded from the case's seed, and σ the noise of the budget in force, σ_n scaled by 5/3, 1, 1/2 and 1/3 at 20, 33, 66 and 200 ms, so the default 3 mm gives 5, 3, 1.5 and 1 mm. A raw range under 45 mm is flagged FLOOR and one within 5 mm of the retainer CEILING, and both are still delivered. A dropout event flags five readings invalid, and each repeats the last valid reading. A change of budget drops the reading in flight, so the command holds until the first reading of the new period.

5.3.5 Stops, Initial State and Numerics

The seat ($z = 0$) and the retainer ($z_{top} = 370$ mm) are perfectly inelastic stops applied after each step, and each arrival is counted. With the retainer parameter off, a ball that reaches the rim faster than 0.5 m/s is ejected and the run pauses. A run starts with the ball at rest at the case's height, the flow at the network's value and the flap at the 0.1° step nearest the true hover angle in Auto, or at the manual station in Manual. The four states are integrated by the framework's fourth-order Runge–Kutta step (Section 1.6.1) at $\Delta t = 1$ ms, the controller is called every step and runs once per delivered reading, and the historian logs at 50 Hz. Halving the step moves the challenge's IAE by 6.8×10^{-9} of itself and its peak height by 1.3×10^{-7} mm. The application passes its 49 verification checks, among them the inversion of the static map, the drop speed, the band ratio and a grep-level test that the controller's source never reads the true height, the true mass or the stop flags.

5.3.6 Assumptions

A nominal rig. The blower curve, the ball's mass, the flap's shape and the orifice are design proposals. They break down as soon as the rig is measured, and Section 5.7 gives the procedure that replaces

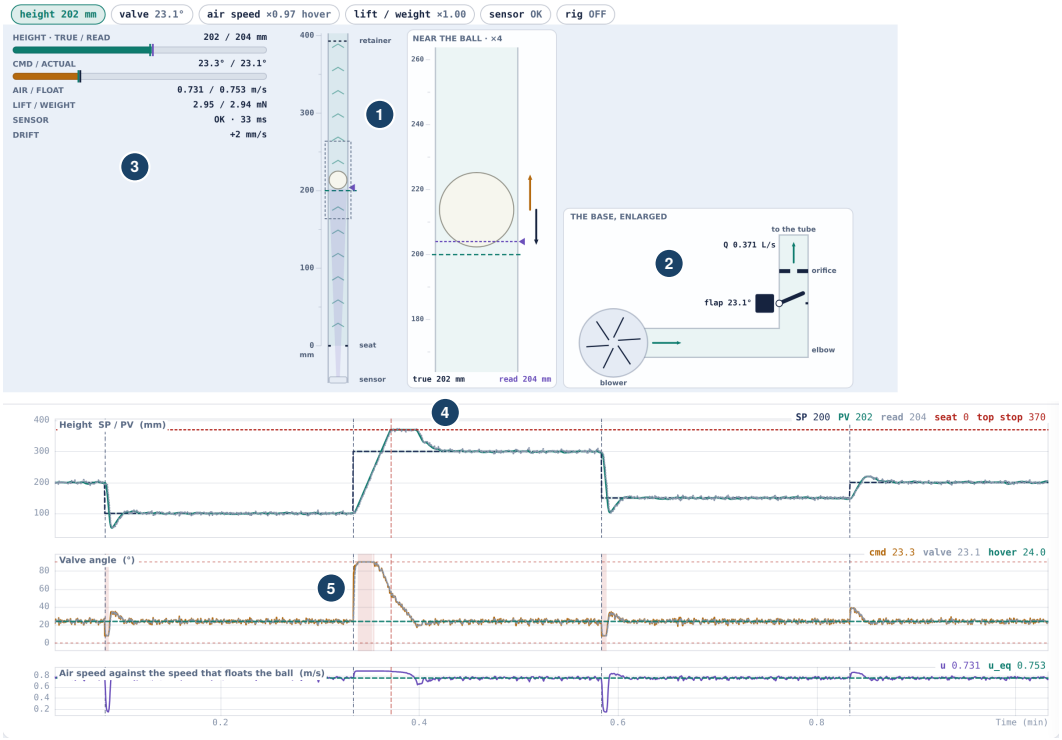


Figure 5.2 The application in case 10, *A setpoint program, on the twin and on the rig*, as shipped at $t = 62$ s on the Rig view with a 1 min trend window and no rig connected, cropped to a composite of the Rig view above the trends. (1) The tube and the zoom near the ball show the ball at a true height of 202 mm read as 204 mm against the 200 mm setpoint. (2) The enlarged base shows the flap at 23.1° and a flow of 0.371 L/s. (3) The readouts give a command of 23.3°, an air speed of 0.731 against 0.753 m/s and a lift of 2.95 against 2.94 mN. (4) The height trend shows the four steps and the rise to 300 mm reaching the retainer. (5) The command is shaded where the PID asks past the open flap during the rise, as it does past the seat after the down steps, and the air speed drops on each down step. The scorecard, cut from this crop, reads an IAE of 554.9 mm s, a valve travel of 3358° in both tiles, one retainer contact and 2.4 % of the time saturated, and its overshoot and settling tiles, 36.5 % and 9.56 s, describe the last step from 150 to 200 mm.

them.

One degree of freedom. No spin, no sideways motion and no rattling against the wall. A fan-driven rig was given a honeycomb to reduce the Magnus effect on its ball [15], a force this model leaves out.

A drag law with a fitted coefficient. At a gap Reynolds number near 290 the flow is below the quadratic range [16], so the true law may lie nearer linear, which the exponent n covers. The linearized gain ng/u_{eq} is what the fits see either way.

Quasi-static air. No pressure capacitance below the ball, one lag for the air column and the blower, a parabolic fan curve at one speed, and a flap whose opening has a fixed form. The supply sag is the only blower change.

A plain sensor. Rounding, a bias, colored noise and one sample of latency. The 25° cone of the VL53L0X is 133 mm wide at 300 mm and always sees the wall, and the cross-talk this adds beyond the bias, ambient light and range-dependent noise are not modeled.

Hard stops. The seat and the retainer absorb all energy. Temperature and humidity are constant.

5.4 An Integrator Behind a Valve

Linearized about hover in deviation variables (Section 1.4.3), Equation (5.1) gives $\tau_v\,\delta\dot{v} = \delta u - \delta v$ with $\delta\dot{z} = \delta v$, and the valve adds its slope, so the plant from valve angle to height is the transfer function of Equation (5.6).

$$G(s) = \frac{Z(s)}{\Theta(s)} \approx \frac{K_i\,\mathrm{e}^{-\theta_d s}}{s\,(\tau_v s + 1)}, \qquad K_i = \frac{1}{A_t}\frac{dQ}{d\theta}\bigg|_{\theta_{hov}} \tag{5.6}$$

where $\theta_d = T_s + \tau_a + \tau_q = 88\,\mathrm{ms}$ at the 33 ms budget lumps the sample, the servo and the flow lag as dead time (Section 1.4.6), and $K_i = 16.88\,\mathrm{mm/s}$ per degree at the hover angle of 23.96°. The coefficient $1/\tau_v = ng/u_{eq}$ is the $2g/v_{eq}$ of the fan-driven tube [2]. There is no restoring term. A valve 0.5° off hover drifts the ball at 8.4 mm/s, and even the firmware's 0.1° step nearest hover, 24.0°, drifts it upward at 0.75 mm/s.

🕮 **Pause.** Try Exercise 5.1 before reading on.

The gain is not one number, as shown in Figure 5.3. K_i is 30.4, 16.8, 5.2 and 1.0 mm/s per degree at 20, 24, 32 and 46°, because the opening grows as $(\Delta\theta)^{1.5}$ while the blower, orifice and tube take a rising share of the pressure. A ball 20 % heavier hovers at 31.14°, where K_i is 5.78, 2.92 times less, which is case 6. The usable band, defined as the travel from the hover angle to the angle that gives $1.15u_{eq}$, is 22.34° wide, from 23.96 to 46.29°. The fully open valve gives $1.17u_{eq}$, so the ball rises at 0.125 m/s at most.

5.4.1 The Series Restriction

Without the orifice the same blower must be throttled by the flap alone. At the hover flow the flap then drops 287.7 Pa of the blower's 295.1 Pa, 97.5 %, its opening is 0.0495 against the 0.0100 of a seated flap, and the hover angle falls to 17.57°, 9.57° above the seat. The band shrinks to 1.19°, about the width of the SG90's dead band, and K_i at hover rises to 92.5 mm/s per degree, 5.48 times the restricted rig's. With the orifice the plate takes 209.7 Pa (71 %) and the flap 78.0 Pa, the opening at hover is 0.095, and the open valve gives $1.17u_{eq}$ instead of $7.08u_{eq}$. The band widens 18.8 times, and the cap on the open valve's air speed also bounds how fast a windup can throw the ball at the retainer. This is the result the bench has to confirm.

🕮 **Pause.** Try Exercise 5.2 before reading on.

5.4.2 Tuning an Integrator With a Lag

The half rule of Equation (1.16) folds half of τ_v into the dead time, $\theta' = \theta_d + \tau_v/2 = 0.107\,\mathrm{s}$, and the SIMC rules for an integrating process with $\tau_c = \theta'$ give Equation (5.7) [5]. Volume 2, Chapter 5 builds the tuning of an integrating level loop from the same rules.

$$K_c = \frac{1}{K_i(\tau_c + \theta')} = \frac{1}{2K_i\theta'}, \qquad \tau_I = 4(\tau_c + \theta') = 8\theta' \tag{5.7}$$

which gives 0.276 °/mm and 0.858 s. The application adds a derivative time equal to the lag the half rule folded away, $\tau_D = \tau_v = 0.0384\,\mathrm{s}$, with a filter of $\tau_D/2$, and rounds the four to the sliders, 0.28 °/mm, 0.86 s, 0.038 s and 19 ms. On an integrating plant a PI loop overshoots a setpoint step by design, because the integral must carry the valve past the hover angle to stop the ball. With the noise off this tuning overshoots a 50 mm step by 19.9 mm and settles in 1.97 s, and a 10 mm step by 3.1 mm in 0.23 s, so an overshoot of 5 % estimated before the twin was built does not hold. On case 4's step, integral times of 0.3, 0.4 and 0.5 s give three, two and one crossings of the new setpoint, so the SIMC value is 1.7 times the 0.5 s at which the ball first crosses only once. A sensor at 200 ms makes $\theta' = 0.274\,\mathrm{s}$, and the reading's age alone adds 0.167 s, which is case 7.

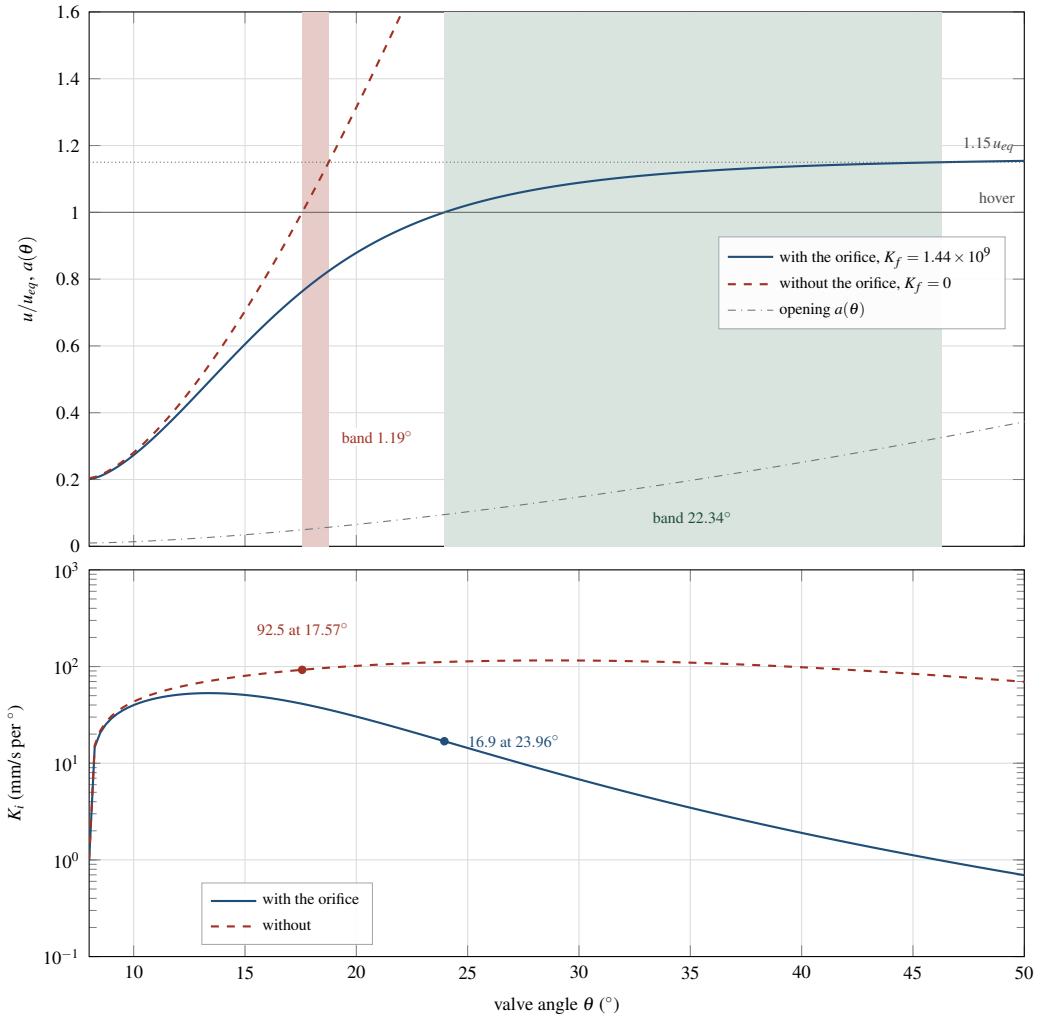


Figure 5.3 The static map of the nominal rig with the 0.30 g ball, from the application's own network functions. Top: air speed as a ratio to the hover speed with the orifice (solid) and without it (dashed), and the flap's opening $a(\theta)$. The shaded bands run from the hover angle to $1.15u_{eq}$, 22.34° with the orifice and 1.19° without. Bottom: the integrating gain $K_i(\theta)$ on a logarithmic axis, with each hover angle marked.

5.5 The Controller as Coded

The controller reads the delivered reading $\hat{z}_k$, its status, the setpoint and its own state. It never reads the true height, the true mass, the true flow or the stop flags, and a test in the application's suite reads its source to enforce that. Table 5.3 lists its settings.

5.5.1 The Shared PID Once per Reading

The engine calls the controller every 1 ms step, and the law runs only on the step at which a reading is delivered, with a period of T_s, so the sample period follows the timing budget without changing the engine's own control period. The law is the shared velocity-form PID of Section 1.6.2, written in Equation (5.8) in this chapter's symbols: $\hat{z}_k$ is the PV, θ_b the bias u_b, $\hat{u}_k$ the output v_k before the

Table 5.3 Controller Settings on the ZC Faceplate

Symbol	Setting	Default	Range
z_{SP}	height setpoint	200 mm	50 to 350
θ_{man}	manual valve station	24.0°	0 to 90, 0.1°
K_c	proportional gain	0.28 °/mm	0 to 2
τ_I	integral time, 0 is off	0.86 s	0 to 10
τ_D	derivative time	0.038 s	0 to 0.2
τ_f	derivative filter (Advanced)	19 ms	0 to 100
r_{SP}	setpoint ramp, 0 is off (Adv.)	0	0 to 300 mm/s
—	hover-angle feedforward (Adv.)	on	on, off

limits, and the action a is $+1$.

$$\begin{aligned} e_k &= z^*_{SP,k} - \hat{z}_k, \qquad P_k = K_c e_k \\ D_k &= -K_c \tau_D \left(\hat{z}_{f,k} - \hat{z}_{f,k-1}\right)/T_s \\ S_k &= S_{k-1} + (P_k - P_{k-1}) + (D_k - D_{k-1}) \\ \hat{u}_k &= \theta_b + S_k + I_{k-1}, \qquad \theta_{cmd} = \mathrm{clamp}(\hat{u}_k,\ \theta_0,\ 90) \\ I_k &= I_{k-1} + (K_c/\tau_I)\, e_k T_s + (\theta_{cmd} - \hat{u}_k)\, T_s/\tau_I \end{aligned} \tag{5.8}$$

where $\hat{z}_f$ is the reading through the exact first-order lag $\hat{z}_{f,k} = a\hat{z}_{f,k-1} + (1-a)\hat{z}_k$ with $a = \mathrm{e}^{-T_s/\tau_f}$, the gains act on errors in meters with K_c stored in degrees per meter, and the integral is updated only when $\tau_I > 0$. The action is direct, more valve for a ball below its setpoint. The back-calculation's tracking time is τ_I.

The lower limit is the flap's seat angle $\theta_0 = 8°$ of the fitted map, not 0°. Below the seat the flap is already closed, and with a limit at 0° the integral would wind through 8° of dead travel before the back-calculation acted, which deepened the overshoot of case 10's down steps until the application's technical review moved the limit. The saturation flag compares $\hat{u}_k$ with the two limits in Auto and is cleared every sample.

The bias is the static map's hover angle at a declared mass, $\theta_b = \theta_{hov}(0.30\,\mathrm{g}) = 23.955°$, computed on the parameters in force but never from the ball's true mass, so after a mass change the feedforward is wrong by a known amount and the integral finds the rest. The setpoint ramp rate-limits the setpoint the PID sees, $z^*_{SP,k} = z^*_{SP,k-1} + \mathrm{clamp}(z_{SP} - z^*_{SP,k-1}, \pm r_{SP} T_s)$, in Auto only. In Manual the PID tracks the station at every reading, so the return to Auto is bumpless (Section 1.6.2), and in Auto the station mirrors the command, which lets a change to Manual leave the valve where it was. The shared PID's slew limit is not used.

5.5.2 The Bench Tests as Helpers

Three helpers run in Manual. *Set valve to hover angle* applies θ_b rounded to 0.1°. *Step test* $\pm 2°$ needs a ball that has flown at least 1 s in Manual, steps the station and fits up to 3 s of readings before the step and 6 s after it, stopping 30 mm short of either stop, by least squares on the simulation (output) error of Section 1.4.8 with Equation (5.9).

$$y(t) = y_0 + v_0 t + K_i \Delta\theta \left[(t - t_d) - \tau\left(1 - \mathrm{e}^{-(t-t_d)/\tau}\right) \right]_{t > t_d} \tag{5.9}$$

where v_0 absorbs the drift the ball already had. For each (τ, t_d) the three linear parameters follow from linear least squares, and τ in 3 to 600 ms and t_d up to 400 ms come from a grid refined by a coordinate search. The fit reports K_i, τ, t_d and the SIMC tuning they imply with $\theta' = t_d + \tau/2$. *Drop test* opens the valve until the reading passes 300 mm, closes it and fits a straight line to the fall from 0.25 s after the close until 30 mm above the seat. With the valve closed only the seated leakage flows,

Table 5.4 Firmware Protocol, One Line per Command

Command	Reply	Meaning
`VER`	identity	Ball Levitation Firmware 1.0.0 Leonardo
`Z`	`h,s`	latest height (mm) and status
`V 24.3`	`24.3`	valve request, clamped to 10 to 90°, the angle given
`S 1/S 0`	on/off	stream `t_ms,h,s,theta,flags` per sample
`P 33`	`33`	timing budget, 20, 33, 66 or 200 ms
`L 360`	`360`	ceiling, 50 to 400 mm
`C`	`r0=41`	record the resting range as r_0 in EEPROM
`X`	`Closed`	valve closed, stream off
`A`	`Start`	clear the watchdog and ceiling latches

so $u_{eq} = v_{fall} + u_{leak}$, with $u_{leak} = 0.151\,\mathrm{m/s}$ from the fitted map, and $C_{eff} = 2mg/(\rho A_b u_{eq}^n)$ at the declared mass.

The models stay apart, as Section 1.4.9 keeps the simulation, the identified model and the controller's model apart. The simulation runs the true rig, the PID's bias and the drop test use the declared 0.30 g ball on the fitted map, the linearization of Equation (5.6) is the tuning model, and a fourth model, an integrator behind a lag about one angle, comes out of each step test.

5.5.3 The Hardware Session

The same page drives the physical rig over Web Serial from a Chromium-based browser on HTTPS or localhost, as the TCLab application does. The rig follows the app: the mode, the setpoint, the station, the tuning, the ramp and the timing budget are shared, while the rig's PID is a second instance built by the twin's factory with its own integral and filter state, its lower limit raised to the firmware's closed angle of 10°. Its loop is driven by the wall clock. Each stream line from the firmware runs the law once with the period from the firmware's own clock and sends one valve request, a newer request replacing one still waiting. On connect the PID starts in tracking with the valve closed and the ball on its seat, so the first move is the controller's, and while the twin is paused the rig holds the valve closed. Rows are logged on the twin's clock, the speed is held at 1×, and the rig's CSV carries the twin's columns without the true height and with the firmware's clock and a flag word added. The helpers run on the rig's readings as well. Table 5.4 gives the protocol at 115 200 baud.

The firmware protects the rig whatever the browser does. The servo's pulse is set to the closed angle before it is attached, so the valve starts closed, and closed means 10°, 2° short of the seat, so the servo never stalls against it. In stream mode a command must arrive at least once a second, and after 1000 ms without one the valve closes and every stream line carries `WD` until `A`, so a browser that stalls and recovers does not reopen the valve. Two valid readings in a row above the ceiling close the valve and set `CEIL`, which releases 20 mm below it, and an invalid reading neither trips nor releases the catch. The shipped image is the I^2C build, 14 244 bytes of program, 49 % of the 28 672 below the boot loader, and 541 bytes of RAM, 21 % of 2560, compiled with arduino-cli 1.5.1 on the Arduino AVR core 1.8.8 with an open-source VL53L0X library (version 1.3.1, MIT license) included unmodified. The page installs it through the Leonardo's boot loader with a write-and-verify flasher whose three transport files are copied byte for byte from the TCLab application.

5.6 Ten Cases and a Challenge, Measured

Table 5.5 sets out the cases and Table 5.6 what the engine measured. Each run was stepped outside the browser one 1 ms step at a time at the app's default seed, 20261025, with any change the text names made before Run, which gives the same run as a change carried through Reset (Section 1.6.1). Each case also carries a completion rule judged on the twin, so no reader needs a rig, and the exercises of Section 5.11 use the twin alone. Heights are the true heights. A settling time runs from the change until the ball stays within ±5 mm for at least 1 s, as the scorecard's tile reads it, valve travel sums the

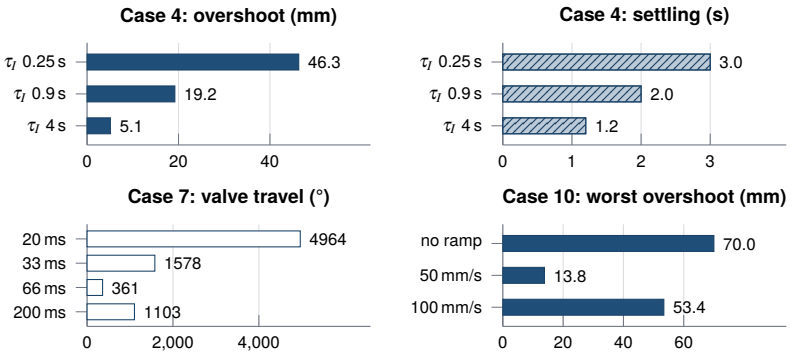


Figure 5.4 Four rows of Table 5.6 as bars: the overshoots of case 4 at three integral times and of case 10's setpoint program at three ramps (solid), the settling of case 4 (hatched) and the valve travel of case 7 at four timing budgets of the sensor (open). Notice that the longest integral time is the fastest to settle, that the 66 ms budget moves the valve least, and that a 50 mm/s ramp gives a smaller worst overshoot than no ramp or a faster one.

Table 5.5 Scenario Sequence. Every Case Starts at 200 mm Under the Suggested Tuning Unless Its Row Says Otherwise

#	Setup	Teaching point
1	Auto, 3 mm noise, switch to Manual	no angle holds
2	Manual at 24.0°, noise off, drop and step tests	u_{eq} and K_i
3	P with τ_D 0.04 s, noise off, SP 250 at 5 s, +20 % mass at 20 s	offset under load
4	τ_I 0.25 s, noise off, SP 250 at 5 s	slow ringing
5	K_c 0.80 °/mm, τ_I 1.5 s, noise off, SP 220 at 5 s, tap at 15 s	derivative as damping
6	+20 % mass, noise off, SP 220 at 5, 200 at 20 s, sag at 35 s	gain falls with the angle
7	3 mm noise, budgets 20 to 200 ms, SP 250 at 5, tap at 15 s	period against noise
8	dead band 1°, noise off	limit cycle
9	Manual, $K_f = 0$ at 17.6°, noise off, Valve view	a one-degree band
10	3 mm noise, SP 100, 300, 150, 200 at 5, 20, 35, 50 s	ramp against windup
C	P with τ_D, 3 mm noise, SP 150, 250 at 5, mass at 18, 200 at 30, tap at 45 s	everything at once

moves of the Auto command, and an IAE is the height loop's from $t = 0$ over the window named. Figure 5.4 draws four of the table's rows as bars.

5.6.1 By Hand, and Why the Valve Cannot Be Left

Case 1 opens on the ball hovering in Auto with the rig's 3 mm of noise at 33 ms, which the loop turns into a command with a standard deviation of 1.58° from 2 to 30 s. Switched to Manual at 3 s, the valve stays at 22.56°, 1.40° under hover, and the ball comes down on its seat 8.1 s later. In ten switches between 3 and 15 s the valve was left 0.04 to 3.92° from hover, and nine balls reached a stop 3.3 to 23.2 s later, while the tenth, left at 23.99°, was still flying 40 s later. Over 364 switches, one at every reading from 3 to 15 s, half reached a stop within 10.4 s, 300 within 30 s and 323 within 60 s, and the rest crept at 0.8 to 2.7 mm/s. One press of the arrow key, 0.5° the way a creeping ball went, sent each of those to a stop 14 to 20 s later.

Table 5.6 Outcomes Read From the Running Engine

Case and condition	Result
1, switch at 3 s	valve 22.56°, on its seat 8.1 s later
1, 364 switches, 3 to 15 s	median 10.4 s to a stop, 323 within 60 s
2, drop test	fall 0.592 and u_{eq} 0.744 m/s, C_{eff} 24.6
2, step tests + / −2°	K_i 14.3–14.5 / 19.6–19.8, lag 118–134 ms
3, P only, mass step	settles 25.45 mm low
3, τ_I 0.9 s	mass peak 19.9 mm, back in 2.0 s
4, τ_I 0.25 / 0.9 / 4 s	over 46.3 / 19.2 / 5.1 mm, 3.0 / 2.0 / 1.2 s
5, τ_D 0.04 / 0, noise off	travel 646 / 931°, tap 0.21 s / never
5, 3 mm, τ_D 0.04 / 0	travel 4786 / 2334° by 30 s
6, up / down / sag, shipped	2.64 / 1.75 / 2.88 s (0.30 g: 1.05 / 0.64 / 0.55)
6, twice / three times K_c	1.78 / 0.77 / 1.91 s; over 9.4 / 25.5 mm
7, 20 / 33 / 66 / 200 ms	IAE 83 / 76 / 72 / 1880, 4964° / 1578 / 361 / 1103
7, 200 ms at 0.17 °/mm	settles 8.3 s, within 8 mm for 13.4 s
8, dead band 1 / 0.5 / 0°	±2.49 / 0.85 / 0.57 mm
9, Manual at 17.6 / 17.7 / 18.1°	retainer at 58.2 / 14 / 3.5 s
9, Auto at 0.28 / 0.05 °/mm	saturated 46.9 / 0 %, error 99.1 / 1.3 mm
10, ramp 0 / 50 / 100 mm/s	worst over 70.0 / 13.8 / 53.4 mm, IAE 555 / 868 / 618
C, shipped / suggested	worst 27.9 / 5.5 mm, 3262° / 1371, fail / pass

TRY IT IN THE APP **1 · Hover by Hand (Start Here)**

Load the case and press Run. The ball hovers at 200 mm in Auto. Switch to Manual while it hovers, so the valve stays at the controller's last command, and watch the ball. Switched at 3 s, the valve stays at 22.56° and the ball comes down on its seat 8.1 s later. At other moments it may rise to the retainer or only creep.

5.6.2 A Drop and Two Steps

Case 2 hovers in Manual at 24.0° with the noise off. The drop test's fall fits 0.592 m/s where the ball truly falls at 0.601 m/s at 150 mm, and with the leakage it gives $u_{eq} = 0.744\,\mathrm{m/s}$ and $C_{eff} = 24.6$, within 1.3 % and 2.6 % of the twin's 0.753 and 24. Over three step instants the +2° step fits 14.3 to 14.5 and the −2° step 19.6 to 19.8 mm/s per degree, chords on either side of the hover angle, and their mean, 17.0, lies within 1 % of the tangent, 16.9. The total lag $\tau + t_d$ fits 118 to 134 ms against the 126 ms of $T_s + \tau_a + \tau_q + \tau_v$, and the reading's 1 mm steps decide how it splits between lag and dead time. With 3 mm of noise over ten seeds in both directions the fitted gain stays within 12 % of its chord while the lag ranges from 67 to 210 ms, so only the gain is identified on a noisy sensor.

5.6.3 Proportional, Then Integral

Case 3 runs 0.28 °/mm with $\tau_D = 0.04\,\mathrm{s}$ and no integral. At 19 s the ball is at 250.4 mm, since the plant's own integrator removes a setpoint offset. After the mass step it settles 25.45 mm low, because the heavier ball hovers at 31.14°, 7.19° above the feedforward, and 7.19° at 0.28 °/mm is 25.7 mm of error. With $\tau_I = 0.9\,\mathrm{s}$ the mass step costs a peak of 19.9 mm and the ball is back within 5 mm 2.0 s later, while the setpoint step now overshoots 19.2 mm. Case 4 shortens the integral time to 0.25 s. After the 50 mm step the ball swings 46.3 over, 21.5 under, 13.3 over, 6.5 under and 3.7 mm over, about 1.2 s from the first to the third swing, and settles in 3.0 s. At 0.9 s it overshoots 19.2 mm once and settles in 2.0 s, and at 4 s 5.1 mm and 1.2 s, the slow cycle of two integrators in a loop giving way as the integral time grows [4].

5.6.4 Derivative, and the Noise It Buys

The ball's lag is short against the loop's dead time, so at the suggested gain the derivative changes the tap's dip by only 0.2 mm (9.0 against 8.8 mm on case 7's setup) while more than doubling the valve travel, 1578 against 691°. Case 5 therefore raises the gain to 0.80 °/mm with $\tau_I = 1.5\,\mathrm{s}$ and the noise off. A tap's 9.8 mm dip is gone in 0.21 s and the valve travels 646° by 30 s, while without the derivative the loop never settles, swinging up to 6.4 mm from 20 to 30 s for 931°. With the rig's 3 mm

of noise this gain is too high either way, 15.3 and 27.6 mm off with and without the derivative, and the derivative doubles the travel, 4786 against 2334°, or 5701° with the filter off.

5.6.5 A Heavier Ball on a Flatter Valve

Case 6 starts 20 % heavy on the suggested tuning with the noise off. The 20 mm step up settles in 2.64 s, the step down in 1.75 s and a supply sag in 2.88 s, against 1.05, 0.64 and 0.55 s for the 0.30 g ball. Twice the gain restores 1.78, 0.77 and 1.91 s, while three times, close to the full ratio of 2.92, overshoots the step down by 25.5 mm against 9.4 mm, because below the hover angle the valve is steep again. Volume 2, Chapter 6 meets a gain that moves too far for one tuning and schedules it.

5.6.6 The Sensor's Bargain

Case 7 steps 50 mm at 5 s and taps at 15 s with noise of 5, 3, 1.5 and 1 mm at 20, 33, 66 and 200 ms. Over 30 s at the suggested tuning the IAE is 83, 76 and 72 mm s at the first three budgets with 4964, 1578 and 361° of valve travel, so the quieter 66 ms budget wins both. At 200 ms the loop oscillates onto its stops six times for an IAE of 1880, because the reading's age alone has grown θ' from 0.107 to 0.274 s. At 0.17 °/mm the 200 ms loop settles the step in 8.3 s after 47.5 mm of overshoot, costs 9.2 mm on the tap and holds within 8 mm for 13.4 s.

> **DID YOU KNOW? A Slower Sensor That Wins**
>
> From the simulator: in case 7 at the suggested tuning, the 66 ms timing budget beat the fastest one, 20 ms, on both counts over 30 s. Its quieter readings gave an IAE of 72 against 83 mm s and moved the valve 361° against 4964° (Table 5.6).

5.6.7 Dead Band

Case 8 gives the servo a 1° dead band and nothing else. Between 15 and 30 s the ball cycles ±2.49 mm every 1.06 s. Half the gain makes it ±6.35 mm every 4.6 s, an integral time of 2 s ±2.70 mm and twice the gain ±1.52 mm, so retuning moves the cycle around. A 0.5° dead band halves it to ±0.85 mm, and none leaves the ±0.57 mm that the firmware's 0.1° steps make. Adding 0.5 mN of wall friction to the 1° dead band turns the cycle into a stick-slip with one swing every 7.2 s.

5.6.8 No Series Restriction

Case 9 takes the orifice out and opens on the Valve view of Figure 5.3. In Manual at 17.6° the ball takes 58.2 s to reach the retainer, at 17.7° 14 s and at 18.1° 3.5 s. In Auto the shipped gain drives the valve against its seat or its open stop 46.9 % of the first 20 s and the ball swings up to 99.1 mm off its setpoint, while 0.05 °/mm, about 16.88/92.5 of the shipped gain, holds the bare rig within 1.3 mm.

> **TRY IT IN THE APP 9 · A Valve With No Series Restriction**
>
> Load the case, which opens on the Valve view with the orifice taken out, and compare the shaded band with and without it. Press Run and move the Valve angle slider a tenth of a degree at a time. On the twin, 17.6° takes the ball to the retainer in 58.2 s, 17.7° in 14 s and 18.1° in 3.5 s.

5.6.9 A Setpoint Program

Case 10 runs setpoints of 100, 300, 150 and 200 mm at 5, 20, 35 and 50 s with the rig's noise, as shown in Figure 5.2. The steps overshoot 46.9, 70.0, 46.9 and 19.2 mm, and the 200 mm rise strikes the retainer once, for an IAE of 555 mm s over 62 s. A 50 mm/s ramp keeps every overshoot to 13.8 mm or less at an IAE of 868, the tracking error the ramp costs. A 100 mm/s ramp cuts the down steps to 16.0 and 18.3 mm but leaves the rise at 53.4 mm, because the open valve lifts the ball at 0.125 m/s at most and a faster ramp still winds the integral at the open flap. Over ramps of 20, 35, 50 and 70 mm/s at the case's seed and two others the worst overshoot is 28.4 mm.

Table 5.7 Predictions of the Nominal Twin for the Bench

Measurement	Twin
first angle that lifts the ball	24° (hover 23.96°)
seat to retainer at 25 / 46 / 90°	22.75 / 3.32 / 3.00 s
band to $1.15u_{eq}$, with / without orifice	22.34° / 1.19°
drop test: fall, u_{eq} (m/s), C_{eff}	0.592, 0.744, 24.6
step tests + / −2°, K_i	14.3–14.5 / 19.6–19.8 mm/s per °
lag $\tau + t_d$, noise off / 3 mm	118–134 / 67–210 ms
case 10, overshoots, no ramp	46.9, 70.0, 46.9, 19.2 mm

5.6.10 The Challenge

The graded case runs 150, then 250 mm at 5 s, a 20 % mass step at 18 s, 200 mm at 30 s and a tap at 45 s with the rig's noise. It judges the true height every millisecond outside the 3 s after each event and the first 3 s, asking for ±10 mm, no stop contact and valve travel under 2000° by 60 s, on the case's own rig. As shipped, proportional with derivative and no integral, the ball sags 27.9 mm after the mass step (worst at 25.4 s) and the valve travels 3262°, and it fails at the case's seed and at seeds 1 to 10. The suggested configuration, PI at 0.28 °/mm with $\tau_I = 0.86$ s, no derivative and a 50 mm/s ramp, keeps the worst error to 5.5 mm at 8.2 s and the travel to 1371°, and it passes at seeds 1 to 10 with 4.6 to 8.5 mm and 1347 to 1401°. The cases' tuning with its derivative holds the band at 6.9 mm but runs the valve 3224°, and the suggested PI without the ramp passes at the case's seed with 7.0 mm and nine of ten other seeds.

5.7 The Rig Beside Its Twin

No physical rig was available when this chapter was written, so this section states predictions as predictions. The twin's numbers are what the bench procedure should find if the nominal table holds, and a mismatch beyond the acceptance limits below is a finding to report rather than a reason to tune the twin until it agrees.

5.7.1 Three Kinds of Evidence

Checks of the code, comparisons with published work and classroom use are three separate matters, and each supports different claims. Verification is the application's 49 checks, the 12 completion-rule walks, an identical trajectory for every case whether it is loaded first, from the selector or after Reset, a second implementation of the static map by hand that matches the hover angle, K_i, the bands and the leakage, and the step-halving result of Section 5.3. Agreement with the literature is limited to form: the linearization has the coefficient of the fan-driven tube [2], and the nominal hover speed lies inside the design bracket of 0.449 to 0.775 m/s. No measured rig confirms a number in this chapter. Educational use has not been measured, and no learning gain is claimed.

5.7.2 What the Bench Should Find

Table 5.7 lists the predictions. For the hover map the servo steps from 10° in 1° increments with the ball on its seat. On the twin the ball does not lift at 23° and lifts at 24°, the first step above 23.96°, drifting up at 0.75 mm/s. From the seat it reaches the retainer in 22.75 s at 25°, 12.53 s at 26°, 3.32 s at 46°, the top of the band, and 3.00 s with the valve open. The drop and step tests are the helpers of Section 5.5 run on the rig's own readings. The acceptance limits set for the bench are a hover angle within 2°, a drop speed within 15 % and a step-test K_i within 25 % of the twin's after the twin is refitted.

5.7.3 What Was Checked Without Hardware

The hardware half was checked against software stand-ins, and its 26 tests pass. The firmware's protocol, watchdog and ceiling logic are plain C compiled on the host, where the watchdog does not

trip at 1000 ms and trips at 1001 ms, and a model of the firmware used by the transport and session tests agrees with that C state for state over a 600-action random script. The transport frames lines split across chunks and keeps stream lines apart from replies, the session's connect sequence closes the valve before streaming, the rig's law sends identical commands when the twin's truth is tampered with, and the installer is exercised against a model of the boot loader, writing and verifying the image. The shipped HEX file is 40 084 bytes with SHA-256 `2ceaa311...71f4a`, the build records report it byte-identical on clean rebuilds, and the three flasher files compare equal to the TCLab application's.

5.7.4 Not Run Without a Rig

The installation from the browser on a real Leonardo, the watchdog closing the valve within 1.2 s of a pulled cable, the ceiling catch at a set height, a servo stalled by hand without a board reset, a ten-minute hover without a dropped sample, a clean disconnect mid-run and the twin-to-rig comparison of case 10 are not run. The procedure for each is recorded with the application, and no hardware number appears in this chapter.

5.8 Technology in Practice

Figure 5.5 draws the loop of Section 5.3 with the devices of the rig.

What Measures It

The ranger below the seat is an optical time-of-flight device, an echo instrument that times a pulse of light to the ball and back [18]. Equation (5.5) stands for it with a reading rounded to 1 mm, colored noise, a residual bias and one sample of latency, and a FLOOR flag marks a raw range under 45 mm. Echo instruments can be misled by structures near the beam, and the Assumptions note that the ranger's cone always sees the tube wall, which the model leaves out beyond a constant bias. No range, accuracy or response time is given here, because none was found in an open source that names no maker.

> **TECHNOLOGY** **Sensor: An Optical Time-of-Flight Ranger**
>
> An echo instrument times a wave's trip to a surface and back. Pulsed laser light does this for distances between objects, and mass-produced precision electronics have made it affordable. The target must reflect enough light, dust or vapor in the path weakens the echo, structures near the path can return false echoes, and a target too close to the head falls in a dead zone [18]. The reflective infrared ranger of one levitation rig could be affected by the color, shape and motion of the ball, and reflections from the tube folded its calibration curve back on itself until colored strips were fitted inside the tube [2].

What Moves It

The blower runs at full power and a flap throttles it, so the final element is a small damper on an electric position actuator, not a fan whose speed changes. Its opening $a(\theta)$ in Equation (5.3) seats at 8° with 1 % leakage, the tight shutoff a butterfly disk lacks without a soft seat [18]. Equation (5.4) gives the servo its lag, its rate limit and its dead band, and on a lost command stream the firmware drives the flap to its closed angle (Section 5.5).

> **TECHNOLOGY** **Final Element: A Flap Turned by a Position Servo**
>
> A butterfly valve is a disk that turns across the flow: edge-on it barely obstructs, square-on it blocks, and a tight shutoff is hard to obtain without a soft seat. Dampers, with several vanes moved together, throttle large flows of air at low pressure in furnace and boiler draft and in building ventilation. Electric actuators reduce a fast motor through gears, and one that throttles needs a sensor of its own position, such as a potentiometer, a variable differential transformer or a pulse encoder [18]. A hobby servo is a small actuator of this kind, and its gear backlash and pulse dead band are the δ of Equation (5.4).

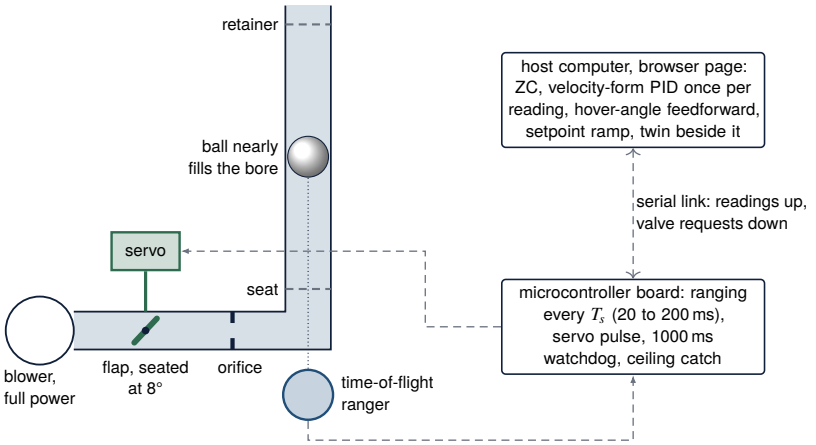


Figure 5.5 The height loop on the rig. The ranger below the seat times light to the ball and back, each reading goes from the board to the controller in the browser page, which returns one valve request, and the servo turns the flap between the full-power blower and the orifice. Notice that nothing in the uniform tube pulls a drifting ball back: the controller's only handle is the share of the blower's air that the flap and the orifice let through.

Where the Controller Runs

The control law runs in a browser page on a host computer. The microcontroller board below the tube times the ranging, sets the servo pulse and keeps its own watchdog and ceiling catch, and each reading it streams over the serial link runs the law once and returns one valve request. That sequence is the cycle time of a standards glossary, in which sensor signals are read into memory, the algorithm is executed and the actuator signals are sent [19]. Plant loops run with cycle times from milliseconds to minutes [19]. Here the cycle is the sensor's timing budget, 20 to 200 ms, while the engine steps every 1 ms, and case 7 of Section 5.6 shows how much the budget alone changes. In a plant the velocity-form PID of Equation (5.8), with its lower limit at the seat, would be the PID function block of a programmable controller or a distributed control system.

> **IN PRACTICE** **Judge a Valve by Its Installed Curve**
>
> A valve tested in a laboratory with a constant pressure drop shows its inherent characteristic. Installed, it sees less pressure drop as the flow rises, so its curve droops: steep near closed and flat near open. A loop on such a valve over-reacts at low flows, where it may oscillate, and under-reacts at high flows. The pressure drop that causes this generally cannot be removed, so the remedy is chosen in the hardware, for example a trim whose inherent curve complements the droop [18]. The orifice of this rig is a different hardware remedy for the same steep start.

What Is Modernizing It

Three developments reach this rig. The first is cheap optical ranging: mass-produced precision electronics made laser distance measurement affordable enough that consumers could buy laser tape measures in 2008 [18]. The second is open hardware: a levitation rig built as a compact shield for a microcontroller board from off-the-shelf parts and 3D-printed parts, with its design files and software published for anyone to replicate [3]. The third is the digital twin, which joins sensor data to physics-based or data-based models. About 40 % of the 31 studies in one review of the process industry combined the two kinds of model [20]. Running a twin beside its own rig under one controller is the subject of the open challenge after Section 5.9.

5.9 Limitations

Every number here is a prediction about a rig, not a measurement of one, because the parameter table is nominal until the hover map, the drop test and the step tests replace it. The lift law is a drag law with a fitted coefficient, not a derivation of the annular flow, and the ball has one degree of freedom with no spin or contact with the wall. The fan curve is a placeholder parabola at constant speed, and the flap's opening has a fixed form fitted at two points. The sensor model omits ambient light, the wall seen inside the cone beyond a constant bias and noise that changes with range, and the serial and pulse-width readers the firmware also carries are untested against a real module. The wall friction has a regularized sign, so a ball under it creeps rather than sticks outright. Every threshold is a teaching value, the ± 5 mm settled band, the challenge's ± 10 mm, 3 s windows and 2000° among them. Web Serial confines the hardware session to Chromium-based browsers on a desktop.

> **OPEN CHALLENGE** **A Twin and a Rig With No Settled Norms**
>
> A survey of control education reports that turning a classic control experiment into an interactive web laboratory is not yet an easy task, nor an art with established norms. Virtual and remote laboratories are commonly accepted to work better when offered together, yet a simulation cannot fully replace real equipment in control engineering, where the behavior of the real elements is decisive [21]. Progress would be a twin and rig pair whose agreement is measured on the bench and whose effect on learning is measured in a course.

5.10 Conclusion

The twin of this chapter floats a ball at a diameter ratio of 0.902 on 0.382 L/s of air throttled by a servo flap, and it measures what that geometry does to a loop. The height integrates the air's mismatch with a 38.4 ms lag and a gain of 16.88 mm/s per degree, so in 323 of 364 switches to Manual the ball reached a stop within a minute. An orifice in series widens the valve's usable band from 1.19 to 22.34° and caps the open valve at 1.17 times the hover speed. The SIMC tuning overshoots a 50 mm step by 19.9 mm, a ball 20 % heavier works on a valve 2.92 times flatter, a 200 ms sensor loses the ball that a 66 ms one holds, a 1° dead band produces a tuning-dependent limit cycle (± 2.49 mm at the shipped tuning, between ± 1.52 and ± 6.35 mm on retuning), and a 50 mm/s ramp holds a program's overshoots to 13.8 mm where steps reach 70 mm. None of these numbers has met the hardware yet. What remains is a weighed ball and a measured blower, the hover map, drop test and step tests on the rig, a refitted table, the twin-to-rig comparison of case 10, and a lift law tested at exponents between 1 and 2. The rig is a desktop integrator whose twin and hardware run under one controller.

KEY POINTS

- No fixed valve angle holds the ball: left in Manual at the controller's last command, it reached its seat or the retainer within a minute in 323 of 364 switches.
- The height integrates the air's mismatch with a 38.4 ms lag and a gain of 16.88 mm/s per degree at hover.
- An orifice in series widens the valve's usable band from 1.19 to 22.34° and caps the open valve at 1.17 times the hover speed.
- A 200 ms sensor loses the ball that a 66 ms one holds, and a 1° dead band produces a limit cycle that retuning only moves.

Where else this applies. Volume 2, Chapter 5 tunes an integrating level loop from an identified slope, Volume 11, Chapter 14 holds a header pressure with no self-regulation, and Volume 10, Chapter 6 meets a gas valve with no authority left when its header pressure rises.

References

[1] S. R. Jernigan, Y. Fahmy, and G. D. Buckner, "Implementing a remote laboratory experience into a joint engineering degree program: Aerodynamic levitation of a beach ball," *IEEE Transactions on Education*, vol. 52, no. 2, pp. 205–213, 2009.

[2] J. Chacón, J. Sáenz, L. de la Torre, J. M. Díaz, and F. Esquembre, "Design of a low-cost air levitation system for teaching control engineering," *Sensors*, vol. 17, no. 10, p. 2321, 2017.

[3] G. Takács, P. Chmurčiak, M. Gulan, E. Mikuláš, J. Kulhánek, G. Penzinger, M. Vdoleček, M. Podbielančík, M. Lučan, P. Šálka, and D. Šroba, "FloatShield: An open source air levitation device for control engineering education," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 17 288–17 295, 2020.

[4] B. D. Tyreus and W. L. Luyben, "Tuning PI controllers for integrator/dead time processes," *Industrial & Engineering Chemistry Research*, vol. 31, no. 11, pp. 2625–2628, 1992.

[5] S. Skogestad, "Simple analytic rules for model reduction and PID controller tuning," *Journal of Process Control*, vol. 13, no. 4, pp. 291–309, 2003.

[6] M. A. A. S. Choudhury, N. F. Thornhill, and S. L. Shah, "Modelling valve stiction," *Control Engineering Practice*, vol. 13, no. 5, pp. 641–658, 2005.

[7] J. Ma and J. V. Nickerson, "Hands-on, simulated, and remote laboratories: A comparative literature review," *ACM Computing Surveys*, vol. 38, no. 3, p. 7, 2006.

[8] M. Rábek and K. Žáková, "Building of the fan driven ball levitation system," *IFAC-PapersOnLine*, vol. 52, no. 27, pp. 283–287, 2019.

[9] A. Tootchi, S. Amirkhani, and A. Chaibakhsh, "Modeling and control of an air levitation ball and pipe laboratory setup," in *2019 7th International Conference on Robotics and Mechatronics (ICRoM)*, 2019, pp. 29–34.

[10] A. F. Caldeira, C. Rech, C. F. Maidana, S. F. Venturini, and A. C. d. O. Pedra, "Feedback control applied to a low cost ball-in-tube air levitation plant," *Ciência e Natura*, vol. 45, no. esp. 1, p. e84132, 2023.

[11] G. Takács, E. Mikuláš, M. Gulan, A. Vargová, and J. Boldocký, "AutomationShield: An open-source hardware and software initiative for control engineering education," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 9594–9599, 2023.

[12] 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.

[13] T. Tkáčik, M. Tkáčik, S. Jadlovská, and A. Jadlovská, "Design of aerodynamic ball levitation laboratory plant," *Processes*, vol. 9, no. 11, p. 1950, 2021.

[14] A. R. L. Zachi, J. A. Gouvêa, C. A. M. Correia, M. F. Pinto, M. R. Petraglia, D. B. Haddad, and F. A. A. Andrade, "Analysis and control of ball displacement in air levitation experiment by using a disturbance rejection-based technique," *IEEE Access*, vol. 13, pp. 71 530–71 550, 2025.

[15] C. P. Estrada and W. Sheng, "Modeling and control of a fan-driven ball levitation system," in *2025 IEEE 21st International Conference on Automation Science and Engineering (CASE)*, 2025, pp. 2005–2010.

[16] P. Timmerman and J. P. van der Weele, "On the rise and fall of a ball with linear or quadratic drag," *American Journal of Physics*, vol. 67, no. 6, pp. 538–546, 1999.

[17] R. Clift, J. R. Grace, and M. E. Weber, *Bubbles, Drops, and Particles*. Academic Press, 1978.

[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, Gaithersburg, MD, Special Publication 800-82r3, 2023.

[20] M. Mayr, G. C. Chasparis, and J. Küng, "Learning paradigms and modelling methodologies for digital twins in process industry," arXiv preprint 2407.02275, 2024.

[21] J. A. Rossiter, B. Pasik-Duncan, S. Dormido, L. Vlacic, B. Jones, and R. Murray, "A survey of good practice in control education," *European Journal of Engineering Education*, vol. 43, no. 6, pp. 801–823, 2018.

5.11 Exercises

5.1 Concept check · 2 min

Case 1 hovers the ball in Auto with 3 mm of sensor noise. The reader switches to Manual, and the valve stays at the controller's last command. In most trials the ball reaches its seat or the retainer within a minute. Which statement explains why?

(a) The servo's dead band leaves the flap up to a degree from the command, and the ball drifts at the speed that error gives

(b) The air speed in the uniform tube does not depend on the ball's height, so a valve off the hover angle drifts the ball and nothing pulls it back

(c) The air jet slows as the ball rises, so a ball above its hover height falls back until it comes to rest on the seat

(d) Manual mode switches off the hover-angle feedforward, so the flap closes by the 23.96° that the bias was adding

5.2 Concept check · 1 min

Case 9 takes out the orifice plate that sits in series with the flap. What does the plate do on the rig that has it?

(a) It narrows the usable band, because it takes pressure that the flap would otherwise use to set the flow

(b) It leaves the band as it is and only moves the hover angle further from the seat, to 23.96°

(c) It widens the usable band from 1.19 to 22.34° and lowers the gain at hover from 92.5 to 16.88 mm/s per degree

(d) It widens the band by raising the air speed of the fully open valve from 1.17 to 7.08 times the hover speed

5.3 Concept check · 2 min

Case 8 gives the servo a 1° dead band, and the ball cycles ±2.49 mm every 1.06 s. An engineer halves the proportional gain to calm the loop. What does the run show?

(a) The cycle halves to about ±1.2 mm, since a gentler loop moves the valve less at every reading

(b) The cycle stops, since the integral alone walks the command slowly to an angle inside the dead band

(c) The cycle is unchanged, since its amplitude is set by the dead band and not by the tuning

(d) The cycle grows to ±6.35 mm every 4.6 s, since the error must double before the command crosses the dead band

5.4 Theory and calculation · 5 min

The nominal ball of Table 5.2 has $m_0 = 0.30\,\mathrm{g}$, $d = 22.9\,\mathrm{mm}$, $C_{eff} = 24$ and $n = 2$, in air of density $1.05\,\mathrm{kg/m^3}$ inside a tube of bore 25.4 mm, with $g = 9.81\,\mathrm{m/s^2}$.

(a) Find A_b, A_t, the hover speed u_{eq} of Equation (5.2) and the air flow that holds the ball. *Checkpoint:* 411.9 mm^2 and 506.7 mm^2, 0.753 m/s, 0.382 L/s.

(b) Find the lag τ_v and the pressure mg/A_b under the ball. Repeat for a ball 20 % heavier, with its new hover speed. *Checkpoint:* 38.4 ms and 7.15 Pa, then 0.825 m/s, 42.0 ms and 8.57 Pa.

(c) The integrating gain at hover is $K_i = 16.88\,\mathrm{mm/s}$ per degree. At what speed does a valve 0.5° above the hover angle lift the ball, and how long does the ball take from 200 mm to the retainer at 370 mm? *Checkpoint:* 8.4 mm/s and 20.1 s.

5.5 Theory and calculation · 8 min

Use the network of Equation (5.3) with the nominal values of Table 5.2: $P_0 = 300\,\mathrm{Pa}$, $Q_0 = 3.0\,\mathrm{L/s}$, $K_f = 1.44 \times 10^9$ and $K_t = 2 \times 10^6$ (both in Pa s^2/m^6), $C_c = 0.65$, $A_d = 507\,\mathrm{mm}^2$, $\rho = 1.05\,\mathrm{kg/m}^3$, $\theta_0 = 8°$, $\theta_1 = 90°$, $p = 1.5$ and $a_{min} = 0.01$. The ball hovers on $Q = 0.3816\,\mathrm{L/s}$ and holds up $\Delta P_b = 7.15\,\mathrm{Pa}$.

(a) Find the blower's pressure at the hover flow and the pressure drops across the orifice and the tube. *Checkpoint:* 295.1 Pa, 209.7 Pa and 0.29 Pa.

(b) Find the flap's pressure drop and its share of the blower's pressure, with the orifice and without it. *Checkpoint:* 78.0 Pa and 26 %, then 287.7 Pa and 97.5 %.

(c) Find the opening a and the hover angle in both cases. *Checkpoint:* 0.0950 and 23.96°, then 0.0495 and 17.57°.

5.6 Theory and calculation · 5 min

The tuning of Section 5.4 uses the half rule and the SIMC rules of Equation (5.7), with $\tau_a = 25\,\mathrm{ms}$, $\tau_q = 30\,\mathrm{ms}$ and $\tau_v = 38.4\,\mathrm{ms}$.

(a) Find θ_d, θ', K_c and τ_I at the 33 ms budget with $K_i = 16.88\,\mathrm{mm/s}$ per degree. *Checkpoint:* 88 ms, 0.107 s, 0.276 °/mm and 0.858 s.

(b) Repeat at the 200 ms budget. *Checkpoint:* 0.274 s, 0.108 °/mm and 2.19 s.

(c) A ball 20 % heavier hovers at 31.14°, where $K_i = 5.78\,\mathrm{mm/s}$ per degree and $\tau_v = 42.0\,\mathrm{ms}$. Find K_c at 33 ms and its ratio to the shipped 0.28 °/mm. Case 6 runs three times the shipped gain: what does it find on its step down, and why? *Checkpoint:* 0.79 °/mm, 2.8 times. An overshoot of 25.5 mm.

5.7 App challenge · 9 min

Load case 2, Drop test and step tests: the integrating gain. It hovers the ball in Manual at 24.0° with the noise off. Use the default seed and 1× speed.

(a) Press Drop test and then Run. When the fit lands, about 2 s later, read the Fall speed tile. Add the seated leakage speed of 0.151 m/s to find u_{eq}, and find C_{eff} from u_{eq} at the declared 0.30 g. *Checkpoint:* 592 mm/s within 2 mm/s, $u_{eq} = 0.744\,\mathrm{m/s}$ and $C_{eff} = 24.6$, each within 0.1.

(b) Press the reload button, press Run, and press Step test +2° between 1.5 and 3 s. Read the Fitted K_i tile when it appears, about 5 s later. Repeat from a fresh reload with Step test −2°. *Checkpoint:* 14.3 to 14.5, then 19.6 to 19.8 mm/s per degree.

(c) Compare the mean of the two fits with the live K_i that Show Model lists, and say why the two fits differ. *Checkpoint:* A mean of 17.0 against 16.88, within 1 %.

5.8 App challenge · 8 min

Load case 3, Proportional only: no offset to a setpoint, an offset under load. It runs $K_c = 0.28\,°/\mathrm{mm}$ with $\tau_D = 0.04\,\mathrm{s}$ and no integral, steps the setpoint to 250 mm at 5 s and adds 20 % to the ball's mass at 20 s, with the noise off. Run to 40 s at the default seed and export the CSV file.

(a) Read the true height, column `height_mm`, at 19 and 40 s. *Checkpoint:* 250.4 mm and 224.55 mm, each within 0.2 mm.

(b) Read `hover_deg` and `bias_deg` at 40 s. Predict the error the proportional term needs to hold the heavier ball, and compare it with (a). *Checkpoint:* 31.14° and 23.96°, an error of 25.7 mm against 25.45 mm.

(c) Press Reset, set Integral time to 0.9 s and run to 40 s again. Find the lowest true height after the mass step and read the Since the added mass tile. *Checkpoint:* 230.15 mm, a dip of 19.9 mm, settled in 1.98 s within 0.05 s.

5.9 App challenge · 8 min

Load case 7, The sensor's bargain: 20, 33, 66 and 200 ms. It steps the setpoint 50 mm at 5 s and taps the tube at 15 s, with 3 mm of noise at the 33 ms budget, under the suggested tuning. Use the default seed and 2× speed. Run to 30 s, read the IAE tile and export the CSV file. Press Reset, set Sensor timing budget (Advanced) to 66 ms and repeat. Repeat at 200 ms, and finally at 200 ms with Proportional gain 0.17 °/mm.

(a) At 33 and 66 ms, give the IAE at 30 s and the valve travel, column `valve_travel_deg` at 30 s. *Checkpoint:* 76 and 72 mm s, each within 2, and 1578 and 361°, each within 5°.

(b) At 200 ms with the shipped gain, give the IAE, the stop contacts in column `top_contacts` and the travel at 30 s, and explain them with θ'. *Checkpoint:* About 1880 mm s within 5 %, 6 contacts, 1103° within 5°.

(c) At 200 ms and 0.17 °/mm, give the IAE and the travel at 30 s. *Checkpoint:* 174 mm s within 2 and 89.5° within 2°.

5.10 Application area · 3 min

A control valve on a pumped line passes its design flow only a few percent open, because the pump makes far more head than the line needs. Case 9 is the same situation on the rig.

(a) Name two consequences for the loop that case 9 measures. *Checkpoint:* A band about one dead band wide, and a gain that bangs the valve between its stops.

(b) Name the fix the chapter applies and what it costs. *Checkpoint:* A fixed restriction in series. It burns pressure and caps the largest flow.

5.11 Application area · 4 min

A rig is built to the nominal table, and its drop test fits a fall of 0.48 m/s. The acceptance limits of Section 5.7 allow the drop speed of rig and twin to differ by 15 %.

(a) Is the fall within 15 % of the nominal twin's 0.592 m/s? What u_{eq} and C_{eff} does the helper report at the declared 0.30 g, with the seated leakage of 0.151 m/s? *Checkpoint:* No, 19 % slow. $u_{eq} = 0.631\,\mathrm{m/s}$ and $C_{eff} = 34.2$.

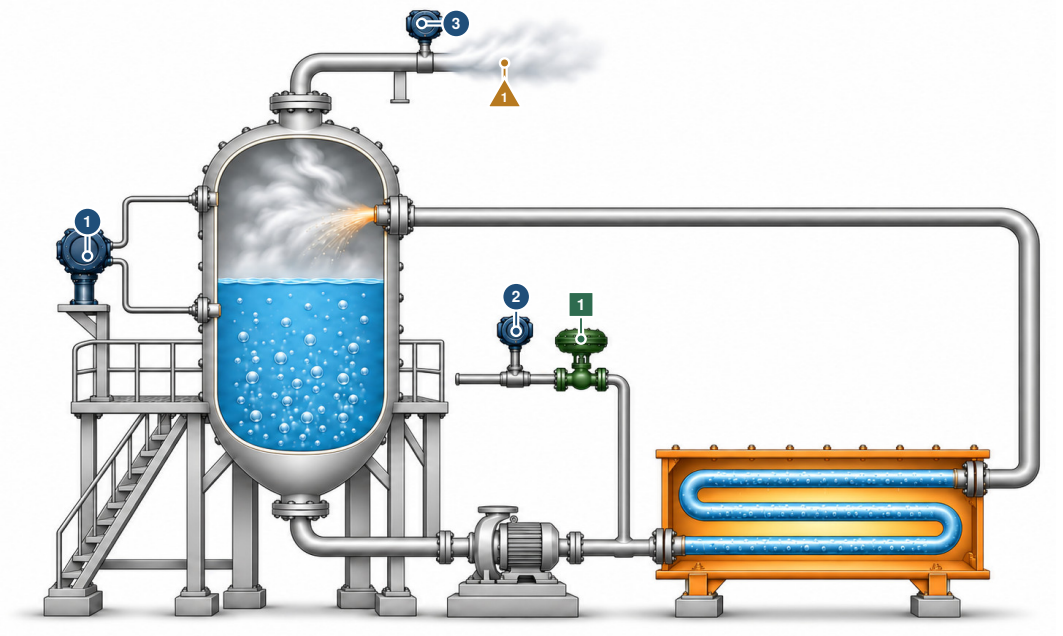
(b) The ball is then weighed at 0.36 g. What C_{eff} does the same fall imply, and what does a drop test identify on its own? *Checkpoint:* 41.0. Only u_{eq}, the ratio of the mass to C_{eff}.

5.12 Review the agent · 5 min

An AI assistant reads the results of case 9 and writes: "The orifice plate only rescales the gain. Without it K_i at hover is 5.48 times larger, and case 9 shows that 0.05 °/mm, about 16.88/92.5 of the shipped gain, holds the bare rig within 1.3 mm. Remove the plate, save the 209.7 Pa it burns, and divide the gain by 5.5." Find the flaw and say how you would check the claim in the app.

CHAPTER 6

A Flash Steam Vessel Whose Level Rises While It Loses Water



The controller moves 1 feedwater valve

The controller reads 1 level of the surface 2 feedwater flow 3 steam flow

What disturbs it 1 more steam taken by the users

Notice the bubbles below the surface, which lift the level the transmitter reads, although in the model the coil water stays liquid and the steam goes to a plant header rather than to the air. Illustration, not to scale.

IN PLAIN TERMS A Steam Vessel Whose Level Can Mislead

The main idea. Hot water under pressure is let down into a tall tank at a lower pressure, and part of it boils into steam at once. The rest stays as water, and a pump sends it back through a coil of pipes in hot exhaust gas to pick up more heat.

The challenge. When the users take more steam, the pressure in the tank falls and water boils all through it. The bubbles lift the surface, so the level reads higher just as the tank is losing water.

Why it matters. A controller that trusts the level alone cuts the water supply at the worst moment.

THREE WAYS TO READ THIS CHAPTER

Explore	11 min	the first page, the history of the steam tables (Section 6.2), the fact on fast property calculation (Section 6.3), the try-it segments of cases 2 and 3 (Section 6.4.2), the fact on the swell peak (Section 6.6) and the key points
Study	50 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	15 min	the first page, the results (Section 6.6), Technology in Practice (Section 6.7), Limitations (Section 6.8) and the open challenge

The control question A rise in steam demand makes the level of a flash vessel rise while the vessel loses water. Which disturbances move the indicated level against the inventory in a heat-recovery unit, and what can single-element, three-element and predictive feedwater control do about it?

Learning objectives

- Write the water and energy balances of the vessel and its loop, and explain why the circulation cancels from the water balance.
- Compute the integrating gain of the inventory, the flash quality and the pressure settling time at the design point.
- Explain swell and shrink as an inverse response through pressure, and say why colder feedwater does not shrink this vessel.
- Compare single-element, three-element and predictive control on a demand step by error, inventory dip and valve travel.
- Judge which plant changes an inventory estimator can follow and which ones, such as foaming or a transmitter bias, it cannot.

Prerequisites Sections 1.4.2 (balances), 1.4.4 (integrating response), 1.4.7 (inverse response) and 1.6.2 (velocity form), and Chapter 2 for a tank level loop. Section 1.6.3 is needed only for the estimator and the predictive controller.

The app https://apmonitor.com/apps/flash-steam/

See also Volume 3, Chapter 2 for flow feedforward and a level-to-flow cascade, Volume 3, Chapter 9 for an inverse response met with an inner loop. Volume 5, Chapter 12 for an observer of an inventory no instrument measures, and Volume 5, Chapter 11 for a soft sensor of a state no instrument reaches.

6.1 Introduction

A flash vessel makes steam without a burner. Water pumped through a heat-recovery coil at 18 bar comes back at 190.3 °C, well above the 170.4 °C at which water boils in a vessel held at 8 bar absolute, and about 4.3 % of it flashes to vapor as an orifice drops it to that pressure. In the generic unit simulated here, 39 kg/s of circulating water recovers 4.08 MW from 18 kg/s of flue gas at 450 °C and delivers 1.75 kg/s of saturated steam to a plant header, while feedwater from a deaerator replaces what leaves. Opening the steam user's valve by a quarter raises the indicated level from 55.0 to 58.0 % of the vessel height in ten seconds, while the water in the vessel falls from 51.90 to 51.78 % of the height as a mass-equivalent level.

Steam recovered from waste heat is one of the larger levers on industrial energy use, and the equipment that recovers it takes many forms, among them recuperators, waste heat boilers and heat-recovery steam generators [1]. Flash separation of a pressurized hot-water stream is the working principle of the single-flash geothermal power plant [2], and the flash evaporation of a liquid pool after a pressure drop depends on how fast the pressure falls and how deep the pool is [3]. The level of any vessel in which vapor forms inside the liquid swells during a depressurization, a transient that nuclear safety analysis treats with drift-flux models [4], and the drum of a natural-circulation boiler shows the same inverse response that the drum-boiler model of Åström and Bell carries as an explicit state of steam below the surface [5]. Level control of such vessels is a long-standing problem of boiler and heat-recovery practice [6–8], and an integrating process with inverse response needs identification and tuning rules of its own [9].

The simulator is built to correct a specific belief, that a level transmitter measures how much water a vessel holds. In this model the transmitter reports the surface above the liquid and submerged vapor. Pressure-driven changes in that vapor can move the indicated level opposite to the liquid inventory. A single-element controller that trusts the level reads a swell as too much water. On the 20 % demand step of case 5 it closes the feedwater valve from 43.7 % to 34.0 % within fourteen seconds, at the moment the vessel is losing water. The same plant also contradicts a belief carried over from drum boilers. The tested cold-feedwater step produces no initial shrink in this model: feedwater mixes with

circulating water upstream of the coil, rather than entering the vessel directly. This result depends on the loop holdup, heat supply and bubble closure.

The usual sources treat the parts of this problem apart. Drum-boiler models are written for natural circulation and a firing rate that the plant controls [5], heat-recovery level control is reported for large multi-pressure units [7, 8], and flash evaporation is studied as heat and mass transfer without a level loop [3]. This chapter works with a browser simulator of a forced-circulation flash unit in which the indicated level, the water inventory and the vapor below the surface are logged side by side. It measures how single-element, three-element and predictive control, with and without an inventory estimate, deal with the gap between them. The balances, the IAPWS-IF97 properties [10], the velocity-form PID, the Kalman filter [11] and the offset-free MPC [12] are established methods. The results are measurements on one plant and seed, and the central one is that in a heat-recovery unit both the swell and the shrink come from the steam side through pressure.

Section 6.2 follows the domain from flash tanks and geothermal separators to drum-level control. Section 6.3 sets out the unit as a ten-state model whose mass and energy ledgers close to the integrator's precision, with properties from an IAPWS-IF97 module checked against the release's verification values and two independent implementations, and it states every equation, constant and default needed to build the simulator again. Section 6.4 reads the dynamics from the model: an integrating inventory, and in Section 6.4.2 the swell and shrink that make this plant the volume's example of an inverse response. Section 6.5 describes the controllers and the measurement boundary that keeps every one of them away from the simulator's truth. Section 6.6 measures the fifteen cases, from a hand-closed water balance to a graded challenge, Section 6.7 names the instruments, valves and platform of a real unit, and Sections 6.8 and 6.9 state the limits of the model and what the cases show.

6.2 Flash Steam, Recovered Heat and the Level Problem

When a pressurized liquid at its boiling point is let down, part of it flashes, and recovering that vapor is the basis of flash tanks on boiler blowdown and condensate returns and of the separators of geothermal plants. In a single-flash geothermal plant a separator divides the two-phase flow from the wells into steam for the turbine and liquid that is returned to the reservoir, and the separator pressure is chosen to balance the steam fraction against the energy each kilogram of steam carries [2]. The same trade appears in the unit simulated here, where the loop pressure keeps the hot return liquid and the vessel pressure decides how much of it flashes.

Industrial waste heat became a design target in its own right as energy prices and emission limits rose. A review of recovery technologies surveys them by the temperature of the source, from recuperators and regenerators to waste heat boilers that raise steam from exhaust and flue gas [1]. A forced-circulation loop with a separate flash vessel suits a source that is variable or dirty, because the coil can remain liquid-filled at elevated pressure while vapor separates in the downstream vessel. The simulation assumes this operating regime and does not model isolation, drainage or bypass equipment.

The level problem came to these units from boilers. Feedwater regulation on boiler drums began with a single measurement of level and moved to two-element control with a steam-flow feedforward and then to three-element control with an inner feedwater-flow loop, which became the standard arrangement on industrial and utility boilers because the drum level swells and shrinks with load [6]. Model-based designs followed. A cascade generalized predictive controller for drum level was reported in 2005 [13], and heat-recovery steam generators, whose load follows a gas turbine or a process rather than a burner, have since been given robust drum level controllers for wide-range operation [7] and three-element schemes with fractional-order PID and fuzzy elements [8]. The physical understanding underneath these designs rests on models in which the steam held below the surface is a state of its own [5], and on two-phase level swell during depressurization [4]. The flash unit simulated here adds two features of the modeled arrangement: its tested feed-temperature

Table 6.1 Variables by Role, With Their Values at the Design Point of Case 1

Variable	Role	Design	Range
L_m	CV, measured surface	55.0 %	SP 40–65 %
v	MV, feedwater valve	43.75 %	0–100 %
$\dot{m}_{f,SP}$	MV of the outer loops	1.75 kg/s	0–4.0 kg/s
d	DV, steam demand	100 % of design	20–150 %
$\dot{m}_g$	DV, flue-gas flow	18 kg/s	0–130% design
T_g	DV, gas temperature	450 °C	300–550 °C
T_{fw}	DV, feedwater	105 °C	20–140 °C
Δp_f, a_{line}	DV, feed delivery	100, 100 %	20–130, 10–100
ϕ	DV, water-side fouling	1	0.4–1
$\dot{m}_b$	DV, blowdown	0	0–0.2 kg/s
σ_L, b_L	DV, transmitter noise, bias	0.2 %, 0	0–1.5, ±6
M, U	states, vessel	1269.0 kg, 1.0943 GJ	—
p	algebraic, pressure	8.00 bar	p_{hdr}−12.5
h_h, T_{hot}	state, loop	809.1 kJ/kg, 190.3 °C	subcooled ≥ 3 K
T_m	state, coil metal	200.5 °C	—
β	state, submerged share	0.0646	0–0.35
$\dot{m}_r$	state, circulation	39 kg/s	≥ 45 %
$\dot{m}_s$	output, steam	1.75 kg/s	—

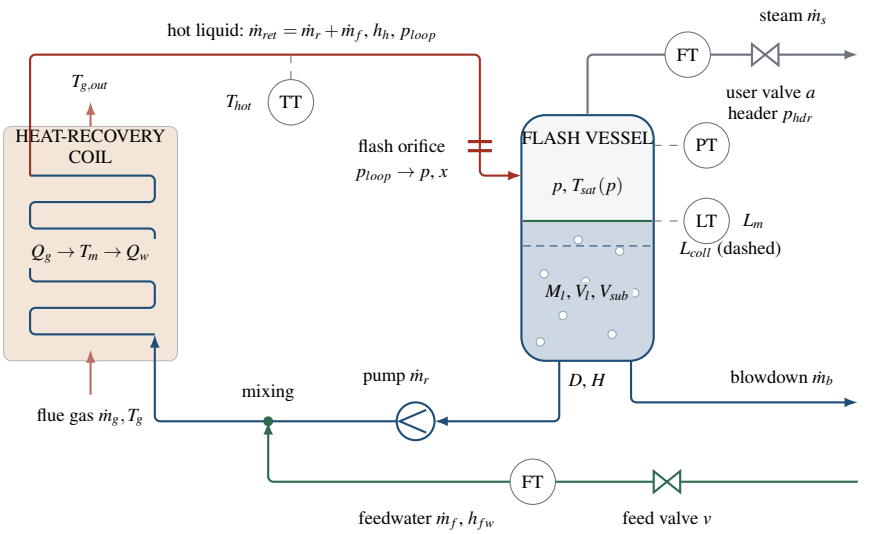
disturbance does not cause an initial shrink, and its available upstream heat constrains sustained steam production.

> ⧖ HISTORY **Steam Tables Become Equations, 1929 to 2003**
>
> International agreement on the properties of water and steam began with steam-table conferences in 1929, 1930 and 1934, and the 1934 meeting agreed on skeleton tables of pressure, enthalpy and volume on a coarse grid. After the Second World War new power plants were designed for higher temperatures and pressures, and new skeleton tables approved in 1963 reached 100 MPa and 800 °C. As computers spread through industry, equations took the place of tables. An industrial formulation adopted in 1966 and revised in 1967 served for three decades, until IAPWS-IF97 replaced it in 1997 [14]. A familiar book of steam tables was first printed from the 1967 equations [15]. The last skeleton tables were withdrawn in 2003, and the recommendations have been equations ever since [14].

6.3 The Unit as Modeled

The simulated unit is a generic industrial design with no proprietary source, as shown in Figure 6.1. Flue gas passes a finned-tube coil whose metal is one node at T_m. A circulation pump draws saturated water from the bottom of a vertical flash vessel 1.2 m across and 2.4 m tall (2.71 m^3), raises it to the loop pressure p_{loop}, and sends it through the coil, where it is lumped as one well-mixed holdup M_h at enthalpy h_h. Feedwater enters the loop at the pump discharge. An orifice drops the hot return to the vessel pressure p, steam leaves through the steam user's valve of opening a to a header at p_{hdr}, and a continuous blowdown $\dot{m}_b$ can take liquid out. The level transmitter LT reads the surface, the feedwater flow FT and a steam flow meter supply the three-element and predictive controllers, the thermowell TT reads the hot return, and PT reads the vessel pressure.

Each variable in Table 6.1 carries its role in the loop, controlled (CV), manipulated (MV), disturbance (DV) or state. Ten states are integrated. Eight are physical, the vessel's water mass M and internal energy U, the loop enthalpy h_h, the coil metal temperature T_m, the share β of the vessel's vapor mass held below the surface, the feedwater valve position v, the circulation flow $\dot{m}_r$ and the steam user's valve opening a. Two are ledgers that integrate the flows and energy crossing the plant boundary, so that conservation can be checked exactly. Units are seconds, kilograms, kg/s, cubic meters, bar absolute, °C, kJ/kg and kW.



What the level means

Surface (solid line)
$L_{surf} = L_{coll} + L_{sub}$

Collapsed liquid volume (dashed line)
$L_{coll} = 100 M_l v_l(p)/V$

Submerged vapor contribution
$L_{sub} = 100 V_{sub}/V$

Liquid inventory at fixed density
$L_{mass} = 100 M_l/(\rho_{ref} V)$

Density changes separate L_{coll} from L_{mass}.

$M = M_l + M_v$ includes *all* vessel water. Submerged vapor is part of M_v.

Dashed links: instrument taps.
FT: flow, TT: temperature,
PT: pressure, LT: level.

Figure 6.1 Forced-circulation flash steam unit as modeled. The coil at T_m receives Q_g from the gas and gives Q_w to the loop, which holds M_h at h_h and returns $\dot{m}_{ret}$ through the orifice with flash quality x. The vessel of diameter D and effective height H holds total water mass M and water-plus-shell energy U at pressure p, liquid M_l of volume V_l and submerged vapor V_{sub}, so the surface L_{surf} stands above the collapsed level L_{coll}. Feedwater $\dot{m}_f$ at h_{fw} passes the valve v and mixes at the pump discharge. The steam and feedwater FT measurements support three-element control. The panel below the drawing distinguishes volume-based levels from mass-equivalent inventory. Process model only, not a construction or safety piping diagram. Not to scale.

6.3.1 The Vessel and Its Saturation Closure

The vessel holds saturated liquid and vapor at one pressure in a fixed volume V, and its shell, of heat capacity C_d, stays at the saturation temperature, as shown in Equation (6.1).

$$\begin{aligned} M &= M_l + M_v, \qquad V = M_l\, v_l(p) + M_v\, v_v(p) \\ U &= M_l\, u_l(p) + M_v\, u_v(p) + C_d\, T_{sat}(p) \end{aligned} \tag{6.1}$$

where v_l, v_v, u_l and u_v are the specific volumes and internal energies of saturated liquid and vapor. Given M and U, the three equations fix p, M_l and M_v. Eliminating the masses leaves one equation in p, which is solved by Newton's method at every stage of the integrator with the exact slopes of the property interpolant. The iteration always starts at the design pressure, so that the pressure is a function of (M, U) alone and two runs that reach the same state agree to the last bit. It stops when the step falls below 10^{-12} of the pressure. The same solution supplies the two partial derivatives that turn

the balances into a pressure rate, $(\partial U/\partial p)_M$ and $(\partial U/\partial M)_p = (v_v u_l - v_l u_v)/(v_v - v_l)$. At design $(\partial U/\partial p)_M = 34.9\,\text{MJ/bar}$, 1263.6 kg of the water is liquid and 5.43 kg is vapor.

6.3.2 Water and Steam Properties

All properties come from the IAPWS Industrial Formulation 1997 [10], Region 1 for the liquid, Region 2 for the vapor, Region 4 for the saturation line in both directions, and the Region 1 backward equation $T(p,h)$ refined by one Newton step on the forward equation. The implementation reproduces the release's computer-program verification values to better than 2×10^{-9} relative. Between 1 and 20 bar its saturation temperature, specific volumes and enthalpies agree with the IAPWS-IF97 implementations of the *iapws* package and of CoolProp [16] to 4×10^{-15}, and with the scientific formulation IAPWS-95 [17] to 7×10^{-5}, which is the consistency IAPWS-IF97 is designed to have. Because the closure is solved several hundred thousand times per simulated hour, the saturation properties are read from a cubic Hermite interpolant built from the equations at load, with nodes every 0.025 bar from 0.5 to 25 bar. Its largest relative error against the equations is 2.0×10^{-8} for the values and 2.7×10^{-6} for the slopes.

> **+ DID YOU KNOW? A Steam Table Called Millions of Times**
>
> Steam properties are often called millions of times within design calculations, so the steam power industry needs a formulation built for speed [15]. IAPWS-IF97 computes the industrially most important properties on average more than five times faster than the 1967 formulation it replaced [18].

6.3.3 Mass and Energy Balances

The water in the vessel changes by feedwater minus steam minus blowdown, and its energy by the hot return minus the liquid drawn to the pump, the steam, the blowdown and a small loss, as shown in Equation (6.2). Each line is the accumulation balance of Equation (1.1) written for one inventory, as Section 1.4.2 does for a tank.

$$\begin{aligned} \frac{\mathrm{d}M}{\mathrm{d}t} &= \dot{m}_f - \dot{m}_s - \dot{m}_b \\ \frac{\mathrm{d}U}{\mathrm{d}t} &= \dot{m}_{ret} h_h - \dot{m}_r h_l - \dot{m}_s h_v - \dot{m}_b h_l - Q_{loss} \\ M_h \frac{\mathrm{d}h_h}{\mathrm{d}t} &= \dot{m}_r h_l + \dot{m}_f h_{fw} - \dot{m}_{ret} h_h + Q_w \\ C_m \frac{\mathrm{d}T_m}{\mathrm{d}t} &= Q_g - Q_w \end{aligned} \tag{6.2}$$

where $\dot{m}_{ret} = \dot{m}_r + \dot{m}_f$ because the loop holdup is constant, h_l and h_v are the saturated enthalpies at p, h_{fw} is the feedwater enthalpy at its temperature and the feed pump's discharge pressure, and $Q_{loss} = UA_{loss}(T_{sat} - 25\,^\circ\text{C})$, 7.3 kW at design. The circulation leaves and returns, so it cancels from the mass balance, and adding the three energy balances leaves feedwater enthalpy plus gas heat minus steam, blowdown and losses. Flashing is not written as a separate loss. The closure of Equation (6.1) decides how much of the returning energy becomes vapor. Over a 360 s run with blowdown and a demand step, the vessel mass differs from its ledger by less than 10^{-10} of the throughput and the stored energy from its ledger by less than 10^{-9}.

6.3.4 Heat Transfer on Both Sides of the Coil

The coil metal is heated by the gas through an effectiveness and gives heat to the loop water through a conductance that scales with the circulation, as shown in Equation (6.3).

$$\begin{aligned} Q_g &= \varepsilon\, \dot{m}_g c_{p,g} (T_g - T_m) \\ \varepsilon &= 1 - \exp\!\Big(-\frac{UA_g}{\dot{m}_g c_{p,g}}\Big) \\ UA_g &= UA_{g,0}\Big(\frac{\dot{m}_g}{\dot{m}_{g,0}}\Big)^{0.6} \\ Q_w &= UA_{w,0}\Big(\frac{\dot{m}_r}{\dot{m}_{r,0}}\Big)^{0.8} \phi\,(T_m - T_{hot}) \end{aligned} \tag{6.3}$$

where $T_{hot} = T(p_{loop}, h_h)$ and ϕ is the water-side fouling factor. The gas-side conductance is not a parameter but is sized so that the design steam flow comes out at the design pressure. At design the vessel balance fixes $h_{h,0} = 809.1\,\mathrm{kJ/kg}$, the loop balance gives $Q_{w,0} = 4078\,\mathrm{kW}$, the metal sits at $T_{m,0} = T_{hot,0} + Q_{w,0}/UA_{w,0} = 200.5\,^\circ\mathrm{C}$, the required effectiveness is 0.826, and therefore $UA_{g,0} = 34.6\,\mathrm{kW/K}$ with the gas leaving at 244.0 °C. The exponents 0.6 and 0.8 are the usual dependences of gas-side and water-side film coefficients on velocity.

6.3.5 Flashing at the Inlet and Boiling in the Bulk

The hot return carries more enthalpy than saturated liquid at the vessel pressure, and the excess over the latent heat is the equilibrium flash quality. A falling pressure also boils the liquid already in the vessel, because the liquid and the shell must give up the sensible heat that their lower saturation temperature no longer allows, as shown in Equation (6.4).

$$\begin{aligned} x &= \frac{h_h - h_l(p)}{h_v - h_l}, \qquad \dot{m}_{flash} = \dot{m}_{ret} \max(0, x) \\ \dot{m}_{bulk} &= -\frac{\big(M_l\, h_l'(p) + C_d\, T_{sat}'(p)\big)\,\dot{p}}{h_v - h_l} \\ \dot{p} &= \frac{\dot{U} - (\partial U/\partial M)_p \dot{M}}{(\partial U/\partial p)_M} \end{aligned} \tag{6.4}$$

where the primes denote derivatives along the saturation line. At design $x = 4.30\,\%$ and $\dot{m}_{flash} = 1.754\,\mathrm{kg/s}$. Neither rate enters the balances. Both are diagnostics that drive the bubble closure below, and $\dot{m}_{bulk}$ turns negative when the pressure rises, which is how the closure condenses bubbles.

6.3.6 Vapor Held Below the Surface

A share of the vapor generated stays in the liquid for a residence time, and the held share relaxes toward that, as shown in Equation (6.5).

$$\begin{aligned} \dot{m}_{sub} &= \max\big(0,\ \eta\, \dot{m}_{flash} + \eta_b\, \dot{m}_{bulk}\big) \\ \beta^* &= \min\big(\beta_{max},\ \tau_{res}\, \dot{m}_{sub}/M_v\big) \\ \tau_\beta \frac{\mathrm{d}\beta}{\mathrm{d}t} &= \beta^* - \beta, \qquad V_{sub} = \beta\,(V - V_l) \end{aligned} \tag{6.5}$$

where $V_l = M_l v_l$ and βM_v is the vapor mass below the surface. The submerged vapor is part of M_v, so it is counted once, and because $M_v v_v = V - V_l$ it never exceeds the vapor volume and the freeboard never becomes negative. Most of the inlet flash separates at a tangential inlet ($\eta = 0.1$), while vapor formed in the bulk must rise through the liquid ($\eta_b = 0.8$). The two shares are a synthetic teaching closure. With the inlet flash alone, at $\eta = 0.2$, a 25 % demand step swelled the level by only 0.8 points, and the bulk term raises the swell to 3.1 points while keeping the bubble contribution at design at 3.1 points.

6.3.7 Three Levels

Three levels are computed from the state, as shown in Equation (6.6).

$$\begin{aligned} L_{surf} &= 100\,\frac{V_l + V_{sub}}{V}, \qquad L_{coll} = 100\,\frac{V_l}{V} \\ L_{mass} &= 100\,\frac{M_l}{\rho_{ref} V}, \qquad L_{sub} = 100\,\frac{V_{sub}}{V} \end{aligned} \tag{6.6}$$

where $\rho_{ref} = 897.03\,\mathrm{kg/m^3}$ is the saturated liquid density at the design pressure. The simulated transmitter idealizes a surface-level instrument. Actual radar, float and differential-pressure readings depend on foam, phase density and instrument configuration. The collapsed level removes the bubbles but still moves with liquid density. The mass-equivalent level expresses liquid inventory at a fixed reference density. It changes with both boundary flows and internal evaporation or condensation, since $\dot{M}_l = \dot{M} - \dot{M}_v$. Total water mass M, rather than liquid mass M_l, exactly integrates feedwater minus steam and blowdown. Thus $L_{surf} = L_{coll} + L_{sub}$, while $L_{mass} + L_{sub}$ generally differs from the surface away from the design pressure. At design $L_{surf} = 55.00\,\%$, $L_{mass} = 51.89\,\%$ and the bubble contribution $L_{sub} = 100 V_{sub}/V$ is 3.11 points. One percentage point of mass-equivalent level is 24.3 kg of liquid water.

6.3.8 Valves, Pump and Steam User

The feedwater valve follows its target with a first-order lag and a stroking-rate limit, and its flow depends on the feed pump's pressure drop and the line's capacity, as shown in Equation (6.7).

$$\begin{aligned} \frac{\mathrm{d}v}{\mathrm{d}t} &= \mathrm{clip}\Big(\frac{v^* - v}{\tau_v},\ \pm r_v\Big) \\ \dot{m}_f &= \dot{m}_{f,max}\, a_{line}\, f(v)\sqrt{\Delta p_f} \\ \dot{m}_s &= K_s\, a\,\sqrt{\max(0, p^2 - \max(p_{hdr}, 0.55p)^2)} \\ \tau_r \frac{\mathrm{d}\dot{m}_r}{\mathrm{d}t} &= \dot{m}_{r,0}\, u_{pump} - \dot{m}_r, \qquad \tau_a \frac{\mathrm{d}a}{\mathrm{d}t} = a_0 d - a \end{aligned} \tag{6.7}$$

where $f(v) = v$ for the linear trim and $(R_v^v - 1)/(R_v - 1)$ for the equal-percentage option, Δp_f and a_{line} are fractions of design, d is the steam demand as a fraction of the design opening $a_0 = 0.6$, and the valve is blocked at its stops. The target v^* is the controller's command passed through a stiction band S, which holds the stem until the command has moved more than S from the last target and then lets it slip all the way. The compressible steam law chokes when the header falls below 0.55 of the vessel pressure, and $K_s = 0.625\,\mathrm{kg/(s\,bar)}$ is sized so that the design opening passes the design flow at the design pressure. Steam demand is a valve opening rather than a flow, so the pressure decides how much steam leaves.

6.3.9 Instruments

The level transmitter reads the surface through a chain, as shown in Equation (6.8).

$$\begin{aligned} y &= L_{surf} + k_{cal}(p - p_0) + b_L \\ \tau_{LT} \frac{\mathrm{d}y_f}{\mathrm{d}t} &= y - y_f \\ L_m(t) &= q\,\mathrm{round}\Big(\frac{y_f(t - \theta_L) + \sigma_L n(t - \theta_L)}{q}\Big) \end{aligned} \tag{6.8}$$

where $n(t)$ is a unit normal deviate drawn every 0.25 s, $q = 0.05\,\%$ is the quantization, θ_L is a communication delay, and the noise follows the lag so that σ_L is the noise a trend shows. A lost signal is reported as missing with a separate stale value, never as zero. The pressure, feedwater flow, steam

Table 6.2 Model Parameters. All Are Adjustable in the Parameters Panel Unless Marked Derived

Symbol	Meaning	Value	Unit
D, H	vessel diameter, effective height	1.2, 2.4	m
C_d	vessel shell heat capacity	1000	kJ/K
UA_{loss}	vessel loss conductance	0.05	kW/K
$\dot{m}_{s,0}$	design steam flow	1.75	kg/s
p_0	design vessel pressure	8	bar
a_0	steam valve opening at design	60	%
$T_{fw,0}$	design feedwater temperature	105	°C
p_{hdr}	header pressure	6.5	bar
p_{loop}	loop pressure	18	bar
$\dot{m}_{r,0}$, M_h	circulation, loop holdup	39, 2000	kg/s, kg
C_m	coil metal heat capacity	5000	kJ/K
$UA_{w,0}$	water-side conductance	400	kW/K
$\dot{m}_{g,0}$, $T_{g,0}$	design gas flow, temperature	18, 450	kg/s, °C
$c_{p,g}$	flue-gas heat capacity	1.10	kJ/(kg K)
$UA_{g,0}$	gas-side conductance (derived)	34.6	kW/K
K_s	steam valve coefficient (derived)	0.625	kg/(s bar)
η, η_b	submerged shares, inlet and bulk	0.1, 0.8	—
τ_{res}, τ_β	residence, redistribution time	2, 6	s
β_{max}	largest submerged share	0.35	—
$\dot{m}_{f,max}$	feedwater valve capacity	4.0	kg/s
Δp_{ref}	reference valve pressure drop	6	bar
R_v	equal-percentage rangeability (option)	30	—
τ_v, r_v	valve lag, stroking limit	2.5, 6	s, %/s
S	valve stiction band	0	%
τ_r, τ_a	circulation, steam valve response	3, 2	s
τ_{LT}, θ_L	level lag, delay	1.5, 0	s
q	level quantization	0.05	%
k_{cal}	level calibration error	0	%/bar
τ_{FT}, τ_{TT}	flow, thermowell lags	0.5, 5	s

flow and hot-return measurements have default lags of 1, 0.5, 1 and 5 s. The flow lags are τ_{FT} and $2\tau_{FT}$. Pressure noise has standard deviation 0.004 bar, and each flow meter has 0.004 kg/s noise. The hot-return signal has no added noise. Every execution draws the same four deviates whatever the noise settings, so runs that differ in one setting stay comparable sample by sample.

6.3.10 Parameters, Initialization and Numerics

Table 6.2 gives every parameter with its default. Every value is a generic teaching value for a typical industrial unit, chosen so that the design point is consistent rather than measured on a plant. Each case starts in the steady state of its own disturbances. A bisection on the vessel pressure finds where the heat the gas gives up equals the heat the water needs to make the steam the valve passes at that pressure, every other quantity follows from the balances, and the liquid is then placed so that the surface, bubbles included, sits on the setpoint. At 65 % load this gives 8.25 bar and 1.24 kg/s. The derivatives at the initial state are below 10^{-9} of each state's scale. The integrator is the classical fourth-order Runge-Kutta method with a step of 0.05 s, inside the engine of Section 1.6.1. Halving the step changes the peak level of case 5 by less than 10^{-4} points and its integrated error by 0.003 %. A simulated hour costs about 3 s of computing in a browser engine.

6.3.11 Assumptions

The vessel is in saturated equilibrium, so stratification, subcooled pockets and delayed nucleation are absent, and the model is least reliable for a very fast depressurization. The loop is one well-mixed holdup at a fixed pressure and the coil one metal node. The water residence time is $M_h/(\dot{m}_r + \dot{m}_f) = 49$ s at design, while the coil metal introduces an additional thermal state. These coupled dynamics are not a single 49 s lag or a distributed exchanger with transport delay. Pump work, pump curves and pipe hydraulics are omitted. The constant-pressure loop uses $M_h h_h$ as its lumped energy store

and neglects changes in $p_{loop}V_{loop}$. The bubble closure of Equation (6.5) is a synthetic teaching closure, not a prediction for a particular inlet device or water chemistry. The vessel has the constant cross-section of a vertical cylinder, so a horizontal vessel's changing area is not represented. The run stops rather than continue outside the model: when the vessel pressure leaves the range from the header pressure to 12.5 bar, the hot return comes within 3 K of boiling at the loop pressure, or the circulation falls below 45 % of design. These are conservative model-domain checks, not predictions of boiling onset or equipment damage.

6.4 Dynamic Character

6.4.1 An Integrator in the Water

Neglecting the small change in vapor mass, liquid inventory integrates the flow imbalance, as the pumped tank of Equation (1.9) does, with a gain of $100/(\rho_{ref}AH) = 0.041\,\%/\mathrm{s}$ per kg/s. One percent of valve passes 0.04 kg/s at the design pressure drop, so the single-element process gain is 0.001 64 %/s per percent of valve. In case 1 a four-point valve step adds 0.16 kg/s and the level climbs from 55.0 to 56.0 % in 150 s. The circulation of 39 kg/s, 22 times the steam flow, appears nowhere in that figure. The level of Chapter 9 is an integrator for the same reason.

6.4.2 Swell and Shrink From the Steam Side

A step in steam demand changes the outflow at once and the heat supply not at all. In case 2 the steam flow jumps from 1.75 to 2.09 kg/s, the pressure falls at about 0.018 bar/s, and bulk boiling reaches 0.31 kg/s four seconds after the step. With the vapor that holds below the surface, the surface reaches 58.0 % ten seconds after the step while the inventory has fallen to 51.78 %, as shown in Figure 6.2a. As the pressure settles toward 7.52 bar the bulk boiling stops and the bubbles clear, and by 600 s the surface is at 53.8 % and the inventory at 50.65 %, still falling because the coil's steam exceeds the fixed feedwater. The decay follows the pressure settling time, the vessel's energy capacity divided by the latent heat times the steam valve's slope, $(\partial U/\partial p)_M/\left(h_{fg}\,\partial \dot{m}_s/\partial p\right) \approx 26.5\,\mathrm{s}$ at design.

> **TRY IT IN THE APP** **2 · More Steam Demand and a Rising Level (Manual)**
>
> Load the case, set the Teaching overlay to show the simulator truth, and press Run. At 90 s the steam user's valve opens a quarter further while the feedwater valve stays put. Watch the level and the water inventory: ten seconds after the step the level stands at 58.0 % while the inventory has fallen from 51.90 to 51.78 %.

This is the inverse response of Section 1.4.7. The steam valve acts on the indicated level through two paths of opposite sign. The inventory path integrates the water lost, while the bubble path answers within seconds through the pressure and decays with it, so the level first moves the wrong way. Chapter 7 has the other classical case, in which the fast path is the water column of a penstock.

Pause. Try Exercise 6.1 before reading on.

A step down does the reverse. In case 3 a 25 % demand cut raises the pressure, the negative bulk rate condenses the bubbles, and the surface drops from 55.0 to 52.4 % in 18 s while the inventory rises from 51.90 to 52.13 %, as shown in Figure 6.2b. The feedwater step that follows at 330 s makes no shrink at all. Thirty degrees colder feedwater, mixed into 39 kg/s of loop water, lowers the coil inlet by about 1.3 °C. It reduces the steam the coil can make, the pressure eases, and the bubble contribution grows slightly, from 2.57 to 2.81 points, while the level keeps rising with the inventory.

Pause. Try Exercise 6.2 before reading on.

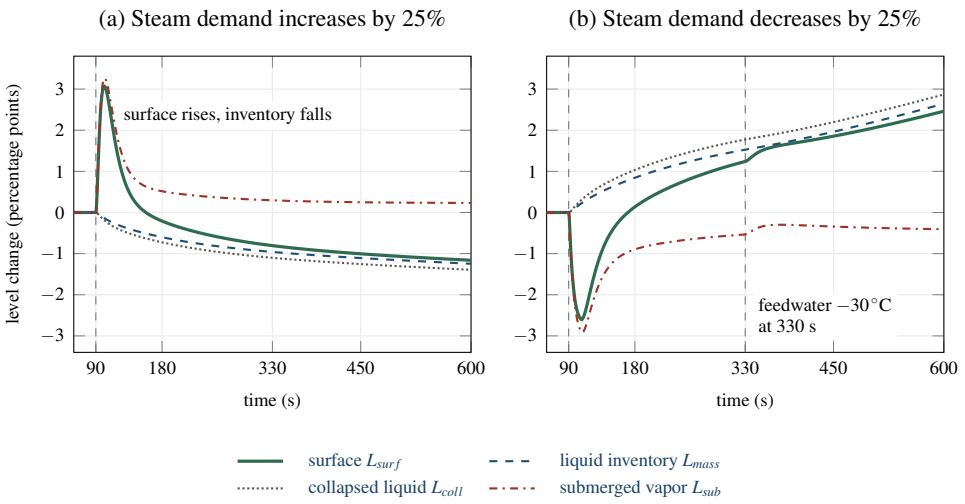


Figure 6.2 Open-loop swell and shrink, from the simulator in Manual. Deviations from the values at 60 s of the surface (thick solid), the mass-equivalent inventory (dashed), the collapsed liquid level (dotted) and the bubble contribution (dash-dotted). Both panels use the same axes, and dashed vertical lines mark the input changes. At every instant $\Delta L_{surf} = \Delta L_{coll} + \Delta L_{sub}$, while density changes separate collapsed level from mass-equivalent inventory. In (a) the surface rises 3 points while the inventory falls, and in (b) it falls 2.6 points while the inventory rises, and the colder feedwater at 330 s produces no shrink.

> **TRY IT IN THE APP** **3 · A Demand Cut Shrinks; Cold Feedwater Does Not (Manual)**
>
> Load the case and run it past 330 s. The demand cut at 90 s drops the level from 55.0 to 52.4 % in 18 s while the inventory rises to 52.13 %. Then watch the bubble trend through the feedwater step at 330 s: no shrink appears, and the bubble contribution creeps up from 2.57 to 2.81 points.

6.4.3 A Steam Production the Heat Decides

In a fired boiler the firing rate follows the load. In a heat-recovery unit the heat is set upstream, so a larger valve opening draws on stored energy only until the pressure has fallen far enough for the opening to pass what the heat supports. After the 25 % step of case 2 the steam flow returns to 1.77 kg/s within ten minutes, against a design of 1.75. A loss of heat is the disturbance that changes production for good. With the flue gas at 75 % in case 6, the heat to the water falls from 4.08 to 3.29 MW, the pressure from 8.00 to 7.52 bar and the steam from 1.75 to 1.42 kg/s, less than in proportion to the gas because the coil's effectiveness rises as the gas slows. These two facts decide the control problem. The inverse response comes from the pressure, so the steam flow is the measurement worth feeding forward, and the level controller can only match the feedwater to the steam that the heat makes.

6.5 Implemented Control System

The controllers execute every $\Delta t_c = 0.25$ s, the scan of the inner flow loop and of the instruments. The level loops run on their own period T_o of 1, 3 or 6 s and the MPC every 2 s, within the engine step of Section 1.6.1, whose plant-side bookkeeping here covers ramps, stiction and protection. The only interface between the simulator and the controllers is a measurement packet holding the transmitter readings of Equation (6.8), their quality flags and their time, and a test of the application fails if the packet ever carries anything else.

6.5.1 Single-Element PID

The single-element loop is the velocity-form PID of Section 1.6.2 with the derivative on the measurement, run every T_o on the measured level filtered by a first-order lag of $\tau_f = 2\,\mathrm{s}$, as shown in Equation (6.9).

$$\begin{aligned} e_k &= L_{SP} - L_{f,k}, \qquad P_k = K_c e_k \\ D_k &= -K_c \tau_D \dot{L}_{f,k} \\ s_k &= s_{k-1} + (P_k - P_{k-1}) + (D_k - D_{k-1}) \\ \hat{u}_k &= s_k + I_{k-1}, \qquad u_k = \min\big(100,\ \max(0,\ \hat{u}_k)\big) \\ I_k &= I_{k-1} + \Big[\frac{K_c}{\tau_I} e_k + \frac{u_k - \hat{u}_k}{\tau_I}\Big] T_o \end{aligned} \tag{6.9}$$

where u is the valve command in percent. The symbols map onto Equations (1.24) to (1.27) with $s_k = S_k$, $\hat{u}_k = v_k$, the bias folded into I, and the tracking time T_t equal to τ_I. A low level opens the valve, so the loop is reverse acting in the instrument sense. Setting the integral time to its maximum of 3000 s removes integral action. The defaults are $K_c = 4\,\%$ per percent, $\tau_I = 600\,\mathrm{s}$ and $\tau_D = 0$. For an integrating process of gain K a PI loop is critically damped at $\tau_I = 4/(KK_c)$, which is 609 s at the default gain.

6.5.2 Three-Element Cascade

The three-element loop sets a feedwater-flow setpoint from the filtered steam flow, the metered blowdown and a level trim, and an inner PI makes the valve deliver it, as shown in Equation (6.10).

$$\begin{aligned} \dot{m}_{f,SP} &= \underbrace{\bar{\dot{m}}_s + \dot{m}_b}_{\text{feedforward}} + \underbrace{\mathrm{PI}_3(L_{SP} - L_f)}_{\text{trim}} \\ v^{cmd} &= \mathrm{PI}_F(\dot{m}_{f,SP} - \dot{m}_{f,m}) \end{aligned} \tag{6.10}$$

where $\bar{\dot{m}}_s$ is the measured steam flow through a 2 s lag. The outer loop is the same velocity-form PID with the feedforward as its bias, an output range of 0 to $\dot{m}^{SP}_{f,max}$ and a rate limit, both shared with the MPC so that comparisons between them are fair, at defaults of 3.5 kg/s and 0.2 kg/s per second. When the inner loop pins the valve, the outer range collapses onto the measured flow, so the outer integral cannot wind up against a flow that is not coming. The gains keep their own units: 0.15 kg/s per percent of level with $\tau_I = 600\,\mathrm{s}$ for the trim, and 20 % of valve per kg/s with $\tau_I = 2.5\,\mathrm{s}$ for the flow loop, which reaches 63 % of a flow step in 2.8 s with 0.8 % overshoot. With the level signal lost the trim is frozen and the setpoint keeps following the steam flow. Volume 3, Chapter 2 compares flow feedforward and a level-to-flow cascade on a tank whose two disturbances enter by different routes.

Every controller that is not driving the valve tracks the one that is, by solving its integral for the present command, so Manual to Auto, a change of architecture and an MPC fallback are all bumpless. In case 14 each of three transfers moves the valve by less than 0.4 %.

6.5.3 Inventory Estimator

A four-state Kalman filter [11] reconstructs the inventory from the measurements alone. Volume 5, Chapter 12 estimates another inventory that no instrument measures, the ammonia stored on a catalyst. Its state is $\hat{x} = [\hat{M},\ \hat{b},\ \hat{d},\ \hat{y}]$, the water mass, the bubble contribution in level points, an unexplained imbalance such as a meter bias, and the transmitter's lagged view of the surface, predicted every

0.25 s as shown in Equation (6.11).

$$\begin{aligned}
\hat{M}^{+} &= \hat{M} + \Delta t_c(\dot{m}_{f,m} - \dot{m}_{s,m} - \dot{m}_b + \hat{d}) \\
\hat{d}^{+} &= \hat{d} \\
\hat{b}^{+} &= \hat{b} + (1 - \mathrm{e}^{-\Delta t_c/\tau_\beta})(b^{*} - \hat{b}) \\
\hat{y}^{+} &= \hat{y} + (1 - \mathrm{e}^{-\Delta t_c/\tau_{LT}}) \\
&\qquad \times \left(L_{coll}(\hat{M}, p_m) + \hat{b} - \hat{y}\right)
\end{aligned} \tag{6.11}$$

where b^{*} is the bubble contribution that the controller's own copy of Equation (6.5) expects from the measured pressure, its filtered rate, the hot-return temperature and the flows. The covariance is propagated with a local linearization that holds the calculated b^{*} fixed, with process noise of $(0.02\,\Delta t_c)^2\,\mathrm{kg}^2$ on the mass, $0.25^2\Delta t_c$ on the bubbles and $0.0006^2\Delta t_c$ on the imbalance, plus 10^{-6} on the transmitter-state variance per scan. Each good level sample corrects the state against $\hat{y}$ as it was one declared delay earlier, with $R = \max(0.05, \sigma_L)^2 + q^2/12$. Delayed samples use a buffered predicted reading with the current covariance, an approximation rather than a full out-of-sequence Kalman update. The quality flag reads lost after 5 s without a level, inconsistent for 120 s after an innovation larger than five standard deviations, degraded when the inventory is uncertain by more than 1.5 points or the filtered normalized innovation exceeds 6, and good otherwise. The estimator's parameters are a copy of the plant's taken when the case starts, and they change only when the reader updates the controller model.

6.5.4 Model Predictive Control

The MPC sets the feedwater-flow setpoint every 2 s through the same inner flow loop. It is the option of Section 1.6.3 with a model of its own. That model is a local linear one, rebuilt at every execution from the measurements, the estimate and the controller's copy of the parameters, as shown in Equation (6.12).

$$\begin{aligned}
\dot{x}_1 &= x_3 - (\dot{m}_{s,m} + \dot{m}_b - \hat{d}), \qquad \dot{x}_3 = (u - x_3)/\tau_{fl} \\
\dot{x}_2 &= \left(k_b\dot{m}_{s,m} + K_{bf}(x_4 - \dot{m}_{s,m} - \dot{m}_b) - x_2\right)/\tau_b \\
\dot{x}_4 &= (x_3 - x_4)/\tau_{th} \\
y_1 &= L_{coll}(\hat{M}) + c_{coll}x_1 + x_2 \\
y_2 &= L_{mass}(\hat{M}) + c_{mass}x_1
\end{aligned} \tag{6.12}$$

where x_1 is the water mass relative to the estimate, x_2 the bubble contribution starting at $\hat{b}$, x_3 the delivered feedwater with $\tau_{fl} = 4\,\mathrm{s}$, and x_4 the feedwater as the loop's thermal holdup has seen it, with $\tau_{th} = M_h/(\dot{m}_r + \dot{m}_f)$, 49 s at design. The steady bubble contribution at the measured steam flow is $k_b\dot{m}_s$ with $k_b = 100\tau_{res}\eta v_v/V = 1.77\,\%$ per kg/s, extra feedwater lowers it by $K_{bf} = -k_b(h_l - h_{fw})/h_{fg} = -0.24\,\%$ per kg/s, and an estimated swell decays in τ_b, the pressure settling time plus τ_β. Its steam-flow slope is $\partial\dot{m}_s/\partial p = \dot{m}_s p/(p^2 - p_{hdr}^2)$ before choking and $\dot{m}_s/p$ when $p_{hdr} \le 0.55p$. That decay is the piece of physics the PID loops do not have. The model is discretized exactly for a zero-order hold, with the affine term mapped step by step so that a forecast of the steam flow can enter. The quadratic program over a 120 s horizon of 60 steps, in twelve blocks of 10 s, is Equation (6.13).

$$\begin{aligned}
\min\ & \textstyle\sum_{k=1}^{60} w_L(y_{1,k} - L_{SP})^2 + \sum_j \frac{w_{\Delta u}}{0.1^2}(u_j - u_{j-1})^2 \\
& \quad + 10^3 s_L^2 + 10^2 w_M s_M^2 \\
\text{s.t. } & 0 \le u_j \le \dot{m}_{f,max}^{SP}, \qquad |u_j - u_{j-1}| \le r_f\,2\,\mathrm{s} \\
& L_{LA} - s_L \le y_{1,k} \le L_{HA} + s_L \\
& \bar{L}_m - \Delta_M - s_M \le y_{2,k} \le \bar{L}_m + \Delta_M + s_M, \quad s \ge 0
\end{aligned} \tag{6.13}$$

Table 6.3 The Fifteen Cases

#	Case and events	Teaching point
1	Manual, valve +4 points at 90 s	the water balance
2	Manual, demand +25 % at 90 s	swell
3	Manual, demand −25 % at 90, feed −30 °C at 330 s	shrink, and none from feed
4	Single, P-only; SP 55 to 59 and back; blowdown at 540 s	P, PI and PID
5	Demand +20 % at 120, back at 480 s	three-element control
6	Flue gas to 75 % over 120–240 s, back at 600 s	heat sets the steam
7	Single; feed pressure drop to 65 % at 120, back at 420 s	flow loop, stiction
8	Single, noise 1 %; demand +20 % at 180 s	filter, sampling, delay
9	MPC; fouling, circulation 70 %, foaming at 1100 s	model mismatch
10	Level bias +4 at 120, signal lost 480–510 s	lying transmitter
11	Demand +25 % at 120, back at 660 s, shared limits	MPC against PI
12	MPC; published gas schedule +25 % at 240 s	forecast value, error
13	Feed line capped at 1.0 kg/s at 120 s	an infeasible balance
14	Start at 65 % load, ramp to design; pump trip	transfers, protection
15	Challenge: four upsets in 1800 s, P-only as shipped	the whole problem

where $w_L = 1$, $w_{\Delta u} = 2$, $w_M = 1$, the level band is the warning band from 30 to 75 %, and the inventory zone is $\pm\Delta_M = \pm 5$ points around $\bar{L}_m$, the estimator's inventory when the case started. The program is solved by the dual active-set method of Goldfarb and Idnani [19] in the solver module the apps share, with 14 variables and 290 constraint rows, and the first move is applied, as in industrial MPC [20]. The estimated imbalance $\hat{d}$ in the model removes offset [12]. A solution that is not optimal, finite and inside its bounds is rejected, and the tracked three-element trim takes the flow setpoint without a bump. If the level has been missing for more than 60 s the MPC is not run and the setpoint follows the steam flow. Over the 599 executions of case 11 no constraint was active, the program never failed, and a whole solve took a median of 0.07 ms.

6.5.5 Protection and the Three Models

A protection layer outside every controller reads the simulator's truth and stops the run at a surface of 18 or 88 %, an inventory of 20 %, the model-domain limits of the assumptions, or an operator's controlled stop. The app names this an exercise stop. A stop latches its reason, marks the trends, freezes the state and pauses the run, and only Reset continues. The application keeps three models apart. The plant is everything in Section 6.3. The identified model is the integrating gain a reader fits from case 1. The controller's model is the estimator's and MPC's copy of the parameters and the reduced structure of Equations (6.11) and (6.12), which knows neither the transmitter bias nor a plant change made after the case started.

6.6 Fifteen Cases, Measured

The fifteen cases appear in Table 6.3. All share one reference configuration (the default tunings above, three-element control, transmitter noise of 0.2 % and the design disturbances) and depart from it only as their rows say. Every number in this section was measured with the engine running headless on the default seed (Section 1.6.1), stepped exactly as the browser steps it, and any setting a row names is entered before the run starts. The exported file carries each measurement beside the truth it measures, in columns whose names begin with `truth_`. Integrated absolute error (IAE) is taken on the true surface against the setpoint from $t = 0$ to the end of the run named, peak deviations on the true surface, and valve travel as the total movement of the stem in percent. Every case also loads identically whether it is selected first, chosen from the list after another case, or rerun with Reset.

6.6.1 Tuning an Integrating Level

Case 4 ships a proportional-only single-element loop at 2 % per percent. The setpoint ramps come out without offset, because the vessel integrates, but once the blowdown takes 0.15 kg/s away the level settles 1.48 % below setpoint over the last 180 s of the 1200 s run. A proportional controller holds

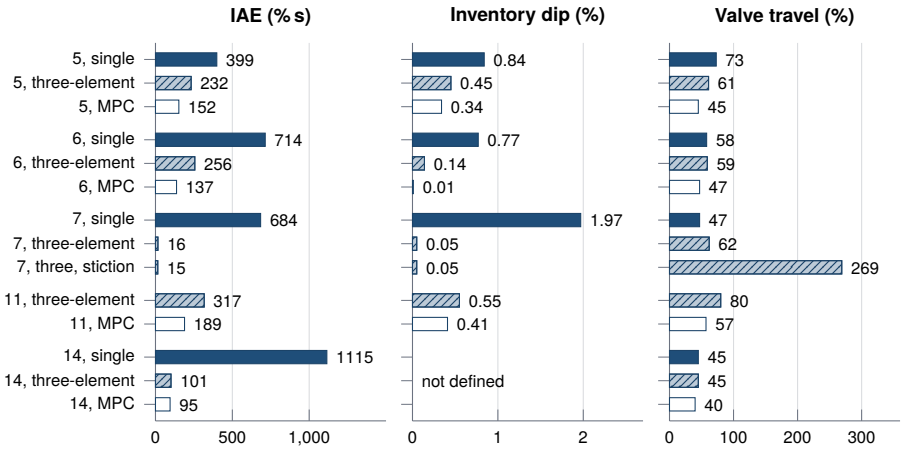


Figure 6.3 The measured comparisons of Table 6.4: single-element control (solid bars), three-element control (hatched) and the MPC (open), with the case number before each controller. Case 14 moves to a new inventory, so it has no dip. Notice that in cases 5, 6 and 11 the MPC has the smallest IAE and dip, and that the stiction band of case 7 more than quadruples the valve travel while the IAE and the dip stay where they were.

the valve away from its starting point only while there is an error. With $K_c = 8$ and $\tau_I = 300\,\text{s}$, near critical damping, the worst error in the same window is 0.19 %, and $K_c = 15$ with $\tau_I = 100\,\text{s}$ reaches 0.04 % at the cost of 39 s saturated and 363 % of valve travel. Adding $\tau_D = 20\,\text{s}$ to $K_c = 6$, $\tau_I = 400\,\text{s}$ raises the travel from 149 to 1947 % with no gain in level, because the derivative differentiates the transmitter noise.

6.6.2 Single Element, Three Element and MPC

The architectures are compared case by case in Table 6.4, and Figure 6.3 draws the same numbers as bars. On the demand steps of case 5 the swell peaks at 57.5 % under every controller, because no feedwater move can stop the pressure from falling within seconds. What changes is the recovery. Single-element control reads the swell as too much water and closes the valve as the vessel empties, with an inventory dip of 0.84 points, while three-element control raises the feedwater with the steam and dips 0.45 points with less valve travel. The MPC, whose model expects the swell to clear, dips 0.331 to 0.349 points depending on the baseline definition, with the least travel. In case 7 the flow loop catches a weaker feed pump within seconds, and the level moves 0.08 points under three-element control against 1.96 points under single-element control. A stiction band of 3 % makes the flow loop hunt, raising the travel from 62 to 269 % while the level still holds within 0.08 points.

> ✦ **DID YOU KNOW? A Swell No Controller Can Stop**
>
> From the simulator: on the demand steps of case 5 the true surface peaks 2.49 points above its setpoint under single-element control, three-element control and the MPC alike. The controllers differ only afterward, with inventory dips of 0.84, 0.45 and 0.34 points (Table 6.4).

The limits of case 11, a flow setpoint of 2.9 kg/s and a rate of 0.06 kg/s per second shared by both controllers, never bind, because a demand step raises the steam only until the pressure has settled. The MPC's advantage there, an IAE of 189 against 317 % s with 57 against 80 % of valve travel, therefore comes from prediction and not from constraint handling. Its inventory-zone slack appears in case 13, where the feed line is capped at 1.0 kg/s against 1.75 kg/s of steam. There the MPC reports a slack from 300 s that grows to 25 points while it keeps asking for 3.5 kg/s it cannot

Table 6.4 Measured Comparisons. IAE on the True Surface Over the Whole Run, Peak Deviation, Inventory Dip and Valve Travel

Case	Controller	IAE (% s)	Peak (%)	Dip (%)	Travel (%)
5, 900 s	single	399	2.49	0.84	73
	three-element	232	2.49	0.45	61
	MPC	152	2.49	0.34	45
6, 1200 s	single	714	1.34	0.77	58
	three-element	256	0.37	0.14	59
	MPC	137	0.41	0.01	47
7, 900 s	single	684	1.96	1.97	47
	three-element	16	0.08	0.05	62
	three, stiction 3 %	15	0.08	0.05	269
11, 1200 s	three-element	317	3.09	0.55	80
	MPC	189	3.09	0.41	57
14, 900 s	single	1115	2.21	—[a]	45
	three-element	101	0.41	—[a]	45
	MPC	95	0.40	—[a]	40

[a]Case 14 moves from the part-load to the design inventory, so no dip is defined.

get, and the low-inventory stop comes at 1060 s, against 1059 s under three-element control and 964 s under single-element control, which opens the valve late. With an imbalance of 0.75 kg/s and 24.3 kg per point the inventory falls about 1.85 % a minute. The stop comes a little sooner than that predicts because the steam rises from 1.75 to 1.83 kg/s as less cold feedwater leaves more of the coil's heat for steam. Closing the steam valve does not help, since the heat still has to leave as steam, and the actions that end the deficit are a gas bypass or a controlled stop.

🕮 **Pause.** Try Exercise 6.3 before reading on.

6.6.3 Noise, Delay, Mismatch and a Lying Transmitter

With 1 % of transmitter noise in case 8, the shipped loop moves its valve 1059 % over 900 s with no level filter, 216 % with the default 2 s filter and 81 % at 8 s, with IAE of 260, 222 and 225 % s. On that slow loop the filter is nearly free. On an aggressive loop ($K_c = 20$, $\tau_I = 60$ s) a 12 s filter with 6 s sampling and a 4 s delay raises the IAE after the step from 156 to 279 % s and deepens the dip after the step from 1.6 to 2.6 points.

In case 9 the MPC runs through water-side fouling to 75 %, a circulation cut to 70 % and, at 1100 s, foaming water that doubles the submerged shares. The estimate follows the first two within 0.02 points of the true inventory, because their effects reach it through measured temperatures and flows. The foam changes a parameter no measurement shows. The estimator explains the higher surface as water, the MPC holds the surface on setpoint, and the true inventory falls to 48.5 % while the estimate stays near 52.0 %, with the residual small and the quality good. Updating the controller model at 1200 s brings the estimate within 0.5 points of the truth by 1450 s. In case 10 a four-point transmitter bias appears at 120 s. It is twenty standard deviations of the noise in one sample, so the estimator flags the level inconsistent for 120 s, and then it absorbs the bias as water. The controller drives the true surface to 50.4 % and the inventory to 47.3 % while the estimate reads up to 53.0 %. During the 30 s loss of signal the estimator carries on with the flows and its $\pm 2\sigma$ band grows only from 0.27 to 0.30 points.

6.6.4 Forecasts and the Load Change

Case 12 publishes the upstream process's production plan, 25 % more flue gas over 30 s from 240 s and back from 780 s. The forecast scales the measured steam by the 0.6 power of the scheduled gas flow, the dependence of Equation (6.3). The MPC's IAE over 1200 s is 119 without preview, 112 with it, and 124 % s when the schedule runs 45 s late, worse than no preview because the MPC feeds water for steam not yet made. The three-element loop given the same forecast as a feedforward 30 s

Table 6.5 The Challenge Over 1800 s

Configuration	IAE (% s)	Travel (%)	Final (%)	Verdict
single, P-only, $K_c = 1.5$ (shipped)	2123	43	3.65	fails
single, $K_c = 8$, $\tau_I = 300$ s	476	233	0.07	fails
three-element, defaults	316	104	0.15	fails
three-element, 0.20, 600 s, $\tau_f = 4$ s	297	104	0.11	passes
MPC, defaults (suggested)	175	72	0.01	passes

ahead scores 228 against 217 % s without it. In case 14, starting at 65 % load and ramping to design over five minutes, three-element control and the MPC hold the surface within 0.41 and 0.40 points, while single-element control strays 2.21 points and ends the run 1.8 points low. A circulation pump trip ends that case on loss of circulation within about 2 s.

6.6.5 The Challenge

The challenge holds 55 % through a 20 % demand rise at 150 s, feedwater 15 °C colder at 510 s, 15 % less flue gas from 870 s and the demand's return at 1230 s. Its target is no exercise stop, an IAE of at most 300 % s after 150 s, valve travel of at most 150 % over the run, and the true surface within ±0.5 % of setpoint over the last 180 s, judged every physics step. Five configurations are listed in Table 6.5. The proportional-only loop it ships with ends 3.65 % off setpoint. Tuning trades IAE against travel, so the three-element loop passes only in a narrow window, while the MPC at its defaults passes with the widest margin on both budgets. The swell of case 2 near its peak is shown in Figure 6.4.

6.7 Technology in Practice

The unit of Section 6.3 reads its level through the chain of Equation (6.8) and moves one feedwater valve. Figure 6.5 draws the three-element loop with plant devices.

What Measures It

A level instrument on a boiling vessel turns a pressure difference or an echo time into a height, and each conversion needs a property of water or steam. A differential-pressure transmitter weighs the column between its taps. In a vessel of constant area that weight follows the mass of liquid whatever its density [21], so the reading sits closer to the chapter's mass-equivalent level than to the surface. A radar transmitter times an echo through the vapor, and its accuracy shifts with the pressure and temperature of that vapor, the gas phase effect [21]. The term $k_{cal}(p - p_0)$ of Equation (6.8) stands for such a pressure-dependent calibration. An orifice meter reads flow as the square root of its pressure drop divided by the density, and a steam table gives that density for steam [21]. No range, accuracy or response time is given here, because none was found in an open source that names no maker.



Figure 6.4 The swell of case 2 near its peak, captured at $t = 104$ s (the clock is in the cut header), on the Level and inventory view with the teaching overlay on. The figure is a composite of the view on top, the controller's readout line in the middle and the trends below, with the other panels cut. The vessel (1) stands beside a zoomed comparison in which the transmitter reads 58.0 % (2) and the estimator splits its 58.1 % surface estimate into collapsed liquid volume and submerged vapor (3). Black ticks show mass-equivalent inventory separately, including the estimate of 51.8 % (±0.3, two standard deviations) and true liquid inventory of 51.7 %, against a surface setpoint of 55 %. The readout line (4) shows Manual with the feedwater valve fixed at 43.8 %. Over a two-minute window the trends show the level rising after the 90 s step (5), the steam flow at 2.03 kg/s against 1.75 kg/s of feedwater, the inventory estimate on top of the truth, and the bubble contribution climbing to 6.3 points from the design value of 3.1 points given after Equation (6.6), which the trend axis does not print.

> **TECHNOLOGY** **Sensor: A Differential-Pressure Level Transmitter on a Steam Vessel**
>
> A differential-pressure transmitter compares the pressure at a tap low on the vessel with the pressure in the vapor space, so the vessel pressure cancels and the output is the head of the liquid alone. On a steam vessel the upper impulse line fills with condensate and becomes a wet leg, whose constant head a zero shift removes. Seal pots are standard on boiler drums, because a wet leg that loses water gives readings in error until steam condenses to refill it. Turning the head into a height needs the density, and hot water at power-boiler temperatures can weigh little more than half as much as cold water [21].

What Moves It

The feedwater valve of Equation (6.7) follows its target with a lag of 2.5 s, a stroking limit and an optional stiction band. A plant valve has a pneumatic actuator, and a positioner corrects the difference

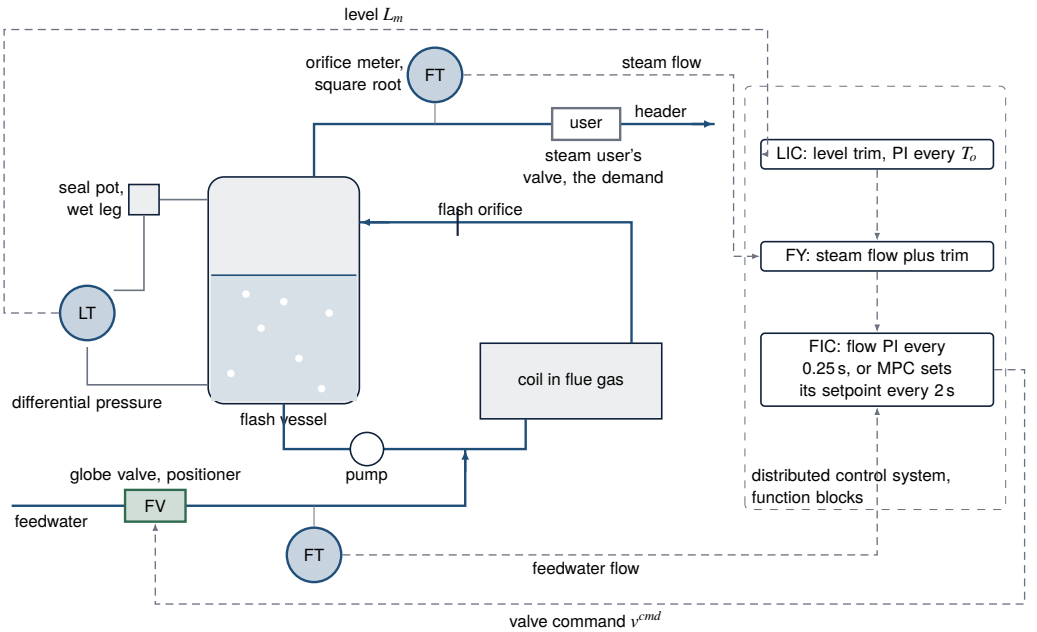


Figure 6.5 The three-element loop of the chapter with plant devices in place of the schematic's symbols. LT reads the level as a differential pressure through a wet leg kept full by a seal pot, an orifice meter reads the steam flow, and on the control platform the steam flow plus the level trim becomes the setpoint of the feedwater flow controller, which moves the globe valve. Notice that the flow loop sits between the level controller and the valve, so a change in feed pressure is corrected before the level sees it.

between the commanded and the actual stem position [21].

> **TECHNOLOGY** **Final Element: Feedwater Valves in Split Range**
>
> A large boiler may feed its drum through two valves in sequence. The small valve opens over the first half of the controller's signal and the large one over the second half. At full power the small valve stands wide open and the large one throttles, while at low power the small one throttles alone. The pair controls the drum level better over the whole range than any single valve. Inside a flow loop, a change in feed pressure is countered before the level is upset [21].

Where the Controller Runs

A distributed control system controls production in industries such as electric power generation and chemical manufacturing. Its loops may be nested or cascaded, with the setpoint of one loop taken from the measurement of another, and they run with cycle times from milliseconds to minutes [22]. The chapter's inner loop and instruments run every 0.25 s, the level loops every 1 to 6 s and the MPC every 2 s, all inside that range. Three-element control is built from standard blocks: a flow controller whose remote setpoint comes from a summing block that adds the steam flow to the level controller's trim [21], the arrangement of Equation (6.10).

IN PRACTICE **Which Water Formulation a Contract Names**

In the steam power industry IAPWS-IF97 is the standard for contracts and tests, while scientific work prefers the more accurate IAPWS-95. The industrial formulation divides its range into five regions, with small discontinuities at their boundaries, and it is weakest near the critical point [15]. Before a property routine is trusted, the release asks that each programmed equation be checked against its tables of test values in eight-byte arithmetic [18]. Because so much rests on it, the industry wants to change its formulation only about every 30 years [15].

What Is Modernizing It

Two developments touch this unit. The first is fast property calculation. An IAPWS guideline describes spline tables that reproduce IAPWS-IF97 at high speed for the simulation of transients and for real-time optimization. It finds the pressure of a two-phase state from its specific volume and internal energy by a one-dimensional Newton iteration [23], the same problem as the closure of Equation (6.1). In flow simulations the spline tables cut the computing time by factors of 6 to 10 [23]. The second is model-based level control. Unplanned reactor trips caused by steam generator level control have been significant contributors to the unavailability of nuclear plants, and there the only true sign of an inventory change is the mismatch between steam and feedwater flows [24]. A predictive controller pairs an estimator of the plant state with an optimizer of future moves [24], the structure of Section 6.5.

6.8 Limitations

The sizes and settings are generic teaching values that make a consistent design point, a 4 MW coil and 6.3 t/h of steam at 8 bar, without being fitted to any unit. The vessel is in saturated equilibrium and the loop and coil are lumped, so the model cannot say where in the vessel the vapor sits or how a temperature front travels through the coil. The bubble closure is synthetic, so the size and timing of the swell are plausible for a teaching unit rather than predicted for a particular vessel, inlet device or water chemistry, and case 9 changes the closure on purpose to show what that means for an estimator. Pump curves, pipe hydraulics, carryover and the chemistry behind foaming are outside the model.

The estimator and the MPC carry a reduced structure with the plant's own declared parameters, so they meet model error only where a case introduces it. The MPC's model omits the gas side and sees heat changes only through the measured steam flow, and its linear prediction is rebuilt every 2 s without any guarantee beyond the horizon. The protection layer reads the truth and is an educational guardrail, not certified trip logic, and a stop freezes the simulation rather than model a shutdown. The disturbances arrive as scheduled steps and ramps rather than as the habits of a real process.

OPEN CHALLENGE **Better Data for the Next Water Formulation**

IAPWS-IF97 was fitted to IAPWS-95 [15], and IAPWS plans to replace IAPWS-95 because new data exist and its functional form has shortcomings. The replacement needs saturation vapor pressures above 373 K measured to 0.02 % or better and liquid densities near saturation above about 400 K to 0.01 %. The metastable fluid inside the two-phase envelope is difficult or impossible to measure, so molecular simulation is asked to fill the gap, and uncertainty estimates for calculated properties are still somewhat ad hoc [25]. Progress would be those measurements, a prescription for the uncertainty of any calculated property, and a formulation built on both.

6.9 Conclusion

This chapter studied a flash vessel on a heat-recovery loop, a plant whose level reads water plus the vapor below the surface. Its simulator logs the indicated level, the inventory and the bubbles side by side from balances closed by IAPWS-IF97. A 25 % rise in steam demand swells the surface 3.0 points in ten seconds while the vessel loses water, a 25 % fall shrinks it 2.6 points while it fills, and the tested 30 °C feedwater cooling causes no initial shrink, because the loop heats it first. Steam production

follows the heat, so a demand step changes the outflow only until the pressure settles. On the demand steps of case 5 three-element control cuts the IAE from 399 to 232 % s against single-element control, and the MPC reaches 152 % s with the least valve travel because its model expects the swell to clear. The inventory estimate follows fouling and lost circulation within 0.02 points but is misled by foaming water and by a persistent transmitter bias, which it flags only in the first 120 s. The model is lumped and its closure synthetic. A fuller model would add a distributed coil with transport delay, a horizontal vessel, a pressure loop with its own actuator, and a bubble closure fitted to a measured transient. The simulator shows that a level controller has to know what its transmitter is measuring before it can act on it.

KEY POINTS

- The level of a boiling vessel reads water plus the vapor below its surface, so more steam demand lifts the level while the vessel loses water.
- IAPWS-IF97 supplies the properties, and the saturation closure turns the vessel's water mass and energy into its pressure.
- Both swell and shrink come from the steam side through pressure, and colder feedwater, heated in the loop first, causes no initial shrink.
- On the demand steps of case 5 three-element control cuts the IAE from 399 to 232 % s against single-element control, and the MPC reaches 152 % s with the least valve travel.

Where else this applies. Chapter 7 has an inverse response whose fast path is a water column, and Volume 5, Chapter 12 estimates an inventory that no instrument measures.

References

[1] H. Jouhara, N. Khordehgah, S. Almahmoud, B. Delpech, A. Chauhan, and S. A. Tassou, "Waste heat recovery technologies and applications," *Thermal Science and Engineering Progress*, vol. 6, pp. 268–289, 2018.

[2] R. DiPippo, "Single-flash steam power plants," in *Geothermal Power Plants: Principles, Applications, Case Studies and Environmental Impact*. Oxford, UK: Butterworth-Heinemann, 2016, pp. 107–142.

[3] D. Saury, S. Harmand, and M. Siroux, "Flash evaporation from a water pool: Influence of the liquid height and of the depressurization rate," *International Journal of Thermal Sciences*, vol. 44, no. 10, pp. 953–965, 2005.

[4] L. W. Ward, "Application of drift flux to transient two-phase level swell," *Nuclear Technology*, vol. 45, no. 1, pp. 68–76, 1979.

[5] K. J. Åström and R. D. Bell, "Drum-boiler dynamics," *Automatica*, vol. 36, no. 3, pp. 363–378, 2000.

[6] S. G. Dukelow, *The Control of Boilers*, 2nd ed. Research Triangle Park, NC: Instrument Society of America, 1991.

[7] P. U. Sunil, J. Barve, and P. S. V. Nataraj, "A robust heat recovery steam generator drum level control for wide range operation flexibility considering renewable energy integration," *Energy*, vol. 163, pp. 873–893, 2018.

[8] M. A. Elhosseini, A. S. El-din, H. A. Ali, and A. Abraham, "Heat recovery steam generator (HRSG) three-element drum level control utilizing fractional order PID and fuzzy controllers," *ISA Transactions*, vol. 122, pp. 281–293, 2022.

[9] W. L. Luyben, "Identification and tuning of integrating processes with deadtime and inverse response," *Industrial & Engineering Chemistry Research*, vol. 42, no. 13, pp. 3030–3035, 2003.

[10] W. Wagner, J. R. Cooper, A. Dittmann, J. Kijima, H.-J. Kretzschmar, A. Kruse, R. Mareš, K. Oguchi, H. Sato, I. Stöcker, O. Šifner, Y. Takaishi, I. Tanishita, J. Trübenbach, and T. Willkommen, "The IAPWS industrial formulation 1997 for the thermodynamic properties of water and steam," *Journal of Engineering for Gas Turbines and Power*, vol. 122, no. 1, pp. 150–184, 2000.

[11] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

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

[13] M. Xu, S. Li, and W. Cai, "Cascade generalized predictive control strategy for boiler drum level," *ISA Transactions*, vol. 44, no. 3, pp. 399–411, 2005.

[14] International Association for the Properties of Water and Steam, "Obsolete IAPWS recommendations," IAPWS, Information document, Mar. 2005. [Online]. Available: https://iapws.org/documents/release/Obsolete.download

[15] ——, "FAQs about IAPWS," Web page, https://iapws.org/faqs/faq2, 2026, accessed 21 November 2026.

[16] I. H. Bell, J. Wronski, S. Quoilin, and V. Lemort, "Pure and pseudo-pure fluid thermophysical property evaluation and the open-source thermophysical property library CoolProp," *Industrial & Engineering Chemistry Research*, vol. 53, no. 6, pp. 2498–2508, 2014.

[17] W. Wagner and A. Pruß, "The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use," *Journal of Physical and Chemical Reference Data*, vol. 31, no. 2, pp. 387–535, 2002.

[18] International Association for the Properties of Water and Steam, "Revised release on the IAPWS industrial formulation 1997 for the thermodynamic properties of water and steam," IAPWS, Lucerne, Switzerland, Release R7-97(2012), 2007. [Online]. Available: https://iapws.org/technical-guidance/release/IF97-Rev

[19] 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.

[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] 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.

[22] 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, Gaithersburg, MD, Special Publication 800-82r3, 2023.

[23] International Association for the Properties of Water and Steam, "Guideline on the fast calculation of steam and water properties with the spline-based table look-up method (SBTL)," IAPWS, Stockholm, Sweden, Guideline G13-15, 2015. [Online]. Available: https://iapws.org/technical-guidance/release/SBTL

[24] A. M. H. Basher and J. March-Leuba, "Development of a robust model-based water level controller for U-tube steam generator," Oak Ridge National Laboratory, Oak Ridge, TN, Tech. Rep. ORNL/TM-2001/166, Sep. 2001.

[25] International Association for the Properties of Water and Steam, "New thermodynamic data for ordinary water," IAPWS, Certified Research Need ICRN 31, 2024. [Online]. Available: https://iapws.org/technical-guidance/icrn

6.10 Exercises

6.1 ○ Concept check · 2 min

In case 2 the steam user's valve opens 25 % further at 90 s while the feedwater valve stays where it is. Ten seconds later the surface stands at 58.0 % against 55.0 % before the step. What holds the surface up?

(a) The falling pressure lets the feedwater valve pass more water, so the vessel gains mass for a few seconds

(b) The falling pressure boils liquid in the bulk, and vapor held below the surface lifts it while the water mass falls

(c) Saturated liquid expands as the pressure falls, so the collapsed level rises with the same water mass

(d) The circulation pump returns its 39 kg/s faster, so the vessel holds more of the loop's water for a while

6.2 ○ Concept check · 2 min

In case 3 the feedwater from the deaerator arrives 30 °C colder at 330 s, with the feedwater valve fixed in Manual. In a drum boiler such a step shrinks the level. What does the indicated level of this flash vessel do over the following minutes?

(a) It shrinks at once, because the cold water condenses the vapor held below the surface

(b) It falls slowly, because the colder feed lowers the flash quality and the vessel loses water

(c) It keeps rising with the inventory, because the feed is heated in the loop before it reaches the vessel

(d) It ends the run, because the coil inlet cools below the subcooling limit that protects the loop

6.3 Concept check · 1 min

In case 11 the MPC and the three-element loop share the inner flow loop, the measurements and the limits on the flow setpoint. The MPC scores an IAE of 189 against 317 % s. Where does its advantage come from?

(a) The 2.9 kg/s limit on the flow setpoint binds under the PI loop and not under the MPC

(b) The MPC measures the steam flow, which the three-element loop leaves out of its calculation

(c) The MPC drives a faster inner flow loop than the one under the three-element trim

(d) Its model expects the bubble contribution to decay as the pressure settles, so it does not chase the swell

6.4 Theory and calculation · 6 min

The vessel is a vertical cylinder of diameter $D = 1.2\,\mathrm{m}$ and effective height $H = 2.4\,\mathrm{m}$, and one point of mass-equivalent level is measured at the reference density $\rho_{ref} = 897.03\,\mathrm{kg/m^3}$. One percent of feedwater valve passes 0.04 kg/s.

(a) Find the liquid mass in one point of level and the integrating gain of the inventory in percent per second for each kg/s of imbalance. *Checkpoint:* 24.3 kg and 0.0411 %/s per kg/s.

(b) Find the single-element process gain K in percent per second for each percent of valve, and the time the four-point valve step of case 1 takes to raise the inventory by one point. *Checkpoint:* 0.001 64 %/s per percent, and 152 s.

(c) A PI loop on an integrating process of gain K is critically damped at $\tau_I = 4/(KK_c)$. Find that integral time at $K_c = 4$ and at $K_c = 8$ percent per percent. *Checkpoint:* 609 s and 304 s.

6.5 Theory and calculation · 8 min

At the design point the vessel is at $p = 8\,\mathrm{bar}$, where saturated liquid and vapor have $h_l = 721.0\,\mathrm{kJ/kg}$ and $h_v = 2768.3\,\mathrm{kJ/kg}$. The hot return carries $h_h = 809.1\,\mathrm{kJ/kg}$ in $\dot{m}_r + \dot{m}_f = 39 + 1.75\,\mathrm{kg/s}$. The steam valve has $K_s = 0.625\,\mathrm{kg/(s\,bar)}$, opening $a_0 = 0.6$ and a header at $p_{hdr} = 6.5\,\mathrm{bar}$, and $(\partial U/\partial p)_M = 34.9\,\mathrm{MJ/bar}$.

(a) Find the flash quality and the flash rate at the orifice. *Checkpoint:* 4.30 % and 1.75 kg/s.

(b) Find the steam flow from Equation (6.7) and its slope with respect to the vessel pressure. *Checkpoint:* 1.75 kg/s and 0.643 kg/(s bar).

(c) Find the pressure settling time of Section 6.4.2. *Checkpoint:* 26.5 s.

(d) In case 2 the opening becomes $a = 0.75$, while the heat, and with it the steam the coil can make, stays near design. Find the pressure at which the larger opening passes 1.75 kg/s. *Checkpoint:* 7.50 bar.

6.6 Theory and calculation · 6 min

In case 13 a clogged strainer caps the feed line at 1.0 kg/s at 120 s, while the coil goes on making 1.75 kg/s of steam. The inventory starts at 51.90 %, the low-inventory stop is at 20 %, and one point is 24.3 kg.

(a) Find the rate at which the inventory falls with three-element control, which opens the valve fully within seconds, and the time of the stop. *Checkpoint:* 1.85 % a minute, a stop near 1155 s.

(b) The steam rises to 1.83 kg/s as less cold feed reaches the loop. Repeat part (a), and compare with the measured stop at 1059 s. *Checkpoint:* 2.05 % a minute, a stop near 1056 s.

(c) The restriction acts as $a_{line} = 0.25$ in Equation (6.7), with a linear trim and $\dot{m}_{f,max} = 4.0\,\mathrm{kg/s}$. Find the feedwater that passes at the design opening of 43.75 %, and the inventory loss per minute while a single-element loop leaves the valve there. *Checkpoint:* 0.44 kg/s, and 3.2 % a minute.

6.7 App challenge · 8 min

Load scenario 2, More steam demand and a rising level (Manual). Set the Teaching overlay to Show simulator truth and change nothing else. Run at speed 5 to 600 s with the default seed, then export the CSV file.

(a) Read `truth_level_surface_pct`, `truth_inventory_pct` and `truth_bubble_pct` in the rows at 90 and 100 s. *Checkpoint:* 55.00, 51.89 and 3.11 at 90 s. 58.03, 51.78 and 6.30 at 100 s, each within 0.02.

(b) From `truth_mass_kg` in the same two rows, find the water the vessel lost in those ten seconds. Then find the mass that the rise of the surface would mean if every point of it were 24.3 kg of water. *Checkpoint:* 2.9 kg lost, against 74 kg implied.

(c) Read the surface, the inventory, the pressure `truth_pressure_bara` and the steam flow `truth_steam_kgps` at 600 s, and say why the inventory is still falling. *Checkpoint:* 53.84 and 50.65 %, 7.52 bar and 1.77 kg/s, each within 0.01.

6.8 App challenge · 9 min

Load scenario 5, Add three-element control, which ships on Single. Run it to 900 s at speed 20 with the default seed, read the IAE of the scorecard and the Valve travel tile, and export the CSV file. Press Reset, select Three-element on the level card with Steam-flow feedforward On, as shipped, and run to 900 s again.

(a) In the single-element file, read the valve command `fw_valve_cmd_pct` at 120 s and its lowest value between 120 and 170 s. *Checkpoint:* 43.7 %, falling to 34.0 % at about 134 s.

(b) For each run, give the scorecard IAE, the valve travel, the lowest `truth_inventory_pct` and the highest `truth_level_surface_pct`. *Checkpoint:* Single: 429.9, 73 %, 51.06, 57.49. Three-element: 296.2, 61 %, 51.45, 57.49. IAE within 3, travel within 1, levels within 0.01.

(c) Explain why the peak is the same in both runs, and why the scorecard IAE is larger than the 399 and 232 % s of Table 6.4.

6.9 App challenge · 8 min

Load scenario 4, Tune P, PI and PID on the level, which ships as a proportional-only loop with Level gain (valve) 2 % per percent and Level integral time (valve) 3000 s. Run it to 1200 s at speed 20 with the default seed and export the CSV file. Press Reset, set Level gain (valve) to 8 and Level integral time (valve) to 300 s, run to 1200 s again and export.

(a) In the first file, find the mean and the largest deviation of `truth_level_surface_pct` from the 55 % setpoint over the last 180 s. *Checkpoint:* Mean −1.37 %, largest 1.48 %, each within 0.02.

(b) In the second file, find the largest deviation over the same window, and read the Valve travel tile of each run. *Checkpoint:* 0.19 % within 0.02. Travel 46 and 205 %, within 1.

(c) Explain why the proportional loop leaves no offset after the setpoint ramps but does after the blowdown opens at 540 s.

6.10 Application area · 4 min

A site asks whether the unit could export steam at 10 bar instead of 8 bar by raising the vessel pressure, with the coil, the circulation and the 18 bar loop pressure unchanged. At 10 bar saturated liquid and vapor have $h_l = 762.7\,\mathrm{kJ/kg}$ and $h_v = 2777.1\,\mathrm{kJ/kg}$. Saturation at 18 bar is 207.1 °C, and liquid at 849 kJ/kg and 18 bar is at 199.3 °C.

(a) Find the flash quality of the design return, 809.1 kJ/kg, at the higher vessel pressure, and the flash rate from 40.75 kg/s. *Checkpoint:* 2.30 % and 0.94 kg/s.

(b) Find the return enthalpy that flashes the design 4.30 % at 10 bar. Name two limits of the chapter's model that a loop running that hot approaches. *Checkpoint:* 849 kJ/kg, about 199 °C.

6.11 Application area · 3 min

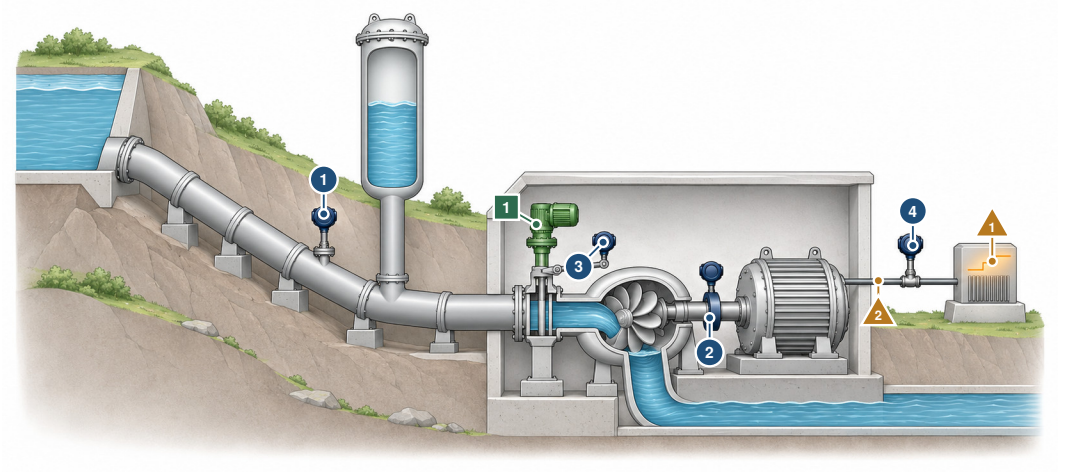
A three-element loop on this vessel holds the level within 0.08 points through the feed-pressure dip of case 7, both with a healthy feedwater valve and with a 3 % stiction band. The level trend gives no sign of the second fault. Which signals show it, and what does it cost?

6.12 Review the agent · 5 min

An AI assistant reviewing the trend of case 2 writes: "After the 25 % demand step the level rose from 55.0 to 58.0 % in ten seconds. At 24.3 kg per point, about 73 kg of water entered the vessel, so the feedwater was too high. In automatic, the level controller should cut the feedwater as soon as the level rises after a demand increase." Find the flaw and say how to check it in the app.

CHAPTER 7

Water Inertia, Transient Droop and Closure Laws in a Hydroelectric Governor Simulator



The controller moves 1 the gate opening, through the gate servo

The controller reads 1 water pressure in the pipe 2 turning speed of the shaft 3 gate position 4 electrical power to the town

What disturbs it 1 a step in the town's load 2 loss of the whole load when the breaker opens

Notice the long column of water between the reservoir and the gate: it must speed up or slow down before the power can follow the gate. In the chapter's model the surge tank stands at the top of the long penstock, the pressure is read at the turbine inlet and the gates are turned by a hydraulic servomotor. Illustration, not to scale.

IN PLAIN TERMS A Water Turbine That First Answers Backward

The main idea. A hydroelectric plant sends water from a lake down a long pipe to a turbine. A governor opens or closes the gates in front of the turbine to hold its speed, and that speed sets the frequency of a town the plant feeds alone.

The challenge. Opening the gates first lowers the power, because the water in the pipe takes time to speed up. A governor that pushes too hard makes the town's frequency swing. Shutting the gates fast when the town is suddenly cut off sends a pressure surge up the pipe.

Why it matters. A governor must be tuned to the water in its pipe before anything else.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the history of the first hydroelectric plants (Section 7.2), the fact on plant size (Section 7.3), the try-it segments of cases 1 and 5 and the fact on the nadir (Section 7.6) 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 results (Section 7.6), Technology in Practice (Section 7.7), Limitations (Section 7.8) and the open challenge

The control question Opening the gate of a hydro turbine first lowers its power, because the water in the penstock must be accelerated by head taken from the turbine. How far does that right-half-plane zero limit the governor of a unit that supplies a town alone, and what does the same water column demand of the gate when the load is lost?

Learning objectives

- Derive the gate-to-power zero and pole of the rigid water column, and the ratio of the first response to the final one.
- Compute the frequency at which an islanded town settles at a fixed gate.
- Read the island's gain limit on the stability map and on the running plant, and tune a transient-droop governor by the rule.
- Explain why a surge tank can oscillate under a governor that holds the power, and compare Thoma's area with the closed-loop boundary.
- Choose a gate closure law that trades water hammer against overspeed after a load rejection.

Prerequisites Sections 1.4.2 and 1.4.3 (balances and linear models), 1.4.7 (inverse response and the ideal turbine), 1.6.2 (velocity form) and 1.6.4 (scenarios and the challenge).

The app https://apmonitor.com/apps/hydro-governor/

See also Chapter 6 for an inverse response from swell and shrink and Chapter 8 for droop on a grid with finite inertia. Volume 7, Chapter 13 for a stability boundary that moves with the operating point and Volume 7, Chapter 3 for one tuning that must serve several plants. Volume 17, Chapter 7 for droop governors, shedding and a synchronizer on an island.

7.1 Introduction

A 400 m steel penstock that delivers 80 MW to a Francis turbine on a 100 m head holds about 9400 tonnes of water moving at 3.8 m/s. In the simulator of this chapter, raising the gate from 65.6 to 75 % by hand at that load cuts the turbine power by 4.4 MW within 0.74 s. The power is back above 80 MW only 1.37 s after the move, and the frequency of the islanded town it supplies passes 62 Hz 10.9 s after the move. The extra opening lowered the head at the runner before the column could speed up, which is the inverse response of a water column and the starting point for every hydro governor. It is the plant behind the ideal turbine of Section 1.4.7, and Section 7.6.1 measures it.

Hydropower is an established technology that is still developing, as a recent survey of the emerging technologies of the sector shows [1]. Its governors are tuned against the water rather than against the machine. The standard dynamic models of the turbine, the penstock and the surge tank were collected by an IEEE working group [2], the guide to applying governing systems is an IEEE standard [3], and the classical textbook treatment of the turbine and its governor is Kundur's [4]. The behavior has a wider family. A right-half-plane zero imposes design tradeoffs on any feedback loop around it [5], inverse response was treated as a process control problem in 1962 [6], and the shrink and swell of a boiler drum is the same structure in another plant [7], a structure that Chapter 6 meets in a flash vessel. A reader who has tuned a drum level, a boost converter or a reactor with competing fast and slow effects has met the shape of this problem.

The simulator is built around two misconceptions. The first is that more opening always means more power at once, which the inverse response in the opening paragraph contradicts. The second is that a governor that behaves well on the grid will behave well when the unit supplies a load alone. On the grid the stiff network holds the speed and almost any gain is stable, while on an island the water starting time against the rotor's mechanical starting time sets a hard ceiling. The grid tuning that the challenge ships with, a gain of 8 with a 2 s integral time, sits above the island boundary of 6.04 for that integral time, and the town swings between 56.0 and 62.5 Hz.

A student usually meets the pieces of this problem in separate places, governor stability in power system texts [4, 8], gate closure schedules in hydraulic studies [9, 10], and surge tanks in the literature on hydraulic transients [11, 12]. This chapter works with one browser simulator in which the inverse

response, the gain limit, the transient-droop rule, the load share on a grid, the closure law of a load rejection and the surge tank all act on one plant, and it reports what each does there, with numbers. The model is the IEEE working group's, the governors are the two forms the industry builds, and multi-stage closure laws are established practice. Section 7.2 traces how hydro governing reached its present form. Section 7.3 gives the 26-state model with its optional elastic penstock, and a measurement boundary that keeps the governor away from the simulated flow. Section 7.4 derives the right-half-plane zero of the water column and the stability map, computed from the same equations the engine integrates and checked against the engine's own oscillation onset, and treats the tuning rule, water hammer and the surge tank. Section 7.5 states the two governors, the power controller and the load-rejection sequencer. Section 7.6 measures nine graded cases, from a gate moved by hand to a challenge that islands a town through a load step and a rejection, and checks the code against its own equations. Section 7.7 describes the instruments, gate servomotors and platforms of real governors, Section 7.8 states the limitations and Section 7.9 draws the results together.

7.2 From Dashpot Governors to Closure Laws

Hydro governors have long been tuned in a form peculiar to water, a governor that acts weakly at first, with a large transient or temporary droop, and builds to its full action over a reset time. In 1962 Hovey published a method for adjusting the governors of a utility's hydro units on that basis [13], and five years later Undrill and Woodward revisited the temporary droop calculation with a nonlinear model [14]. The rule that relates the transient droop and the reset time to the water and mechanical starting times reached later engineers through Kundur's textbook, which gives it after Hovey [4]. Interconnection made the governors a system matter as well as a unit matter. The coordination of hydraulic turbine governors for power system operation was addressed in 1966 [15], and governor tests by simulated isolation of hydraulic turbine units were reported two years later [16].

Proportional-integral-derivative governors followed, and the stability of a hydraulic turbine generating unit under a PID governor was analyzed in 1979 [8]. Hydraulic turbine and turbine control models for system dynamic studies were collected by an IEEE working group in 1992 [2], and the application of governing systems to hydroelectric units is now covered by an IEEE guide [3].

The hydraulic side has its own history. The pressure rise when a valve stops a flow faster than a wave can travel the pipe and back is Joukowsky's relation, and the slower closures that hydro plants actually use were treated by the water hammer theory that followed it [17]. A surge tank shortens the water column that the turbine sees, and a governor that holds its power constant can make the tank's slow oscillation grow, which is why tanks are sized against a stability area named after Thoma [11] and why throttling a tank changes governor stability and power control [12]. The closure law, the schedule on which the gate shuts after a load rejection, was treated as an optimal gate closure schedule in 1973 [9], and three-stage closure schemes are now optimized for pumped storage plants [10].

> ⧖ HISTORY **Water Turbines Begin to Make Electricity, 1880 to 1909**
>
> In 1880 a water turbine drove a dynamo that lit a theater and a storefront in Grand Rapids, Michigan, and in 1881 a turbine in a flour mill at Niagara Falls powered street lights. Both ran on direct current, which serves only short distances. Alternating current carried power farther, and the Redlands plant in California, the first commercial hydropower facility in the United States, came online in 1893 [18]. In the arid West, the Bureau of Reclamation first put small generators to work building the Theodore Roosevelt Dam in Arizona, lifting stone blocks into place, and sold the surplus to the nearby community. A 4500 kW plant followed, and by 1909 five generators were pumping irrigation water and supplying the Phoenix area [19].

7.3 The Unit as Modeled

The simulated plant is one generic 100 MW Francis unit with synthetic teaching values, shown in Figure 7.1 in the same arrangement as the application's Plant view. Water leaves a reservoir 100 m above the tailwater through a headrace tunnel, passes a surge tank, and falls through a 400 m penstock

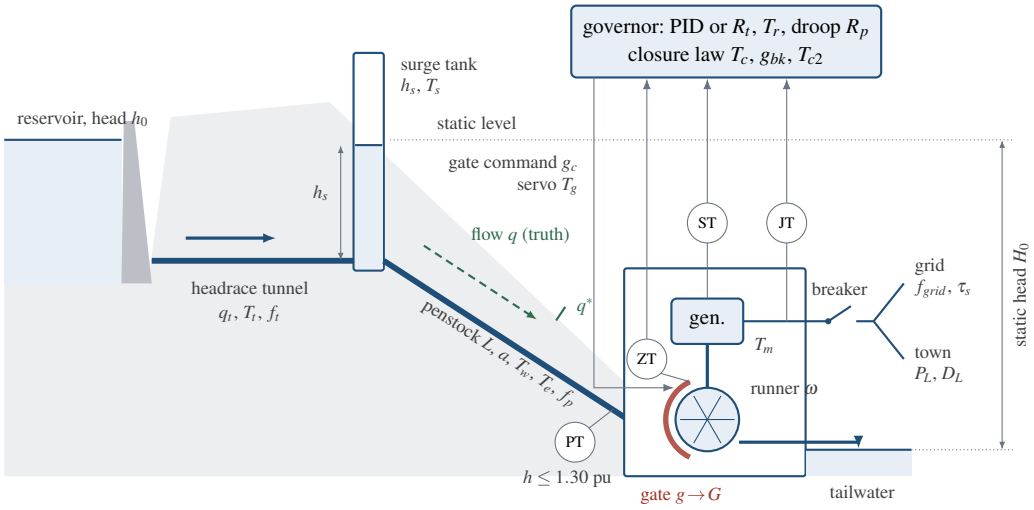


Figure 7.1 Hydroelectric unit as modeled. The reservoir at head h_0 feeds the headrace tunnel (flow q_t, starting time T_t, loss f_t) and the surge tank (head h_s, filling time T_s), which feeds the penstock of length L, wave speed a, water starting time T_w, travel time T_e and loss f_p. The penstock flow q is teaching truth (dashed), and the tick marks q^*, the flow the present gate passes once the column has caught up. The servo moves the stroke g, the ideal gate is G, the runner turns at ω with mechanical starting time T_m, and the breaker connects the generator to the grid or to the town load P_L with frequency sensitivity D_L. PT reads the inlet head h against its 1.30 pu teaching limit, and the speed (ST), gate (ZT) and power (JT) measurements feed the governor and its closure law. Not to scale.

to the turbine's wicket gates. The runner drives a synchronous generator whose breaker connects it either to a stiff grid or to an islanded town. The governor reads the speed, the gate position and the electrical power, and moves the gate through a hydraulic servo. The penstock flow has no meter in the plant and appears in the application only as teaching truth.

> ✦ DID YOU KNOW? **A Powerhouse of 6.8 Million Kilowatts**
>
> A federal pamphlet names Grand Coulee Dam on the Columbia River as the largest single electrical generating complex in the United States, with a capacity of more than 6.8 million kilowatts. A modern hydroelectric plant turns about 90 % of the energy of its falling water into electricity [19].

Every quantity is in per unit of the static head $H_0 = 100\,\text{m}$, of the flow $Q_0 = 115\,\text{m}^3/\text{s}$ through the fully open gate at that head, of the generator rating of 100 MW and of the synchronous speed at 60 Hz. Table 7.1 classifies the variables. At the design point of the island cases, the town draws 80 MW, the gate stands at 65.6 % of its stroke and the inlet head is 0.964 pu, or 9.45 bar above the tailwater.

7.3.1 Water Column and Turbine

The rigid penstock is a column of water accelerated by the difference between the surge tank's head and the head at the turbine, and the turbine passes flow in proportion to the gate and the square root of its head, as shown in Equation (7.1).

$$T_w \frac{\mathrm{d}q}{\mathrm{d}t} = h_s - h - f_p\, q|q|, \qquad h = \left(\frac{q}{G}\right)^2 \tag{7.1}$$

where $T_w = LV_0/(gH_0) = 2.0\,\text{s}$ is the water starting time at the full-gate velocity $V_0 = 4.905\,\text{m/s}$, so that the penstock's area is $23.45\,\text{m}^2$ and its diameter 5.46 m, and $f_p = 0.02$ is its head loss at full flow.

Table 7.1 Variables by Role, With Their Values at 80 MW on the Island

Variable	Role	Value	Range
f	CV, speed probe	60.00 Hz	band 58–62 Hz
P_e	CV on the grid, power transducer	80 MW	dispatch 0–100
g	MV, gate stroke	65.6 %	0–100 %
P_L	DV, town load	80 MW	0–100 MW, +10 % step
f_{grid}	DV, grid frequency	60 Hz	59.5–60.5, −0.2 Hz event
Δh_0	DV, reservoir level	0 m	−20 to 10 m
breaker	DV, load rejection	closed	opens on the event
$\Delta\omega$	state, speed deviation	0	—
q, q_t	states, penstock and tunnel flow	0.778 pu	truth only
h_s	state, surge tank head	0.976 pu	tank level, measured
q_i, h_i	states, elastic sections	mirrored	10 flows, 10 heads
$\Delta\omega_g$	state, grid speed	0	—
G	algebraic, ideal gate	79.2 %	—
h	algebraic, inlet head	0.964 pu	limit 1.30 pu
P_m	algebraic, turbine power	80 MW	—

The ideal gate G is the effective flow area. It follows the servo's stroke g through a declared concave characteristic, and the turbine's mechanical power follows the working group's form [2] as public lecture notes reproduce it (the working group's report itself was not read), as shown in Equation (7.2).

$$\begin{aligned} G &= \max\!\left(\frac{g(1+c)}{1+cg},\; G_{leak}\right) \\ P_m &= A_t\, h\,(q - q_{nl}) - D_t\, G\,\Delta\omega \\ A_t &= \frac{1}{q_1^2(q_1 - q_{nl})}, \qquad q_1 = \frac{1}{\sqrt{1+f_t+f_p}} \end{aligned} \tag{7.2}$$

where $c = 1$ makes a percent of stroke cut more flow near closed than near open, $G_{leak} = 0.01$ is the leakage of the closed wicket gates, $q_{nl} = 0.08$ is the flow that only spins the runner, and $A_t = 1.1893$ makes the power exactly 1 pu at full gate on the static head with both losses included. The term $D_t G\Delta\omega$ with $D_t = 1.0$ is the turbine's loss of power with speed, chosen so that the runaway speed at full gate is near 2 pu. The curvature c and D_t are teaching values, not a measured wicket-gate curve or runner characteristic.

7.3.2 Rotor, Network and Surge Tank

The rotor integrates the mismatch between turbine and electrical power, and the electrical power depends on what the breaker connects, as shown in Equation (7.3).

$$\begin{aligned} T_m \frac{\mathrm{d}\Delta\omega}{\mathrm{d}t} &= P_m - P_e - D\,\Delta\omega \\ P_e &= P_L\,(1 + D_L\Delta\omega) \quad \text{(island)}, \qquad 0 \quad \text{(open)} \\ P_e &= P_m - D\Delta\omega + \frac{T_m}{\tau_s}(\Delta\omega - \Delta\omega_g) \quad \text{(grid)} \\ T_{g,grid}\frac{\mathrm{d}\Delta\omega_g}{\mathrm{d}t} &= \frac{f_{grid}}{60} - 1 - \Delta\omega_g \end{aligned} \tag{7.3}$$

where $T_m = 2H = 8\,\mathrm{s}$, the unit damping D is zero, the town's load rises 1 % for each percent of frequency ($D_L = 1$), and on the grid the rotor is pulled to the grid's speed in $\tau_s = 0.1\,\mathrm{s}$ while the grid follows its own frequency setting with $T_{g,grid} = 2\,\mathrm{s}$. The grid case therefore omits the electromechanical swing of the machine against the network. Chapter 8 models the frequency of a whole grid with finite inertia, where droop and load damping set the offset. The surge tank and the

headrace tunnel close the hydraulic circuit, as shown in Equation (7.4).

$$T_s \frac{\mathrm{d}h_s}{\mathrm{d}t} = q_t - q, \qquad T_t \frac{\mathrm{d}q_t}{\mathrm{d}t} = h_0 - h_s - f_t\, q_t |q_t| \tag{7.4}$$

where $T_s = A_s H_0 / Q_0 = 250\,\mathrm{s}$ is the tank's filling time (a tank of about $290\,\mathrm{m}^2$), $T_t = 10\,\mathrm{s}$ is the tunnel's starting time, $f_t = 0.04$ its head loss and h_0 the reservoir head, 1 unless the reservoir slider moves it. At a fixed gate the tank and tunnel oscillate with a period near $2\pi\sqrt{T_s T_t} = 314.2\,\mathrm{s}$.

7.3.3 Gate Servo and Closure Law

The servo follows the governor's command with a first-order lag, opens no faster than 10 % of stroke per second, and closes no faster than the closure law allows, as shown in Equation (7.5).

$$\begin{aligned} \frac{\mathrm{d}g}{\mathrm{d}t} &= \mathrm{clip}\Big(\frac{g_c - g}{T_g},\ -r_c(g),\ r_{open}\Big) \\ r_c(g) &= \begin{cases} 1/T_{c2} & \text{two slopes and } g < g_{bk} \\ 1/T_c & \text{otherwise} \end{cases} \end{aligned} \tag{7.5}$$

where $T_g = 0.2\,\mathrm{s}$, $r_{open} = 0.10\,\mathrm{pu/s}$, T_c is the time the fast slope would take for a full stroke (6 s as shipped) and the cushion below the breakpoint g_{bk} closes a full stroke in T_{c2}. The stroke is held inside 0 to 1.

7.3.4 Elastic Penstock

An advanced switch replaces the rigid column with ten lumped sections, a staggered method of lines on the water hammer equations with a flow in each section and a head at each node, as shown in Equation (7.6).

$$\begin{aligned} T_w \Delta\xi\, \frac{\mathrm{d}q_i}{\mathrm{d}t} &= h_{i-1} - h_i - f_p \Delta\xi\, q_i |q_i|, \quad h_0 \equiv h_s \\ \frac{T_e^2}{T_w} \Delta\xi\, \frac{\mathrm{d}h_i}{\mathrm{d}t} &= q_i - q_{i+1}, \quad i < N \\ \frac{T_e^2}{T_w} \frac{\Delta\xi}{2}\, \frac{\mathrm{d}h_N}{\mathrm{d}t} &= q_N - G\sqrt{h_N} \end{aligned} \tag{7.6}$$

where $N = 10$, $\Delta\xi = 1/N$ and $T_e = L/a = 0.4\,\mathrm{s}$ for a wave speed of 1000 m/s. A closure faster than $2T_e = 0.8\,\mathrm{s}$ then raises the head by the Joukowsky amount $Z\Delta q$ with $Z = aV_0/(gH_0) = T_w/T_e = 5.0$. The rigid column's sections are mirrored from its flow and a linear friction line, so the switch can be made during a run.

7.3.5 Measurements, Initialization and Numerics

The governor sees seven instruments, read once per 20 ms scan: the speed probe, the bus frequency on the grid side of the breaker, the gate position, the generator power, the inlet pressure, the surge tank level and the head pond level. Each carries band-limited noise from a first-order filtered Gaussian with a 0.05 s correlation time, with standard deviations of 10^{-4} pu for the two speeds, 5×10^{-4} of stroke for the gate, 0.2 MW for power, 2×10^{-3} pu for pressure and 0.02 m for both levels, scaled by an instrument-noise slider. The penstock and tunnel flows are logged only in columns prefixed `truth_`, and a test fails if the controller source reads anything but the measurement packet.

Every case starts lined out. In automatic mode a bisection finds the gate that carries the town's load, or the grid dispatch, at 60 Hz, and in manual mode the speed is set to the deviation at which the manual gate balances the town. The largest derivative at the initial state is 3.2×10^{-17}. The 26 states (five plant states, twenty elastic section states and the grid's speed) advance by fourth-order Runge-Kutta steps of $\Delta t = 10\,\mathrm{ms}$, the governor executes every 20 ms, and a row is logged every 0.1 s, in the engine of Section 1.6.1. A load rejection and a breaker closure are applied once, after a step,

Table 7.2 Model Parameters. A: Adjustable in the Parameters Panel, F: Fixed, H: Hidden, D: Derived

Symbol	Meaning	Value	Unit	
H_0, Q_0	static head, full-gate flow	100, 115	m, m^3/s	F
L, a	penstock length, wave speed	400, 1000	m, m/s	F
T_w	penstock water starting time	2.0	s	A
T_e	wave travel time L/a	0.4	s	A
f_p	penstock head loss at $q=1$	0.02	pu	A
T_t, f_t	tunnel starting time, head loss	10, 0.04	s, pu	A
T_s	surge tank filling time	250	s	A
q_{nl}	no-load flow	0.08	pu	A
D_t	turbine damping	1.0	pu	A
c	gate curvature	1.0	—	A
G_{leak}	closed-gate leakage	0.01	pu	H
A_t	turbine gain	1.1893	—	D
T_g, r_{open}	servo lag, opening limit	0.2, 0.10	s, pu/s	A
T_m	mechanical starting time $2H$	8	s	A
D, D_L	unit damping, load sensitivity	0, 1.0	pu	A
τ_s	grid coupling time	0.1	s	H
$T_{g,grid}$	grid frequency lag	2	s	F
N	elastic switch, sections	off, 10	—	A
h_{max}	inlet head limit	1.30	pu	F
ω_{max}	overspeed limit	1.50	pu	F

with a mark on the trends. The pressure and overspeed limits are annunciated rather than tripped, so that the whole peak a closure law produces is visible.

Table 7.2 lists every parameter. The values make a consistent design point for a generic unit and are not fitted to any plant.

7.3.6 Assumptions

The column is rigid unless the elastic switch is on, which is accurate for the governor's seconds-long motions and for closures slower than $2L/a$. The turbine passes $q = G\sqrt{h}$ at every speed, so the Francis runner's S-characteristic, its efficiency hill and the draft tube are absent, and the overspeed after a rejection is a teaching estimate rather than a prediction for a real runner. The gate characteristic is a declared smooth curve whose concavity gives a cushioned closure a reason to exist. The surge tank has no throttle and cannot overflow, and the reservoir and tailwater are fixed except for the reservoir slider. The stiff grid removes the generator's 1 to 2 Hz swing against the network, voltage and excitation. The pressure limit of 1.30 pu (12.75 bar), the overspeed limit of 1.50 pu and the island band of 60 ± 2 Hz are teaching values, and a real plant takes its limits from the penstock's design pressure, the generator's overspeed rating and the frequency standard it operates under, with the governing system specified as the IEEE guide describes [3].

7.4 Why the Water Limits the Governor

7.4.1 The Right-Half-Plane Zero

Linearizing Equations (7.1) and (7.2) at a constant head upstream, with the ideal gate as the input and the mechanical power as the output, as Section 1.4.3 linearizes a tank, gives Equation (7.7).

$$\frac{\Delta P_m}{\Delta G} \propto \frac{h - 2f_p q(q - q_{nl}) - (q - q_{nl})T_w s}{T_w s + 2h/q + 2f_p q} \tag{7.7}$$

With $f_p = q_{nl} = 0$ at rated flow this is the textbook ideal turbine $(1 - T_w s)/(1 + T_w s/2)$ of Section 1.4.7, whose zero lies at $+1/T_w$ and whose pole lies at $-2/T_w$, so the first response to a gate step is -2 times the final one. The design plan for the application placed the zero at $2/T_w$, which is the pole's magnitude, and the application follows the derivation instead. On the application's plant the framework's eigenvalue routine, applied to the linearized column, returns the zero of Equation (7.7)

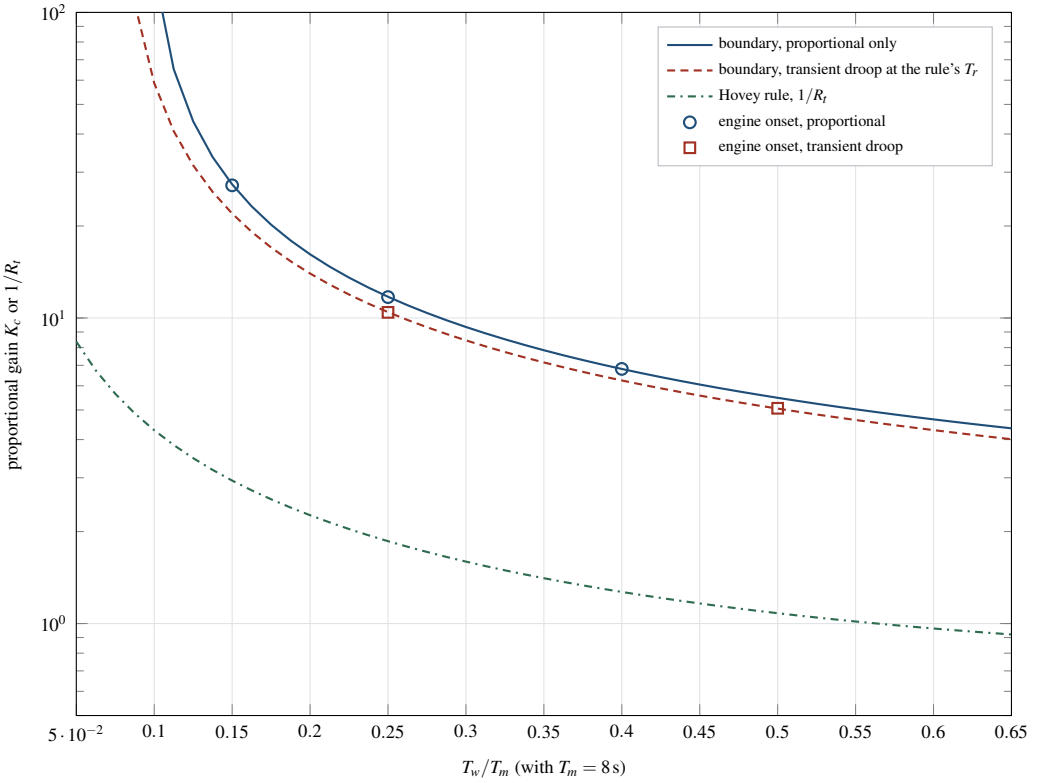


Figure 7.2 Island stability boundary against T_w/T_m at 80 MW with 5 % permanent droop, from the application's map function: the proportional governor (solid), the transient-droop governor at the rule's own reset time at each T_w (dashed) and the gain $1/R_t$ of the rule (dash-dotted). The circles (proportional) and squares (transient droop) are the engine's oscillation onset found by bisection.

within 10^{-6} at every load tested: 1.143, 0.675 and 0.510 rad/s at 50, 80 and 100 MW, with poles at -1.952, -1.254 and -0.991 rad/s. The first response is therefore 1.71 to 1.94 times the final one, and of the opposite sign. Section 7.6.1 shows the dip on the running plant. With the surge tank and tunnel left active the zero at 80 MW moves only to 0.672 rad/s and a slow pair appears near $\pm 0.020i$ rad/s. Doubling T_w to 4 s halves the zero to 0.337 rad/s at 80 MW.

Pause. Try Exercise 7.1 before reading on.

7.4.2 The Gain Limit and the Stability Map

The application's stability map linearizes the island loop by central differences of its own right-hand side (rigid column, servo without rate limits, rotor, town load and the governor in continuous form) with the surge tank held at its steady head, and computes the eigenvalues with the framework's balanced Hessenberg QR routine. The map plots the damping of the least-damped mode over the proportional gain and T_w/T_m, with T_m held and T_w varied, at the load the town draws at the frequency the speed probe has averaged over the last 30 s. The boundary falls steeply with the ratio, as shown in Figure 7.2. For a proportional governor with 5 % droop at 80 MW the boundary gain is 27.4 at $T_w/T_m = 0.15$, 11.73 at 0.25 and 6.81 at 0.40, and the engine's own oscillation onset, found by bisection on the growth rate of a small disturbance with the noise off, lies within 0.75 % of each.

The boundary also depends on the operating point. After case 2's load step a gain of 4 leaves the island near 58.7 Hz, where the town draws about 86 MW instead of the 88 MW it would draw at

60 Hz. At the load the map then uses, the boundary reads 13.04 at 30 s while the averaging window still contains pre-step data, and 12.89 once that window is entirely post-step, against 13.36 at 88 MW and 60 Hz and 11.73 at 80 MW, and the engine's onset is 13.14. The map is a linear prediction of where a small swing begins to grow. It ignores the servo's rate limits and the tank, so it says nothing about the size of a limit cycle or the gain at which a large one dies, which is why the application asks for a margin of about 10 %. A stability boundary that moves with the operating point is also the subject of Volume 7, Chapter 13, where the depth of cut at which a mill chatters moves with the spindle speed.

🕮 **Pause.** Try Exercise 7.2 before reading on.

7.4.3 Transient Droop and the Rule

The transient-droop governor of Section 7.5 acts first with a gain of $1/(R_p + R_t)$ and relaxes over the reset time T_r to its permanent droop. The rule the application suggests is Equation (7.8), which Kundur gives after Hovey [4, 13].

$$R_t = \big[2.3 - 0.15(T_w - 1)\big]\frac{T_w}{T_m}, \qquad T_r = \big[5 - 0.5(T_w - 1)\big]T_w \tag{7.8}$$

The equation was not checked against a primary copy. Kundur's book has no open copy and Hovey's paper is behind a paywall, so the form was taken from two editions of a public university lecture that quote it as Kundur's equations 9.41 and 9.42, and its worked example ($T_w = 1.3$ s, $T_m = 8$ s, giving 0.366 and 6.31 s) is reproduced by the application's function. Undrill and Woodward later proposed an improved calculation of the temporary droop [14], so the application calls Equation (7.8) a teaching rule and nothing more. At the design values it gives $R_t = 53.75$ % (54 % on the slider) and $T_r = 9$ s, and at $T_w = 4$ s it gives 92.5 % and 14 s. The rule sits far inside the linear boundary. At $T_w/T_m = 0.25$ its gain $1/R_t = 1.86$ is 5.6 times below the transient-droop boundary of 10.45 at the same reset time, and at 0.5 the factor is 4.7, as shown in Figure 7.2.

7.4.4 Water Hammer and the Closure Law

Stopping the column raises the head at the gate. For a closure faster than $2L/a = 0.8$ s the rise is the Joukowsky value, $aV_0/g = 500$ m for the full velocity, five times the static head. The governor's closures last seconds, so the rigid column applies and the rise is roughly T_w times the rate at which the flow is cut. Closing slowly lets the water drive the runner with no load, so the closure time trades the inlet pressure against the overspeed. The concave gate characteristic makes the trade uneven along the stroke, because near full opening a percent of stroke cuts little flow and near closed it cuts a lot, and a two-slope law closes fast while the gate is wide open and slowly in the cushion below a breakpoint.

7.4.5 The Surge Tank Under a Governor

When the tank's level falls, the head at the turbine falls with it. A governor that holds the power then opens the gate and draws still more water from a tank that is already low, which is negative damping on the tank's slow mode. Thoma's stability area answers how large a tank must be for that mode to decay anyway [11]. In per unit it is $T_{s,th} = T_t/(2 f_t h_s)$, 128.1 s at 80 MW and 128.8 s at 88 MW, against the plant's 250 s. The linearized closed loop of this plant, with the rule's transient droop and no permanent droop at 88 MW, puts the boundary lower, at $T_s = 85.2$ s. Removing terms one at a time explains the gap. With $D_L = 0$ the boundary rises to 112.6 s, with $D_t = 0$ to 112.4 s, with both zero to 130.4 s (close to Thoma's value), and with $q_{nl} = 0$ to 95.2 s, while $f_p = 0$ lowers it to 81.1 s. Thoma's derivation assumes the power is held exactly, while here the town's and the turbine's frequency sensitivity and the no-load flow let the power give way over the tank's three-minute period, and the penstock's own head loss takes a little of that damping back.

7.5 The Governors as Implemented

Both governors execute every $\Delta t_c = 20\,\text{ms}$ on the measurement packet and act on a load reference g_{ref}, the gate at which the unit runs at its speed setpoint. The permanent droop R_p acts on the measured gate position, as it does in the dashpot governor, with a default of 5 %. The form that is not in use tracks the gate command every scan, so switching forms and switching between Manual and Auto are bumpless.

7.5.1 PID Governor

The PID governor is the framework's velocity-form PID of Section 1.6.2, run on the per-unit speed with the droop folded into the measurement, as shown in Equation (7.9).

$$\begin{aligned} PV_k &= \Delta\omega_{m,k} + R_p\,(g_{m,k} - g_{ref}), \qquad e_k = \Delta\omega_{SP} - PV_k \\ P_k &= K_c e_k, \qquad D_k = -K_c \tau_D \dot{PV}_{f,k} \\ s_k &= s_{k-1} + (P_k - P_{k-1}) + (D_k - D_{k-1}) \\ \hat{u}_k &= u_0 + s_k + I_{k-1}, \qquad g_{c,k} = \min(1, \max(0, \hat{u}_k)) \\ I_k &= I_{k-1} + \left[\frac{K_c}{\tau_I} e_k + \frac{g_{c,k} - \hat{u}_k}{\tau_I}\right]\Delta t_c \end{aligned} \tag{7.9}$$

where $\Delta\omega_{SP} = f_{SP}/60 - 1$ and the derivative is taken on the measurement through a 0.1 s filter. The running sum and the back-calculated integral are those of Equations (1.25) to (1.27), with $s_k = S_k$, $\hat{u}_k = v_k$, $g_{c,k} = u_k$, $u_0 = u_b$ and the tracking time constant equal to τ_I. An integral time of zero removes integral action. The loop is direct acting in this sign convention, so a speed below the setpoint opens the gate. The defaults are $K_c = 3$, $\tau_I = 3\,\text{s}$ and $\tau_D = 0$.

7.5.2 Transient-Droop Governor

The transient-droop governor is the electronic equivalent of the dashpot governor, as shown in Equation (7.10).

$$\begin{aligned} g_{c,k} &= \text{clip}\left(\frac{(\Delta\omega_{SP} - \Delta\omega_{m,k}) + R_p\, g_{ref} + R_t\, z_k}{R_p + R_t},\ 0,\ 1\right) \\ z_{k+1} &= z_k + \frac{\Delta t_c}{T_r}\,(g_{c,k} - z_k) \end{aligned} \tag{7.10}$$

where z is the dashpot state. A speed error first moves the gate by $1/(R_p + R_t)$ per unit, and z then follows the gate so that the action grows over T_r toward the permanent droop $1/R_p$. With $R_p = 0$ the law is a PI controller with $K_p = 1/R_t$ and $\tau_I = T_r$, which is why the stability map plots $1/R_t$ for this form. With an ideal servo the structure equals the textbook dashpot loop exactly, and the transfer function $(1 + sT_r)/[R_p(1 + sT_r R_t/R_p)]$ that the lecture notes quote from Kundur is its approximation for $R_t \gg R_p$. The defaults are the rule's 54 % and 9 s.

7.5.3 Power Controller, Sequencer and Synchronizer

On the grid an optional power controller integrates on the load reference so that the unit follows its dispatch P_{SP} and keeps its droop share of a grid event, as shown in Equation (7.11).

$$g_{ref,k+1} = \text{clip}\left(g_{ref,k} + \Delta t_c K_{pc}\left[\frac{P_{SP}}{100} + \frac{e_k}{R_p} - \frac{P_{m,k}}{100}\right],\ 0,\ 1\right) \tag{7.11}$$

where P_m here is the measured generator power in megawatts, e_k/R_p is the frequency bias and $K_{pc} = 0.05\,\text{pu}$ of gate per per-unit power per second. A load-rejection sequencer acts in both modes because it is protection. When the breaker opens it commands the gate closed, and the servo closes along the closure law of Equation (7.5). Once the speed has peaked and fallen below 1.05 pu, it hands back with the load reference at the no-load gate for the measured reservoir head. A Synchronize helper closes the breaker only when the speed is within 1 % of the network's, and sets the load reference to the gate that carries the present load.

Table 7.3 The Nine Cases as Shipped

#	Setting and scripted events	Teaching point
1	Manual on the island, gate 65.6 %	inverse response, overspeed
2	Proportional, $K_c = 4$, load +10 % at 10 s	gain limit, the map
3	Transient droop 10 %, 3 s, load step at 10 s	the rule, why hydro is slow
4	Grid, PID 8 and 2 s, grid −0.2 Hz at 20 s	droop share, power control
5	Rule, 88 MW, rejection at 10 s, single 6 s	closure time and law
6	Rule with $R_p = 0$, $T_s = 60$ s, load step	surge tank instability
7	Rule, load step at 10 s	nadir and settling
8	$T_w = 4$ s, droop 25 %, 3 s, load step	a longer penstock
9	Challenge: grid PID, single 6 s, step at 10, rejection at 90 s	the whole problem

Table 7.4 Measured Results. Frequencies in Hz, Heads and Speeds in Per Unit

#	Setting	Quantity	Value
1	gate 65.6 to 75 % at 2 s	power dip, at	4.38 MW, 0.74 s
		peak f, f at 120 s	62.30, 61.96
2	map, engine at 80 MW	boundary gain	11.73, 11.70
	$K_c = 14$ at 30 s	limit cycle	4.6–5.5 p-p, 10.0 s
3	10 %, 3 s (shipped)	limit cycle, last 30 s	5.26 p-p
3, 7	rule, 54 %, 9 s	nadir, settle ±0.3	58.40, 28 s
	30 %, 4 s	nadir, settle ±0.3	58.52, 8.3 s
4	5 % droop, 40 s after	gate, power	6.65 %, 4.65 MW
	10 % droop	gate, power	3.33 %, 2.53 MW
	5 %, power control	peak, at 900 s	7.29 MW, 6.78 MW
5	single 6 s	head, speed	1.787, 1.366
	single 15 s	head, speed	1.277, 1.604
	two slopes 6/40 %/15 s	head, speed	1.286, 1.439
6	Auto, swing peak to trough	236 s, 836 s	6.28 m, 11.06 m
	Manual from 491 s	493 s, 877 s	8.01 m, 0.79 m
8	shipped 25 %, 3 s	range	52.57–65.52
	rule for 4 s, rule for 2 s	settle ±0.3	116 s, 41 s

7.5.4 Plant, Map and Governor Models

The three models of Section 1.4.9 take a particular form here. Section 7.3 defines the plant. The stability map is a second model, the island loop linearized with the tank frozen and the servo's limits removed, and it is used only for display, never by a controller. Neither governor carries an internal model. The gate-to-power slope of 0.74 at 80 MW and the gate-to-power zero are properties of the plant that a learner identifies, not parameters the governors use.

7.6 Nine Cases, Measured

The cases appear in Table 7.3, and Section 1.6.4 describes how a case, its completion rule and the challenge work in every app. The measurements below were taken without a display, on the same engine and seed the browser uses and with the same fixed steps, and any change a sentence mentions is made at the time it states, once the run is under way. A headless run of the case regenerates each value. Table 7.4 collects the headline measurements.

7.6.1 The Gate by Hand

Case 1 holds the gate at 65.6 % in Manual with the town at 80 MW. Raising it to 75 % at 2 s drops the turbine power by 4.38 MW, with the minimum 0.74 s after the move, and the power is back above 80 MW 1.37 s after it. Figure 7.3 shows the plant during the dip. This is the inverse response of Section 7.4.1 on the running plant. The surplus then spins the rotor up. Nothing but the town's frequency sensitivity and the turbine's own damping pulls the speed back, so the frequency passes 62 Hz 10.9 s after the move, peaks at 62.30 Hz and reads 61.96 Hz at 120 s. Over the following minutes the surge tank's swing carries it between 62.30 and 62.38 Hz. The case's completion rule

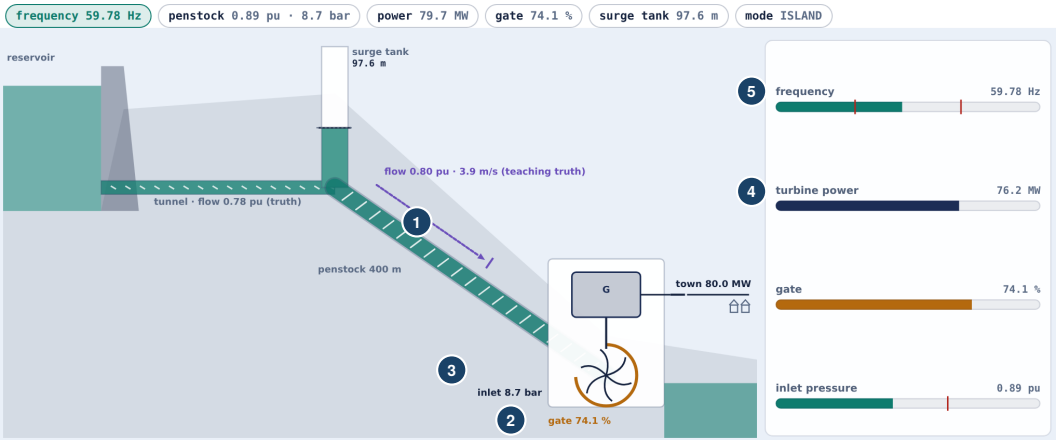


Figure 7.3 The inverse response of case 1 on the Plant view, cropped from the application window, at $t = 2.9$ s (the clock is in the cut header), 0.9 s after the gate station was raised from 65.6 to 75 % in Manual on the island. The servo has reached 74.1 % (2), but the penstock flow (1, dashed, teaching truth) is still 0.80 pu at 3.9 m/s. The tick ahead of its arrow marks the flow the new opening will pass once the column has caught up, 0.83 pu, a value the figure does not print. The inlet pressure has fallen to 8.7 bar (3), the turbine gives 76.2 MW (4) against the 79.7 MW the town draws at that frequency, and the frequency has dipped to 59.78 Hz (5) before the faster water reverses it. The power had passed a minimum of 75.6 MW at 2.74 s, which the cropped view does not show and Table 7.4 gives as a dip of 4.38 MW 0.74 s after the move.

asks for the gate at 74 % or more after Run, then the turbine power below 79 MW, then the frequency above 62 Hz. The gate is moved after Run because a setting changed before Run becomes part of the initial state instead of a step.

> **TRY IT IN THE APP** **1 · Open the Gate by Hand on the Island**
>
> Load the case and press Run. Once the clock passes 2 s, raise Gate opening from 65.6 to 75 % and watch the turbine power on the Power strip. It falls first, by 4.38 MW at its lowest, and is back above 80 MW only 1.37 s after the move. Then keep watching the frequency: with nobody governing it, it climbs past 62 Hz 10.9 s after the move.

7.6.2 The Gain Limit

Case 2 runs a proportional governor with a gain of 4 and 5 % droop through a 10 % load step. Raised to 14 at 30 s, the gain sits above the settled post-step boundary of 12.89, and the island settles into a limit cycle with a period of 10.0 s whose amplitude wanders between 4.6 and 5.5 Hz peak to peak from 60 to 600 s as the surge tank moves the head. The servo's rate limits hold it there. A gain of 13, about 1 % below the engine's onset of 13.14, decays at first (0.74 Hz peak to peak at 160 to 180 s) and then grows into the same limit cycle (4.73 Hz at 580 to 600 s), while 12.5 dies away to 0.02 Hz. Recovering from a running limit cycle needs more margin than the onset. With the gain lowered from 14 at 90 s, the swing falls below 0.2 Hz peak to peak in 110 s at 12 and 60 s at 11, and never at 13 or 12.5, and lowered at 200 s the times are 170 s and 75 s.

7.6.3 Transient Droop, the Nadir and a Longer Penstock

Case 3 ships a transient droop of 10 % with a 3 s reset, the way a steam governor would be set. The boundary at that reset time is $R_t = 12.4$ %, so the island limit-cycles after the load step, 5.26 Hz peak

to peak over the last 30 s of 120 s. The rule's 54 % and 9 s give a nadir of 58.40 Hz 5.4 s after the step with no overshoot, and the frequency stays within 0.3 Hz of its value at 120 s from 28 s after the step. A faster 30 % and 4 s reach a nadir of 58.52 Hz and settle in 8.3 s. The nadir barely moves because it is set in the first five seconds by the rotor and the water. Over 366 calm PID and transient-droop tunings the best nadir is 58.87 Hz, from a PID with a gain of 3, an integral time of 5 s and a derivative time of 2 s, and with the gate held fixed the frequency falls to 57.13 Hz by 15 s. Case 7 repeats the rule's run and adds that over 40 to 900 s the surge tank moves the island between 59.33 and 59.78 Hz, inside the settling band around the 120 s value of 59.61 Hz.

Case 8 doubles T_w to 4 s. The tuning that settled the 2 s penstock in 11.7 s, a transient droop of 25 % with a 3 s reset, now sits beyond a boundary of 35.0 % and drives the island between 52.57 and 65.52 Hz. The rule for 4 s settles within 0.3 Hz in 116 s, and the rule for 2 s still settles, in 41 s. A faster tuning on one water column tolerates less of another.

> ✦ DID YOU KNOW? **A Faster Governor Barely Lifts the Nadir**
>
> From the simulator: after the 10 % load step on the island, the rule's transient droop of 54 % with a 9 s reset settles in 28 s, and a faster 30 % with 4 s settles in 8.3 s. The lowest frequency moves only from 58.40 to 58.52 Hz (Table 7.4).

7.6.4 Droop on the Grid

Case 4 synchronizes the unit to the grid at 80 MW with the grid PID, and the grid falls 0.2 Hz at 20 s. With 5 % droop the gate opens 6.65 % of stroke against the $(0.2/60)/0.05 = 6.67\,\%$ the droop asks for, and 40 s after the event the power is up 4.65 MW, not the 6.67 MW the ratio would give in power, because the droop acts on the gate and the turbine's gate-to-power slope at 80 MW is 0.74. A 10 % droop halves the gate move to 3.33 % and gives 2.53 MW. With the power controller on from the start at 5 %, the power carries the droop's share itself, overshooting to 7.29 MW at 239 s and reading 6.78 MW at 900 s, with the surge tank still moving it between 6.41 and 6.99 MW over the last five minutes.

7.6.5 Load Rejection and the Closure Law

Case 5 opens the breaker at 88 MW, where the penstock carries 98.2 m^3/s at 4.19 m/s. A single 6 s closure raises the inlet head to 1.787 pu (17.5 bar) at 14.5 s while the speed peaks at 1.366 pu. A 15 s closure holds the head at 1.277 pu but lets the speed reach 1.604 pu. No single slope between 1 and 40 s meets both teaching limits at this load, and the closest with the head under 1.30 pu, 14 s, overspeeds to 1.583 pu. The two-slope law the case suggests (6 s above a 40 % breakpoint and a 15 s cushion below it) gives 1.286 pu and 1.439 pu. Halving the time step moves these peaks by less than 1×10^{-3} pu. The case's completion rule asks for the two-slope law and a run past the rejection with both peaks inside their limits. Figure 7.4 sets the three laws against the two limits.

> ⌖ TRY IT IN THE APP **5 · Load Rejection: Closure Time Against Pressure and Overspeed**
>
> Load the case and press Run. The breaker opens at 10 s, and the single 6 s closure drives the inlet head to 1.787 pu while the speed peaks at 1.366 pu. Switch Closure law to Two slopes, press Reset and run again: the head now peaks at 1.286 pu and the speed at 1.439 pu, both inside their limits.

From 100 MW the result changes in an instructive way. The servo offers one or two slopes, and on the plant equations, which match the engine to 1×10^{-3} pu at 88 MW, no single slope from 1 to 40 s in steps of 0.1 s meets both limits, and none of 56,610 two-slope laws (fast 1 to 12 s, breakpoint 5 to 95 %, cushion 2 to 80 s) does either. The closest two-slope law, 5.5 s, 42.5 % and 14 s, holds 1.297 pu with 1.541 pu of speed. The water and the rotor do not forbid it. Of 46,032 three-slope laws, 656 meet both limits, such as 4 s for a full stroke above 60 %, 8 s down to 25 % and 14 s below (1.296 pu and 1.491 pu), and a gate that held the head exactly at 1.30 pu would limit the overspeed to 1.383 pu. The

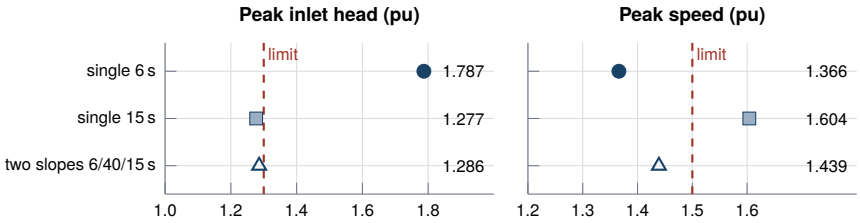


Figure 7.4 The closure laws of case 5 from Table 7.4, rejecting 88 MW: a single 6 s closure (filled circles), a single 15 s closure (squares) and the two-slope law with a 40 % breakpoint and a 15 s cushion (open triangles), with each value printed at the right. The dashed lines are the teaching limits. Notice that each single slope breaks one limit, the fast one the head and the slow one the speed, while the two-slope law stays under both.

infeasibility therefore belongs to the law family the servo offers, not to the plant, and the application says so. Optimal closure schedules were studied in 1973 [9] and three-stage closure schemes are optimized for pumped storage [10], so the counterexample is a measured property of this teaching plant that agrees with that practice.

The same computation sets the limits the application uses. The design plan proposed an overspeed limit of 1.2 pu, but at 88 MW the head-holding gate bounds the overspeed at 1.307 pu, so 1.2 pu with a 1.30 pu head is out of reach for any closure law at $T_w = 2$ s and $T_m = 8$ s. The limit was recalibrated to 1.50 pu, which 1188 of the 56,610 two-slope laws meet at 88 MW, the best at 1.425 pu. The design plan's island band of ± 1 Hz was widened to ± 2 Hz because the best of 366 calm tunings reaches only 58.87 Hz after the 10 % step.

7.6.6 The Surge Tank

Case 6 shrinks the tank to $T_s = 60$ s, an undamped period of $2\pi\sqrt{T_s T_t} = 154$ s, and runs the rule's governor isochronously. After the load step the level swing grows about 8 % each half cycle at a period near 171 s, from 6.28 m peak to trough at 236 s to 11.06 m at 836 s, which the linearized loop matches with a growth of 1.086 per half cycle and a period of 171.0 s. The case's completion rule asks for a swing larger than 8 m under the governor, first true at 491 s, followed by 300 s with the gate held. Switched to Manual at 491 s, the fixed gate passes less water as the head falls, and the swing decays from 8.01 m to 0.79 m by 877 s, about 40 % less each half cycle after the first. On the engine the boundary lies between $T_s = 70$ s, which grows, and 90 s, which decays, consistent with the linear 85.2 s of Section 7.4. The plant's own tank of 250 s decays under the same governor.

Pause. Try Exercise 7.3 before reading on.

7.6.7 The Challenge

The challenge islands the town at 80 MW, raises its load 10 % at 10 s and opens the breaker at 90 s. Its target, judged on the true speed and head every physics step, is 60 ± 2 Hz until the breaker opens, within 0.3 Hz of the final value from 40 s after the step, and a head at or under 1.30 pu and a speed at or under 1.50 pu over the 150 s. Table 7.5 lists four configurations. The verdicts do not change at seeds 1 through 5 or at a time step of 5 ms, and the logged rows are identical when the run is played at $1\times$ and at $60\times$.

7.6.8 Code Verification, Agreement and What Is Not Claimed

Code verification checks that the program solves its own equations. The integral of $P_m - P_e - D\Delta\omega$ over 60 s of case 7 equals T_m times the change in $\Delta\omega$ to 5.8×10^{-6} pu s. At fixed gates of 40, 65 and 90 % the engine's power equals Equation (7.2) evaluated at the column's steady flow within the test's tolerance of 0.1 %. A 2 % gate step at a fixed gate starts a tank oscillation with a period of 313.7 s

Table 7.5 The Challenge Over 150 s at the Default Seed

Configuration	f range (Hz)	settle (Hz)	head (pu)	speed (pu)	verdict
grid PID, single 6 s (shipped)	56.009–62.479	4.394	1.782	1.226	fails
rule, single 6 s	58.396–60.001	0.116	1.776	1.355	fails
PID 4, 3 s, two slopes	58.535–60.002	0.018	1.275	1.431	passes
rule, two slopes (suggested)	58.396–60.001	0.116	1.276	1.428	passes

against $2\pi\sqrt{T_s T_t} = 314.2$ s. On the elastic penstock a closure of 10 % of stroke in 0.3 s raises the head by 0.1383 pu against the Joukowsky estimate of 0.1380 pu from the turbine's own law, a difference of 0.13 %. An independent N-section solver gives errors of 7.0, 0.70, 0.35, 0.13, 0.14 and 0.16 % for 2, 3, 5, 10, 20 and 40 sections, unchanged at time steps of 5 and 2.5 ms, which is why the switch uses ten sections and not the two the design plan proposed. Case 7's frequency at 15 and 60 s is the same to 1×10^{-9} Hz at time steps of 10, 5 and 2.5 ms. A 900 s case costs 134 ms headless, and case 6 plays its fifteen minutes in about 15 s at 60× in a browser.

Agreement with published relations is a separate and weaker claim. With $f_p = q_{nl} = 0$ the application's linearization reproduces the ideal turbine's zero at 0.500 rad/s, pole at −1.000 rad/s and initial to final ratio of −2.000 at rated flow [2], the rule's worked example is reproduced as quoted, and Thoma's area and the Joukowsky constant follow from the declared geometry. No plant data were used, so nothing here validates the model against a real unit, and no learning gain is claimed because no learning study has been made.

7.7 Technology in Practice

The governor of Section 7.5 reads seven instruments and moves one gate servo. Figure 7.5 draws its loop with plant devices in place of the schematic's symbols.

What Measures It

A governor is built to see small changes of speed. In a mechanical governor the pilot valve is made very small so that it responds to the minimal forces of minor speed changes [20]. The gate position comes from a position transducer on the servomotor, calibrated so that it reads 0 to 100 % of stroke, and a lost transducer signal must raise an alarm at the controller [20]. At commissioning the speed is often read a second way, by a proximity probe on a temporary target on the shaft, and the two signals are trended together [21]. The chapter's speed probe, with its band-limited noise of 10^{-4} pu read every 20 ms, stands for the toothed wheel and its probes. No range, accuracy or response time is given here, because none was found in an open source that names no maker.

> **TECHNOLOGY** **Sensor: Speed From a Toothed Wheel**
>
> A mechanical governor senses speed with flyweights in a ball head, spun by a motor fed from a permanent magnet generator on the unit's shaft. A digital governor counts the pulses of magnetic probes facing a toothed wheel. The probe's air gap is set to the maker's specification, and metal filings on a probe are cleaned off [20]. A sensor that misses pulses is replaced, and a frequency generator checks the speed channels and the overspeed trip during pre-commissioning [21]. A damaged drive pin on the permanent magnet generator shows on a trend as gates that move a little every revolution while the grid holds the speed, called on-line hunting [20].

What Moves It

The gate servo of Equation (7.5) follows its command with a lag of 0.2 s, opens no faster than 10 % of stroke per second and closes along the closure law. A plant sets the closing rate with an orifice in the servomotor's hydraulic circuit, and some plants have more than one rate limiter in each direction [21], which is what the chapter's two-slope law with its cushion represents. The closing rate sets both the

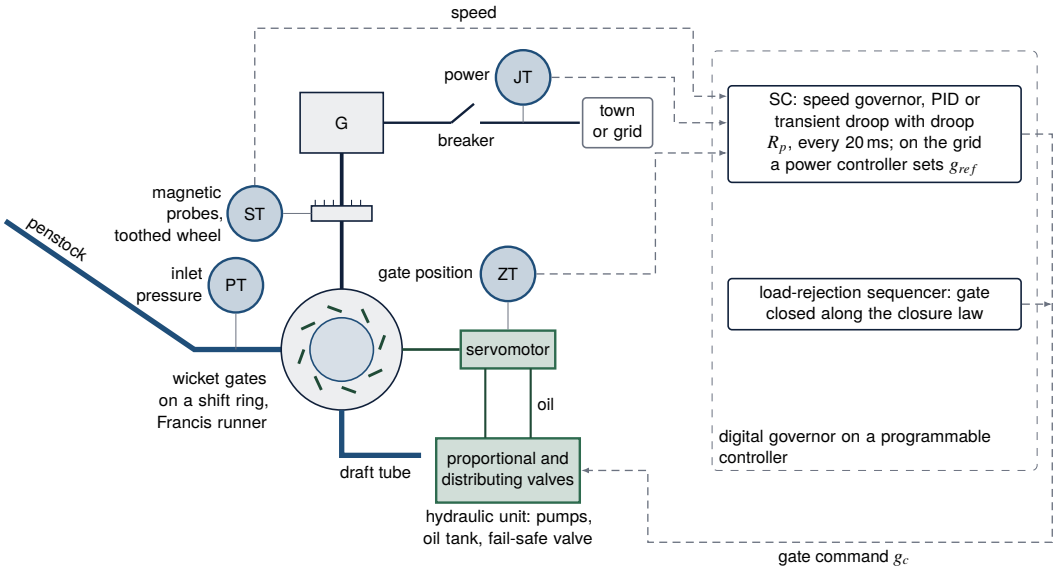


Figure 7.5 The governor loop of the chapter with plant devices in place of the schematic's symbols. ST reads the speed from a toothed wheel on the shaft, ZT the gate position on the servomotor, JT the generator power and PT the inlet pressure. On the digital governor the speed governor, or after a load rejection the sequencer, sets the gate command, which proportional and distributing valves turn into oil flow to the servomotor that turns the wicket gates. Notice that every path from the gate to the speed runs through the water column, the turbine and the rotor.

water hammer and the overspeed after a load rejection, and practice makes it as fast as the design pressure of the penstock or the spiral case allows [20].

> **TECHNOLOGY** **Final Element: Wicket Gates and Their Servomotors**
>
> Hydraulic servomotors drive a shift ring that turns all the wicket gates together. A small pilot valve, sensitive to minute forces, steers a large main distributing valve that meters oil to the servomotors. The oil tank is usually sized so that the gates can still close if the pumps fail. Near the end of the closing stroke the pistons cover their drain ports and a cushioning valve slows them, so that the gates do not slam shut. A digital governor adds a spring-loaded valve that closes the gates at the fastest allowed rate when its controller stops working [20].

Where the Controller Runs

A generating facility is run by a distributed control system that communicates with the grid's supervisory system, and its control loops run with cycle times from milliseconds to minutes [22]. The controller of a digital governor can be a programmable logic controller [20], and the chapter's 20 ms execution sits at the fast end of that range. The transient-droop form of Equation (7.10) began as a hydraulic dashpot whose temporary droop compensates the inertia of the unit and of the water column. On a large grid a bypass lets the dashpot re-center quickly so that load changes go faster, and when the plant is separated to carry a small system alone the bypass is not used [20]. That is the chapter's lesson that grid tuning fails on an island.

IN PRACTICE **Proving a Closure Law by Load Rejection**

The closing rate is set by an orifice in the servomotor circuit and timed from 25 to 75 % of stroke, then doubled, to leave out the cushion. At commissioning the breaker is opened at 25 % power, then at 50, 75 and 100 %, recording the peak speed and the peak penstock pressure each time. After each rejection IEEE Std 125 allows one 5 % underspeed deviation and at most one 5 % overspeed deviation. A pressure above its design limit stops the series, and one close to it goes to the designers before the next step [21].

What Is Modernizing It

Digital governors remove the ball head, the dashpot and the restoring linkage, though their proportional valves need cleaner oil [20]. A review of emerging hydropower technology estimates that digitalizing the world's 1225 GW of hydropower could add 42 TWh of production a year. Variable-speed units, fed through a power converter, control active and reactive power better, at an estimated extra cost of 7 to 15 % of a fixed-speed plant. Flywheels or supercapacitors beside a hydro unit could answer the fastest frequency changes and spare the turbine some wear [1].

7.8 Limitations

All values are synthetic and describe a consistent generic unit, not a measured one. The turbine law omits the S-characteristic and the efficiency hill, so the overspeed numbers of case 5 and the challenge are teaching estimates, and a real runner's flow change with speed would move them. The gate characteristic, the turbine damping and the tank size are declared teaching values. The stiff grid removes the swing of the generator against the network, the governor has no deadband, and the town has no under-frequency load shedding, which Volume 17, Chapter 7 adds to an islanded campus grid. A clean 10 % load step, a town with a fixed frequency sensitivity and a grid held stiff within 0.1 s would not occur on a real system, but each isolates one mechanism. The wave speed of 1000 m/s needs a steel wall near 50 mm on a 5.46 m pipe by Korteweg's formula (30 mm gives 865 m/s), and case 8 doubles T_w without doubling T_e, which matters only with the elastic switch on. The tuning rule was quoted through lecture notes because its primary sources could not be read, and real governors are commissioned by test. The stability map is linear and the limits are annunciated, so neither replaces a transient study or a protection scheme, and a simulated pass is not a certification.

OPEN CHALLENGE **Flexible Duty From Machines Designed for Steady Load**

Most hydropower plants were designed decades ago for conditions unlike today's. Wind and solar generation lower the inertia of the grid, so hydro units are asked for fast frequency reserves and black starts, while hydrodynamic phenomena limit their operating range and frequent starts and stops add wear. Very little has been achieved in modeling and controlling those flow phenomena from standstill to full load. Progress would be a validated digital model of each unit that predicts the stress, wear and fatigue of a start, a stop or a fast change before the unit is asked for it [1].

7.9 Conclusion

This chapter studied a hydroelectric unit, a plant whose actuator first moves its output the wrong way, in a simulator that makes that visible with the IEEE working group's water column, a surge tank, a declared gate characteristic and a rotor on an island or a stiff grid. A 9.4 % gate opening at 80 MW cuts the power by 4.38 MW before the water catches up. The zero sits at 0.675 rad/s near the textbook $1/T_w$ rather than at the $2/T_w$ the design plan proposed, and the island's gain limit of 11.73 from the computed map agrees with the engine's onset to about 0.26 %.

The transient-droop rule settles a 10 % load step in 28 s with a nadir of 58.40 Hz, a governor tuned on the grid oscillates between 56.0 and 62.5 Hz on the island, and a surge tank that Thoma's estimate would call unstable below 128 s is stable under this governor down to about 85 s. Rejecting full load,

no one- or two-slope closure law keeps the head under 1.30 pu and the speed under 1.50 pu, while 656 three-slope laws do.

The model is lumped, its gate and runner are declared rather than measured, and its grid is stiff. A fuller model would add a runner characteristic with an S-shaped flow curve, a generator swing against a finite grid, a governor deadband and a surge tank throttle. The simulator shows that a hydro governor is tuned against its water column before it is tuned against anything else.

KEY POINTS

- Opening a hydro turbine's gate first lowers its power, because the water column must be accelerated by head taken from the turbine.
- On an island the water and mechanical starting times set a gain limit, 11.73 for a proportional governor at $T_w/T_m = 0.25$ and 80 MW.
- The transient-droop rule settles a 10 % load step in 28 s, while a governor tuned on the grid oscillates between 56.0 and 62.5 Hz on the island.
- After a load rejection the closure time trades the inlet pressure against the overspeed, and at 88 MW a two-slope law meets both limits where no single slope does.

Where else this applies. Chapter 6 meets an inverse response from swell and shrink, and Volume 17, Chapter 7 runs droop governors and a synchronizer on an island.

References

[1] I. Kougias, G. Aggidis, F. Avellan, S. Deniz, U. Lundin, A. Moro, S. Muntean, D. Novara, J. I. Pérez-Díaz, E. Quaranta, P. Schild, and N. Theodossiou, "Analysis of emerging technologies in the hydropower sector," *Renewable and Sustainable Energy Reviews*, vol. 113, p. 109257, 2019.

[2] Working Group Prime Mover and Energy Supply, "Hydraulic turbine and turbine control models for system dynamic studies," *IEEE Transactions on Power Systems*, vol. 7, no. 1, pp. 167–179, 1992.

[3] IEEE, "IEEE guide for the application of turbine governing systems for hydroelectric generating units," IEEE Std 1207-2011, 2011.

[4] P. Kundur, *Power System Stability and Control*, N. J. Balu and M. G. Lauby, Eds. New York: McGraw-Hill, 1994.

[5] J. Freudenberg and D. Looze, "Right half plane poles and zeros and design tradeoffs in feedback systems," *IEEE Transactions on Automatic Control*, vol. 30, no. 6, pp. 555–565, 1985.

[6] K. Iinoya and R. J. Altpeter, "Inverse response in process control," *Industrial & Engineering Chemistry*, vol. 54, no. 7, pp. 39–43, 1962.

[7] K. J. Åström and R. D. Bell, "Drum-boiler dynamics," *Automatica*, vol. 36, no. 3, pp. 363–378, 2000.

[8] S. Hagihara, H. Yokota, K. Goda, and K. Isobe, "Stability of a hydraulic turbine generating unit controlled by P.I.D. governor," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-98, no. 6, pp. 2294–2298, 1979.

[9] L. Wozniak and G. H. Fett, "Optimal gate closure schedule for hydroelectric turbine system," *Journal of Fluids Engineering*, vol. 95, no. 1, pp. 47–52, 1973.

[10] X. Fu, D. Li, H. Wang, J. Yang, and X. Wei, "Influence and multi-objective optimization on three-stage guide vane closure scheme of a pumped storage power plant," *Journal of Energy Storage*, vol. 72, p. 108545, 2023.

[11] M. H. Chaudhry, *Applied Hydraulic Transients*. New York: Springer, 2014.

[12] K. Vereide, B. Svingen, T. K. Nielsen, and L. Lia, "The effect of surge tank throttling on governor stability, power control, and hydraulic transients in hydropower plants," *IEEE Transactions on Energy Conversion*, vol. 32, no. 1, pp. 91–98, 2017.

[13] L. M. Hovey, "Optimum adjustment of hydro governors on Manitoba Hydro system," *Transactions of the American Institute of Electrical Engineers. Part III: Power Apparatus and Systems*, vol. 81, no. 3, pp. 581–586, 1962.

[14] J. Undrill and J. Woodward, "Nonlinear hydro governing model and improved calculation for determining temporary droop," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-86, no. 4, pp. 443–453, 1967.

[15] F. Schleif and A. Wilbor, "The coordination of hydraulic turbine governors for power system operation," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-85, no. 7, pp. 750–758, 1966.

[16] F. Schleif and R. Angell, "Governor tests by simulated isolation of hydraulic turbine units," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-87, no. 5, pp. 1263–1269, 1968.

[17] M. S. Ghidaoui, M. Zhao, D. A. McInnis, and D. H. Axworthy, "A review of water hammer theory and practice," *Applied Mechanics Reviews*, vol. 58, no. 1, pp. 49–76, 2005.

[18] U.S. Department of Energy, Water Power Technologies Office, "History of hydropower," Web page, 2026, accessed 21 November 2026. [Online]. Available: https://www.energy.gov/eere/water/history-hydropower

[19] U.S. Bureau of Reclamation, "Hydroelectric power," Power Resources Office, Bureau of Reclamation, Denver, CO, n.d., undated pamphlet; PDF of April 2008. [Online]. Available: https://www.usbr.gov/power/edu/pamphlet.pdf

[20] ——, "Mechanical maintenance of mechanical and digital governors for hydroelectric units," Power Resources Office, Bureau of Reclamation, Denver, CO, Facilities Instructions, Standards, and Techniques Volume 2-3, 2021, fIST 017, minor revision 4 January 2021. [Online]. Available: https://www.usbr.gov/power/data/fist/FIST_2-3_(1-2021).pdf

[21] ——, "Commissioning guide for hydroelectric facilities," Power Resources Office, Bureau of Reclamation, Denver, CO, Facilities Instructions, Standards, and Techniques Volume 4-7, 2023, fIST 024, 30 August 2023. [Online]. Available: https://www.usbr.gov/power/data/fist/FIST_4-7_(8-2023).pdf

[22] 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, Gaithersburg, MD, Special Publication 800-82r3, 2023.

7.10 Exercises

7.1 Concept check · 2 min

In case 1 the unit supplies the town alone and the gate is held by hand. The gate is raised from 65.6 to 75 %, and for about a second the turbine power falls. Why?

(a) The head at the runner falls at once, while the water in the penstock needs time to reach the larger flow

(b) The governor sees the frequency fall and closes the gate before the servo has finished opening it

(c) The rotor speeds up at once, and the turbine's speed damping $D_t G \Delta\omega$ takes the power away

(d) The surge tank level drops at once, so the penstock is fed from a lower head than before

7.2 Concept check · 2 min

The island's stability map gives a boundary gain of 11.73 for a proportional governor with 5 % droop at $T_w/T_m = 0.25$ and 80 MW. Which single change lowers that boundary?

(a) A heavier rotor, with T_m raised from 8 to 16 s and T_w unchanged

(b) A penstock twice as long, with T_w raised from 2 to 4 s and T_m unchanged

(c) A surge tank twice as large, with T_s raised from 250 to 500 s

(d) A town load of 88 MW at 60 Hz instead of 80 MW

7.3 Concept check · 2 min

Case 6 runs a small surge tank, $T_s = 60$ s, under the rule's governor with no permanent droop. After the load step the tank's swing grows about 8 % each half cycle. Switched to Manual, the same swing dies away. What makes the difference?

(a) The tank is smaller than Thoma's area, so its swing grows whatever the gate does

(b) The governor's integral action winds up on the gate's rate limit and overshoots each half cycle

(c) The governor holds the power, so it opens the gate as the head falls and draws more water from a tank that is already low

(d) The water hammer wave in the penstock reflects at the tank and pumps its level each half cycle

7.4 Theory and calculation · 6 min

The gate-to-power transfer function of the rigid column is Equation (7.7), with $T_w = 2$ s.

(a) For the ideal turbine, $f_p = q_{nl} = 0$ and $q = h = 1$, find the zero, the pole and the ratio of the first response to a gate step to the final one. *Checkpoint:* $+0.5\,\text{rad/s}$, $-1.0\,\text{rad/s}$ and -2.

(b) At 80 MW the column carries $q = 0.778$ at $h = 0.964$, with $f_p = 0.02$ and $q_{nl} = 0.08$. Find the zero, the pole and the same ratio. *Checkpoint:* 0.675 rad/s, $-1.25\,\text{rad/s}$ and -1.86.

(c) Find the zero at 80 MW when T_w is doubled to 4 s, and say what it means for the governor. *Checkpoint:* 0.337 rad/s.

7.5 Theory and calculation · 6 min

The plant's surge tank has $T_s = 250\,\text{s}$ and its tunnel $T_t = 10\,\text{s}$ and $f_t = 0.04$, on the bases $H_0 = 100\,\text{m}$ and $Q_0 = 115\,\text{m}^3/\text{s}$. At 80 MW the tank head is $h_s = 0.976$.

(a) Find the tank's cross-section and its undamped period at a fixed gate. *Checkpoint:* $287.5\,\text{m}^2$ and 314.2 s.

(b) Find Thoma's filling time at 80 MW, the matching cross-section, and the plant's margin over it. *Checkpoint:* 128.1 s, about $147\,\text{m}^2$, a factor of 1.95.

(c) Case 6 shrinks the tank to $T_s = 60\,\text{s}$. Find its undamped period, and say whether Thoma's area and the chapter's closed-loop boundary each predict a growing swing under the isochronous governor. *Checkpoint:* 153.9 s. Both predict growth.

7.6 Theory and calculation · 8 min

In case 1 the gate is fixed and the unit supplies the town alone, with $h_0 = 1$, $f_t = 0.04$, $f_p = 0.02$, $q_{nl} = 0.08$, $c = 1$, $A_t = 1.1893$, $D_t = D_L = 1$, $D = 0$ and the town at $P_L = 0.8\,\text{pu}$ at 60 Hz.

(a) From Equations (7.1), (7.2) and (7.4) at steady state, show that a fixed ideal gate G passes $q = 1/\sqrt{1/G^2 + f_t + f_p}$, and check the design point at 65.6 % of stroke. *Checkpoint:* $G = 0.792$, $q = 0.778$, $h = 0.964$ and $P_m = 0.800\,\text{pu}$ at rated speed.

(b) Find the same quantities after the gate is raised to 75 % and the water has settled. *Checkpoint:* $G = 0.857$, $q = 0.839$, $h = 0.958$ and $P_m = 0.864\,\text{pu}$ at rated speed.

(c) With no governor, find the speed deviation and the frequency at which the rotor balances the town, and the power the town then draws. *Checkpoint:* $\Delta\omega = 0.0389$, 62.33 Hz and 83.1 MW.

7.7 App challenge · 7 min

Load scenario 1, Open the gate by hand on the island, at the default seed and the default playback speed. Press Run, pause once the clock has passed 2 s, and note that time as t_m. Raise Gate opening from 65.6 to 75.0 %, run on to 130 s at 5×, and export the CSV file. Read the columns `mech_MW` and `freq_Hz`.

(a) Find the lowest turbine power after the move, its time after t_m, and the dip below the starting 79.99 MW. *Checkpoint:* 75.62 MW about 0.7 s after the move, a dip of 4.38 MW, within 0.05 MW and 0.1 s.

(b) Find how long after t_m the turbine power is back above 80 MW, and how long until the frequency passes 62 Hz. *Checkpoint:* 1.4 s and 11.0 s, each within 0.1 s.

(c) Read the peak frequency on the Frequency nadir · peak tile and the frequency at 130 s, and compare them with the steady state of the previous exercise. *Checkpoint:* 62.30 Hz and 62.00 Hz, each within 0.01 Hz.

7.8 App challenge · 8 min

Load scenario 2, A proportional governor and its gain limit, at the default seed, and keep the Stability map view. The governor is proportional, $K_c = 4$, and the town's load rises 10 % at 10 s.

(a) Run at 1×, pause at 5 s and read the boundary printed beside your marker on the map. Run on, pause again between 30 and 31 s, and read it again. *Checkpoint:* 11.72 and 13.04, each within 0.1.

(b) Set Proportional gain (PID) to 14 at that pause and run to 600 s at 20×. From the trend or the CSV column `freq_Hz`, read the peak-to-peak swing and the time between successive peaks after 300 s. *Checkpoint:* 4.6 to 5.5 Hz, and 10.0 s within 0.2 s.

(c) Press the reload button, run at 1× to a pause between 30 and 31 s, set the gain to 13, and run to 600 s at 20×. Read the peak-to-peak swing over 160 to 180 s and over 580 to 600 s. *Checkpoint:* 0.74 Hz within 0.05 Hz, and 4.7 Hz within 0.1 Hz.

7.9 App challenge · 8 min

Load scenario 5, Load rejection: closure time against pressure and overspeed, at the default seed. The unit carries 88 MW on the island, and the breaker opens at 10 s. For each run below, run to 40 s and read the Peak pressure · peak speed tile. The teaching limits are 1.30 pu of head and 1.50 pu of speed.

(a) Run as shipped, with a single 6 s closure. *Checkpoint:* 1.79 pu and 1.37 pu, each within 0.01 pu.

(b) Set Closure time to 15 s, press Reset and run. *Checkpoint:* 1.28 pu and 1.60 pu.

(c) Set Closure time back to 6 s and Closure law to Two slopes, leaving Cushion breakpoint at 40 % and Cushion closure time at 15 s. Press Reset and run. *Checkpoint:* 1.29 pu and 1.44 pu.

Say which limit each run breaks, and why the third law meets both.

7.10 Application area · 4 min

In case 5 the penstock carries 4.19 m/s when the breaker opens. The penstock is 400 m long with a wave speed of 1000 m/s, on a static head of 100 m.

(a) Find the head rise an instant stop of that flow would cause, in meters and in per unit, and the closure time above which the rigid column applies. *Checkpoint:* 427 m, 4.27 pu, and 0.8 s.

(b) From full load no setting of the servo's one- or two-slope law keeps both teaching limits. Name three remedies a plant has, and how the choice is confirmed on the real unit. *Checkpoint:* A shaped law with more slopes, a pressure relief valve, more rotor inertia or a shorter water column, confirmed by load-rejection tests.

7.11 Application area · 4 min

In case 4 the unit carries 80 MW on a stiff grid with a permanent droop of 5 % acting on the gate position. The grid frequency falls 0.2 Hz.

(a) Find the gate move the droop asks for, and estimate the power change from the gate-to-power slope of 0.74 at 80 MW. Compare both with the measured 6.65 % of stroke and 4.65 MW, and with the 6.67 MW a dispatcher would expect from the droop. *Checkpoint:* 6.67 % of stroke and about 4.9 MW.

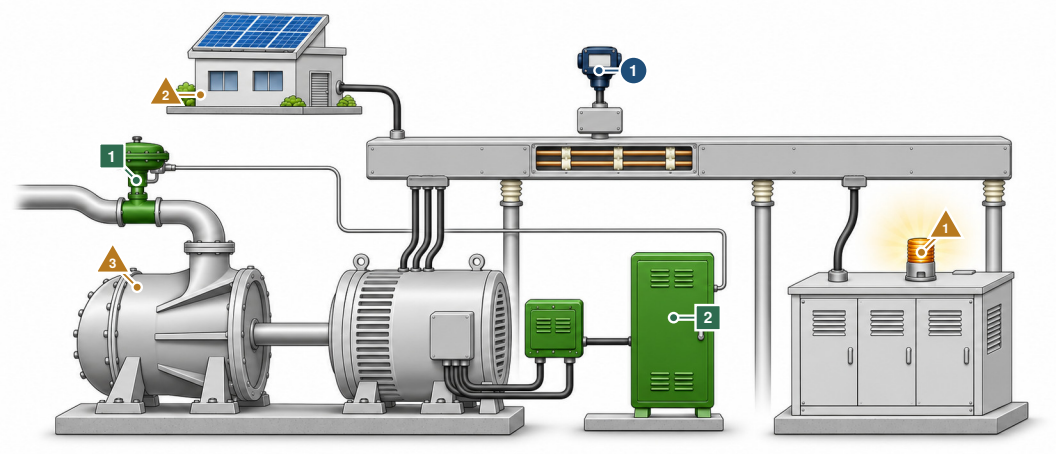
(b) Explain what the frequency-biased power controller changes for the dispatcher, and what it costs. *Checkpoint:* The power, not the gate, carries the share: 6.78 MW at 900 s, after a peak of 7.29 MW.

7.12 Review the agent · 5 min

The penstock of a unit like the chapter's is to be extended, which raises T_w from 2 to 4 s with $T_m = 8$ s. An AI assistant advises: "Keep the transient droop of 25 % with the 3 s reset. On the present penstock it settles a 10 % load step in 11.7 s, and its gain $1/R_t = 4$ is far below the stability map's boundary of 11.73 for this unit, so the margin covers the change. The rule's settings for 4 s would only make the governor slower." Find the flaw and say how you would check the advice before the unit runs on an island.

CHAPTER 8

Grid Frequency Simulator for Inertia, Droop Offset and Integral Restoration



The controller moves 1 the steam valve, opened by the governor as the frequency falls
2 the generation command sent by the control center

The controller reads 1 the frequency of the bus

What disturbs it 1 a step in the load 2 a change in solar output 3 part of the generation lost when a unit trips

Notice that only the spinning turbine and generator store energy between the steam and the load, so every mismatch changes their speed and with it the frequency read on the bus. Illustration, not to scale.

IN PLAIN TERMS **Turning Together, Slowing Together**

The main idea. All the large power plants on a grid spin in step, and the speed they share is the grid's frequency. When more power is used than made, the machines give up some of their stored spin and slow down together.

The challenge. Each plant's governor, a controller on its turbine, opens the steam valve as the speed falls. This stops the fall but leaves the frequency a little low. A slower controller in a control center keeps adding power until the error is gone.

Why it matters. With fewer spinning machines the first fall comes faster, but the resting error and its cure stay the same.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the fact on lost generators (Section 8.1), the history of area control (Section 8.2), the try-it segment of scenario 7 (Section 8.4.1), the try-it of scenario 10 and the fact on a strong integral (Section 8.6.3) 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 results (Section 8.6), Technology in Practice (Section 8.7), Limitations (Section 8.8) and the open challenge

The control question Grid frequency is the level of the kinetic energy stored in the spinning machines. After a load step, which parameters set how fast it falls, where droop arrests it and how integral action brings it back to 60 Hz?

Learning objectives

- Write the swing equation as an energy balance and explain why the frequency integrates a power imbalance.
- Compute the initial rate of change of frequency, the settled droop offset and the borrowed kinetic energy for a load step or a unit trip.
- Test the stability of the primary loop with the Routh condition and find the inertia below which it fails.
- Estimate the time constant of integral restoration and choose an integral gain that meets a restoration target inside a ramp limit.
- Explain why inertia, droop and integral gain each move one of the three numbers of a frequency event and leave the other two alone.

Prerequisites Sections 1.4.2 (balances), 1.4.3 (poles), 1.4.4 (integrating response) and 1.6.2 (velocity form), and Chapter 2 for the offset of a proportional loop.

The app https://apmonitor.com/apps/grid-frequency/

See also Volume 2, Chapter 5 for tuning a loop on an integrating inventory, Volume 2, Chapter 3 for a standing error from proportional control set beside one from an actuator bound. Volume 17, Chapter 7 for droop sharing, virtual inertia and a frequency-restoring integral loop on an islanded bus.

8.1 Introduction

An alternating-current power system stores almost no energy in electrical form. The energy that it does store is held in the rotors of the synchronized machines, and the size of that store is fixed by the speed at which the rotors turn. Generation and load are matched continuously by taking the difference out of those rotors or adding it to them, which slows the rotors or speeds them up, so the frequency of the interconnection is a direct measurement of an energy inventory. A tank whose outflow rises with head reaches a steady state without any control action, because the outflow is a function of the level. A rotating mass has no equivalent mechanism. A low rotor speed does not by itself raise generation or lower load beyond a small damping effect, so the inventory behaves as an integrator and the frequency ramps until a control action arrests it. It is the integrating tank of Section 1.4.4 with a small leak.

The first control action comes from the turbine governor, which is a proportional controller. Each governor measures shaft speed and opens its valve in proportion to the speed error, with a gain fixed by a setting called droop. The governor arrests a frequency excursion within a few seconds and leaves a steady-state offset, for the same reason that a proportional-only level controller leaves an offset, as in Chapter 2. A proportional controller requires a standing error to hold a standing output, and a load change that persists requires a valve position that also persists [1]. The offset is a structural property of the primary loop rather than a tuning defect, and no droop setting removes it.

The arrangement differs from most examples in a first control course in where the proportional controller is located. Droop is built into governor hardware and acts whether or not any computer is running, so the gain of the primary loop is a plant parameter rather than a value in a control system. The term that remains available to a control engineer is the integral term, supplied by automatic generation control (AGC) at a much slower cadence. The result is a two-tier structure in which the fast tier is not retuned by the operator and the slow tier has little effect during the fast event. Inertia sets the rate of change, the composite of droop and load damping sets the arrested offset, and integral action sets the final value, and no one of the three quantities depends on the parameters that set the other two.

The experiments of this chapter cover integrating process dynamics, proportional offset and its dependence on loop gain, the requirement for integral action to remove offset, disturbance rejection

without setpoint motion, actuator saturation against reserve and ramp limits, anti-windup, and the separation of a fast primary loop from a slow secondary loop in the manner of a cascade [1, 2]. The physical fundamentals are conservation of energy applied to the rotating store, in which the swing equation is a per-unit inventory balance, the frequency dependence of aggregate load as a driving force opposing the excursion, and a time-scale separation between a governor lag of 0.20 s and a secondary time constant of 3.5 s [3]. The behavior matters industrially because the rate of change of frequency and the frequency nadir set protection margins and reserve procurement on every interconnected system, and because the displacement of synchronous plant by converter-interfaced generation reduces the one parameter that sets both [4].

> ✦ DID YOU KNOW? **A Lost Generator Almost Every Day**
>
> A 2010 national laboratory report counts how often the sudden loss of a generator is recorded: almost daily in the very large Eastern Interconnection, about once a week in the smaller Texas Interconnection, and the very largest events rarely more than once or twice a year [5].

This chapter works with a single-area load frequency control simulator and states the model, the controller and the measured behavior in enough detail to reproduce them from the text. It separates the three time scales that a frequency event passes through: the rate of change of frequency, the arrested offset and the restored setpoint follow from three expressions with three independent parameter groups, and measured runs show that a change to one group moves one quantity. Section 8.2 describes how frequency regulation acquired its present two-tier form. Section 8.3 writes the three-state model of governor, turbine and rotating store, with its reserve and ramp limits and its protection flags. Section 8.4 derives the rate of change, the offset, the stability limit of the primary loop and the time constant of restoration. Section 8.5 states the secondary controller as the code executes it, and Section 8.6 reports the measured runs of the eleven scenarios. Section 8.7 sets the model beside the sensors, valves and control platforms of real grids. Sections 8.8 and 8.9 state the limits of the model and what the displacement of synchronous plant changes in it, and Section 8.10 sums up. Offset and integral action carry a physical meaning throughout, and a measured value in hertz.

8.2 Background

Early generating stations ran alone or in small groups, and a governor that returned speed exactly to its reference was both possible and desirable. The arrangement fails as soon as two machines are paralleled. Two isochronous governors on one bus each attempt to hold the same speed, the machine with the marginally higher reference takes as much load as its capacity allows, and the division of output between them is indeterminate. Droop was introduced to make that division determinate. A falling speed-output characteristic gives each machine exactly one output at any given frequency, so load divides between machines in inverse proportion to their droop settings, and the consequence is that the common frequency depends on total load.

Interconnection added a second requirement. Once control areas are tied together, a frequency error is shared across the whole interconnection while the schedule violation belongs to one area, and a regulator that responds to frequency alone leaves every area responding to every other area’s imbalance. Tie-line bias control resolves this by forming an area control error that combines the deviation of net tie-line flow from schedule with the frequency deviation weighted by a bias factor, so that an area responds fully to its own imbalance and, in the ideal case, not at all to a neighbor’s [6]. The resulting structure, proportional governing followed by integral restoration of an area control error, remains in use across successive generations of governor, telemetry and control-center hardware.

Analytical treatment of the resulting multi-area problem followed the availability of state-space methods, with optimal control formulations of the megawatt-frequency problem appearing around 1970 and establishing the low-order linear model that is still the standard vehicle for load frequency control studies [7]. Later work condensed the multi-machine response into a single equivalent inertia

and a single reheat time constant to predict frequency nadir directly [8], and successive surveys have catalogued the controller structures applied to the same model [9, 10]. The model has also been carried outward from the transmission system, because droop is now the standard primary control law for parallel inverters in islanded microgrids [11], the setting of Volume 17, Chapter 7, where a droop-governed genset and a battery inverter share an islanded load and an integral loop restores frequency.

Current work in the field concerns the inertia coefficient itself. Converter-interfaced wind and solar generation displaces synchronous plant without contributing rotational energy to the system, so the inertia constant falls while the disturbances it must absorb do not [4, 12]. System operators in small or weakly interconnected systems reached the operational limits of this first and now constrain dispatch to keep inertia above a floor [13], and the stability of a grid dominated by converters is an active research program [14]. That literature concerns the coefficient $2H$ in the model of this chapter, and the analysis of Section 8.4 states which measured quantities that coefficient does and does not set.

> ⧖ HISTORY **Judging the Second Tier, 1962 to 1973**
>
> A committee of North American operators formed in 1962 to recommend an informal operating structure for joining the Eastern and Western Interconnections, which were joined in 1967 and separated a short time later over transmission flow problems. In the 1960s most energy management systems were analog, and a dispatcher relied on the telephone and on strip charts of frequency, tie-line flows and area control error. After the 1965 blackout the industry began to apply digital computers. The first control performance criteria were tested in 1971 and put in place in 1973, and their 10-minute interval was chosen because it was the division most often marked on the strip charts. The attention given to this secondary tier cost the industry much of its memory of why governor response matters [15].

8.3 Process Model

The simulated plant is a single control area reduced to one equivalent machine, one equivalent governor and one equivalent turbine, supplying one aggregate load and receiving one converter-interfaced renewable infeed. That arrangement is drawn in Figure 8.1, which carries every symbol the equations below use. The reduction to a single equivalent machine follows the low-order system frequency response model that is the standard vehicle for load frequency control studies [8, 16].

8.3.1 Per-Unit Conventions

Every power in the model is expressed per unit of a system base S_B equal to the aggregate rating of the synchronous fleet, and every power symbol carrying a Δ is a deviation from the scheduled dispatch. Frequency is normalized differently. The state that the model integrates is the per-unit frequency deviation

$$\Delta f = \frac{f - f_0}{f_0}, \tag{8.1}$$

where f is the system frequency in Hz and $f_0 = 60\,\text{Hz}$ is the nominal value, so Δf is dimensionless and a full hertz of error is 1.667×10^{-2} pu. The inertia constant H is the stored kinetic energy of all synchronized machines at rated speed divided by S_B, which gives it units of seconds and the interpretation of how long the fleet could supply rated output from its rotors alone. The load damping constant D carries units of per-unit power for each per unit of frequency, so $D = 1$ means a 1 % fall in frequency reduces load by 1 % of base. The droop R carries the reciprocal units, per-unit frequency for each per unit of power, so $R = 0.05$ means a 5 % frequency change drives a governor from no load to full load and the primary gain is $1/R = 20$ pu/pu. Rates of change of frequency are reported in Hz/s rather than in per unit, because relay settings are written in those units.

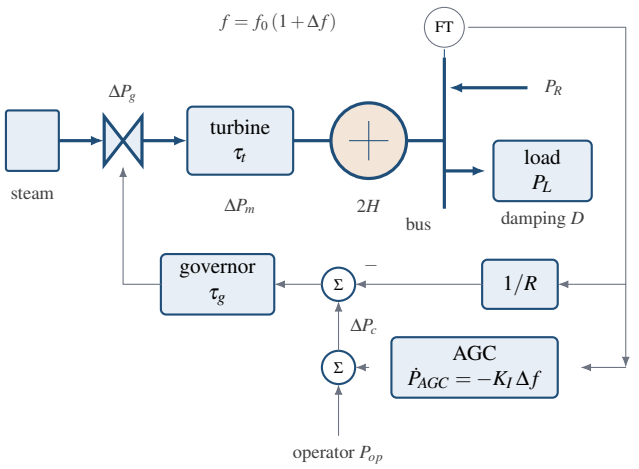


Figure 8.1 Single-area load frequency control arrangement modeled by the application. The governor valve sets ΔP_g, the turbine converts it to mechanical power ΔP_m through the lag τ_t, and any mismatch with the bus load P_L less the renewable infeed P_R is taken from the rotating mass characterized by $2H$ and D. Frequency measured at the bus returns through the droop gain $1/R$ as primary control and through the integrator K_I as secondary control.

8.3.2 Governing Equations

The energy inventory is written as the swing equation with a damping term, as shown in (8.2), the balance of Equation (1.1) applied to the rotating store.

$$2H \frac{\mathrm{d}\Delta f}{\mathrm{d}t} = (\Delta P_m - P_{loss}) - \Delta P_L - D\Delta f \tag{8.2}$$

where ΔP_m is the deviation of turbine mechanical power in pu, P_{loss} is mechanical power removed by a tripped unit in pu, ΔP_L is the deviation of net load in pu and $D\Delta f$ is the self-regulation of the load. Both sides are in per unit of power because $2H$ carries units of seconds and the derivative carries units of reciprocal seconds. Equation (8.2) is conservation of energy applied to the rotating store, with the stored kinetic energy linearized about rated speed, so it reads as an inventory balance in which the level is Δf itself [8]. The only self-correcting term is $D\Delta f$, and it is a property of the connected load rather than of any controller.

The steam path cannot fill instantly, so mechanical power follows valve position through a first-order lag, as shown in (8.3).

$$\tau_t \frac{\mathrm{d}\Delta P_m}{\mathrm{d}t} = \Delta P_g - \Delta P_m \tag{8.3}$$

where ΔP_g is the governor valve output in pu and τ_t is the turbine time constant in seconds.

The governor is the primary control loop and it is written as a lag driven by the difference between the commanded output and the droop feedback, as shown in (8.4).

$$\tau_g \frac{\mathrm{d}\Delta P_g}{\mathrm{d}t} = \Delta P_c - \Delta P_g - \frac{1}{R}\Delta f \tag{8.4}$$

where ΔP_c is the aggregate generation command in pu, τ_g is the governor time constant in seconds and $1/R$ is the primary gain in pu/pu. The term $-(1/R)\Delta f$ is the whole of primary control. It is a pure proportional action on the frequency error, it belongs to the plant model rather than to the controller, and it operates identically whether or not the secondary loop is in service.

Table 8.1 Variable Classification

Symbol	Role	Unit	Range
f	CV, measured	Hz	59.0–61.0 typical
f_{SP}	setpoint	Hz	59.8–60.2
Δf	state	pu	$\pm 1.7 \times 10^{-2}$
ΔP_m	state	pu	± 0.65 reachable
ΔP_g	state	pu	± 0.65 reachable
ΔP_c	MV	pu	-0.30–0.30
P_{op}	MV, manual	% of base	-30–30
P_{AGC}	MV, secondary	pu	signed, unclamped
d_L	DV	%	-10–15
d_R	DV	%	-15–15
d_G	DV	%	0–50
σ_R	DV	%	0–5
RoCoF	derived	Hz/s	± 2 typical
E_{kin}	derived	pu s	signed

Frequency and its rate of change are reported from the state through (8.5).

$$f = f_0\,(1 + \Delta f)\,, \qquad \mathrm{RoCoF} = f_0\,\frac{\mathrm{d}\Delta f}{\mathrm{d}t} \tag{8.5}$$

Net load collects the three disturbance channels and the renewable variability into one signal, as shown in (8.6).

$$\Delta P_L = \frac{d_L}{100} - \frac{d_R}{100} - n(t), \qquad P_{loss} = \frac{d_G}{100} \tag{8.6}$$

where d_L is the commanded load change in percent of base, d_R is the commanded renewable output change in percent of base, d_G is the tripped generation in percent of base and $n(t)$ is band-limited renewable variability in pu. A renewable drop enters net load with a positive sign, because net load is load less infeed, so the swing equation does not distinguish a loss of renewable output from an equal rise in load. A unit trip enters separately as P_{loss} so that mechanical power and renewable infeed remain separately reported quantities.

Renewable variability is an Ornstein-Uhlenbeck sequence generated by the framework's colored-noise primitive, updated once per controller sample with the recursion in (8.7).

$$n_k = a\,n_{k-1} + \sqrt{1 - a^2}\;\sigma_R\,\xi_k, \qquad a = \mathrm{e}^{-\Delta t_c/\tau_R} \tag{8.7}$$

where ξ_k is a unit normal deviate from a seeded mulberry32 generator, σ_R is the requested standard deviation in pu, entered as a percentage of base, and τ_R is the correlation time. The innovation is scaled so that the stationary standard deviation equals σ_R exactly. The value is held constant across the integration step, because the derivative function is evaluated four times per step by the integrator and must be a pure function of its arguments.

8.3.3 Variables and Parameters

Table 8.1 sorts the exposed signals into controlled, manipulated, disturbance, state and derived quantities, and Table 8.2 lists the parameters with their defaults, their adjustable ranges and, where a parameter is fixed by construction, that status.

The eleven parameters above the rule are adjustable while the simulation runs. The six below the rule are fixed in the application. Changing the first five would put the reported axes and the declared actuator limits out of step with the model, and T_t belongs to the secondary controller of Section 8.5 rather than to the plant. The defaults are the customary low-order load frequency control study set [8], and the composite grid stiffness they produce is $1/R + D = 21\,\mathrm{pu/pu}$.

Table 8.2 Model Parameters, Defaults and Adjustable Ranges

Symbol	Meaning	Default	Unit	Range
H	inertia constant	5.0	s	1.0–10.0
D	load damping	1.0	pu/pu	0–3
R	governor droop	0.05	–	0.02–0.20
τ_g	governor time constant	0.20	s	0.05–1.00
τ_t	turbine time constant	0.50	s	0.10–2.00
$\dot{P}_{max}$	generation ramp limit	0.10	pu/s	0.02–0.50
P_{L0}	scheduled load	1.00	pu	0.70–1.20
P_{R0}	renewable infeed	0.20	pu	0–0.60
Δf_n	normal band ($\pm$)	0.10	Hz	0.02–0.30
f_{UFLS}	load-shed relay	59.5	Hz	58.0–59.8
f_{OF}	over-frequency trip	60.5	Hz	60.2–62.0
f_0	nominal frequency	60.0	Hz	fixed
$\Delta P_{c,max}$	reserve up	0.30	pu	fixed
$\Delta P_{c,min}$	reserve down	−0.30	pu	fixed
P_{rated}	turbine rating	1.30	pu	fixed
τ_R	renewable noise correlation	2.0	s	fixed
T_t	anti-windup tracking time	1.0	s	fixed

8.3.4 Initialization and Numerics

Every scenario begins from an exact equilibrium rather than from zero, by solving (8.2) to (8.4) with all derivatives set to zero for whatever command and disturbance the scenario has pinned, as shown in (8.8).

$$\Delta f_0 = \frac{\Delta P_c - \Delta P_{L,\mathrm{tot}}}{1/R + D}, \qquad \Delta P_{m,0} = \Delta P_{g,0} = \Delta P_c - \frac{\Delta f_0}{R} \tag{8.8}$$

where $\Delta P_{L,\mathrm{tot}} = \Delta P_L + P_{loss}$ evaluated with the noise term set to zero. With the default scenario settings this gives $\Delta f_0 = 0$ and a reported frequency of exactly 60.000 Hz, so an undisturbed baseline precedes the event at $t = 2$ s.

The engine of Section 1.6.1 advances the three states with the fixed-step Runge-Kutta method of fourth order and a step of $\Delta t = 0.02$ s, which is a factor of ten below the fastest mode in the model at $\tau_g = 0.20$ s. The secondary controller executes on its own clock at $\Delta t_c = 0.10$ s, first at $t = 0$ and then at every multiple of Δt_c. Data are logged every third integration step, at intervals of 0.06 s. The protection logic of the next paragraph runs after the integration and before the outputs, within the step order of Figure 1.10. The random sequence is seeded at 20260824, so a scenario with variability enabled replays identically.

Two conditions are latched after each step and annunciated. The frequency falling below f_{UFLS} raises an under-frequency load-shedding flag and marks the record, and rising above f_{OF} raises an over-frequency trip flag, each clearing with 0.1 Hz of hysteresis. Neither actuates. A real under-frequency scheme would disconnect a block of feeders and change the model discontinuously, which would replace the transient under study with a step, so the relay is deliberately a flag rather than an event.

8.3.5 Assumptions

The model is a single-area, single-machine equivalent, so inter-machine swings, inter-area oscillations and any spatial variation of frequency are outside it. Real events show frequency differing by tenths of a hertz across an interconnection for the first several seconds, and the model does not reproduce that.

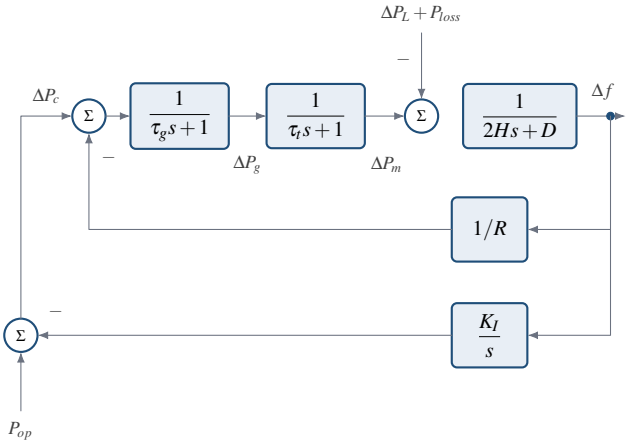


Figure 8.2 Primary and secondary loops around the frequency inventory. The governor droop path $1/R$ closes a proportional loop through two lags inside the plant, while the secondary path K_I/s adds the only integral action in the system and enters at the same summing point as the operator's command P_{op}.

There is no network. Voltages, angles, line impedances and losses are absent, so power injected at the bus arrives at the load unchanged and the model represents neither transmission limits nor voltage stability.

Everything is linearized about the scheduled dispatch. The swing equation uses $2H$ rather than the exact $2H(1+\Delta f)$, and the damping is a single constant. The linearization is accurate for a few percent of imbalance and progressively inaccurate beyond about 10 %, although the disturbance range extends to a 50 % generation loss.

A tripped unit removes mechanical power but not inertia. On a real system the surviving fleet loses that machine's stored energy as well, so a real trip produces a worse rate of change of frequency and a deeper nadir than this model predicts, and the discrepancy grows on a system already short of inertia.

The turbine is a single 0.5 s lag. A reheat steam unit adds a second stage of several seconds, a hydro unit has transient droop and an inverse response from water inertia, treated in Chapter 7, and a gas turbine behaves differently again. Any of these would slow primary response substantially, which is why real primary reserve is specified over tens of seconds rather than the five seconds of this model.

Load damping is one constant. Real frequency-dependent load is a mixture of induction motors, power electronics and thermostatically controlled devices whose aggregate behavior varies with time of day and season.

The generation command is a single aggregate quantity with one reserve limit and one ramp rate. Real reserve is a stack of individual units with individual ramp rates, of which only part is under automatic control at any moment.

Frequency measurement is perfect. There is no transducer noise, no filter lag, no dead band and no telemetry delay, although all four are present in real AGC and are the reason real secondary control runs on a multi-second cycle. The only stochastic element in the application is the renewable disturbance itself.

8.4 Dynamic Analysis

The three quantities that characterize a frequency event, the initial rate of change, the arrested offset and the final value, are set by three disjoint groups of parameters. The block diagram that produces them is shown in Figure 8.2.

8.4.1 Rate of Change of Frequency and Inertia

At the instant a disturbance is applied, ΔP_m has not moved because it is the output of two lags, and Δf is still zero because it is a state. Substituting both into (8.2) and (8.5) gives (8.9).

$$\mathrm{RoCoF}|_{t=0^+} = -\frac{f_0\,(\Delta P_L + P_{loss})}{2H} \tag{8.9}$$

Neither R nor D nor K_I appears. A 5 % load step on the default system therefore produces exactly $-60 \times 0.05/10 = -0.300\,\mathrm{Hz/s}$, and the application measures $-0.299\,\mathrm{Hz/s}$. The residual is the first 0.02 s integration step, because the peak is evaluated after every step and the first evaluation follows the step that applies the load, by which time the depressed frequency has already shed a little load through D. Cutting the inertia to $H = 2\,\mathrm{s}$ multiplies the measured peak by 2.49 to $-0.746\,\mathrm{Hz/s}$ against a predicted factor of 2.5, and doubling it to $H = 10\,\mathrm{s}$ gives $-0.150\,\mathrm{Hz/s}$ exactly.

> **TRY IT IN THE APP** **7 · Low Inertia ($H = 2\,\mathrm{s}$)**
>
> Load the case and press Run. The 5 % load step lands at 2 s. Watch the RoCoF strip. Its peak reaches −0.746 Hz/s, against −0.299 Hz/s in scenario 2, and the nadir falls to 59.658 Hz. Then watch the dial: it still settles at 59.857 Hz, the value of scenario 2.

8.4.2 Steady-State Offset and Composite Stiffness

Setting all three derivatives to zero collapses the model to a single algebraic relation, as shown in (8.10).

$$\Delta f_{ss} = \frac{\Delta P_c - \Delta P_L - P_{loss}}{1/R + D} \xrightarrow{\Delta P_c = 0} -\frac{\Delta P_L + P_{loss}}{1/R + D} \tag{8.10}$$

where the denominator $1/R + D$ is the composite stiffness of the area in pu of power per pu of frequency. Two contributions make up that stiffness and only one of them is a control action. The governors supply $1/R = 20$ and the load supplies $D = 1$ by drawing less power at a depressed frequency. The inertia constant H has dropped out, because a store that is neither filling nor emptying does not enter a steady-state balance whatever its size.

The stiffness is finite, so a finite imbalance forces a finite error. A 5 % load step gives $-0.05/21 = -2.381 \times 10^{-3}$ pu, which is $-0.143\,\mathrm{Hz}$ and a reported frequency of 59.857 Hz. The application settles at 59.857 Hz and its own verification test asserts agreement with (8.10) to 10^{-6} pu. Raising droop from 5 % to 12 % lowers the primary gain to 8.33 and more than doubles the measured offset to 0.321 Hz, while lowering droop to 2 % raises the gain to 50 and reduces the offset to 0.059 Hz. Raising the gain reduces the offset and does not remove it, which is a measured demonstration of the requirement for integral action [1].

> **Pause.** Try Exercise 8.2 before reading on.

8.4.3 Stability of the Primary Loop

Eliminating ΔP_m and ΔP_g from (8.2) to (8.4) with $\Delta P_c = 0$ gives the third-order characteristic polynomial in (8.11), whose roots are the poles of the primary loop in the sense of Section 1.4.3.

$$2H\tau_g\tau_t\,s^3 + [2H(\tau_g + \tau_t) + D\tau_g\tau_t]\,s^2 + [2H + D(\tau_g + \tau_t)]\,s + \left(D + \tfrac{1}{R}\right) = 0 \tag{8.11}$$

The Routh criterion on a cubic requires the product of the two middle coefficients to exceed the product of the outer two, which for the default τ_g and τ_t reduces to (8.12).

$$(1.4H + 0.1D)\,(2H + 0.7D) > 0.2H\left(D + \tfrac{1}{R}\right) \tag{8.12}$$

At the defaults the left side is 75.97 against 21 on the right, which is a wide margin. Holding $D = 1$ and $1/R = 20$ and reducing inertia, the inequality fails below $H = 1.05\,\text{s}$, and the primary loop is then unstable rather than merely oscillatory. The mechanism is the same as high proportional gain on any plant with two lags in series [3], and it is the reason the strong-droop scenario at $R = 2\,\%$ is oscillatory while settling at the smallest offset of the three droop cases.

8.4.4 Restoration by Integral Action

Secondary control acts far more slowly than the governors, so the primary loop can be treated as quasi-static and (8.10) used as its gain. Combining that with the integral law $\dot{P}_{AGC} = -K_I \Delta f$ gives the first-order closed-loop response in (8.13), whose time constant is (8.14).

$$\frac{\mathrm{d}\Delta P_c}{\mathrm{d}t} = -\frac{K_I}{1/R + D}\left(\Delta P_c - \Delta P_L - P_{loss}\right) \tag{8.13}$$

$$\tau_{sec} \approx \frac{1/R + D}{K_I} \tag{8.14}$$

At the default $K_I = 6$ pu/s per pu the secondary time constant is $21/6 = 3.5\,\text{s}$, which is seven times the turbine lag of 0.5 s and 17.5 times the governor lag of 0.20 s. The separation is deliberate and is the same time-scale argument that governs the choice of a cascade inner loop [1], as in the level-to-flow cascade of Volume 3, Chapter 2. Raising K_I above approximately 25 places the secondary loop on the time scale of the lags below it and the two loops then interact. Integral gain barely moves the nadir, because whatever command the integrator has accumulated by the arrest still has to pass through the governor and turbine lags. On the challenge case the nadir is 59.781 Hz at $K_I = 1$ and 59.797 Hz at $K_I = 20$, while at $K_I = 25$ the second swing falls to 59.655 Hz, below the first, so a large enough K_I makes the nadir worse.

The steady state of the complete system follows directly and is what the manual scenario is constructed to demonstrate. Setting $\dot{P}_{AGC} = 0$ requires $\Delta f = 0$, and (8.10) with $\Delta f_{ss} = 0$ then requires $\Delta P_c = \Delta P_L + P_{loss}$. The generation command must equal the imbalance exactly, because at 60.000 Hz the droop term contributes nothing.

🕮 **Pause.** Try Exercise 8.3 before reading on.

8.5 Controller Design

The primary controller is (8.4), and it is a plant parameter rather than an adjustable controller setting. The implemented code is the secondary layer alone. It carries integral action with an optional proportional term and no derivative term, so it is coded separately from the shared PID block of Section 1.6.2, which parameterizes integral action as K_c/τ_I and cannot express a pure integral law at $K_c = 0$. The code adds a secondary contribution to the operator's own setting, which the secondary loop never writes. It otherwise follows the shared block, with the limits on the applied command and anti-windup on the integral term alone. Away from the limits the law it executes is the incremental, or velocity, form of the PID law [17], in which each sample adds a change to the previous command.

🕮 **Pause.** Try Exercise 8.1 before reading on.

8.5.1 Discrete Algorithm

At each controller sample of period $\Delta t_c = 0.10\,\text{s}$ the following is executed in order, with k indexing the sample. The error is formed in per-unit frequency, as shown in (8.15).

$$e_k = \frac{f_{SP} - f_0}{f_0} - \Delta f_k \tag{8.15}$$

A positive e_k means the frequency is below setpoint and generation must rise, so the command falls as the measured frequency rises and the loop is reverse acting in the ISA sense. The two per-sample

increments are given by (8.16).

$$\begin{aligned} \Delta u_k^P &= K_p\,(e_k - e_{k-1}) \\ \Delta u_k^I &= K_I\,e_k\,\Delta t_c \end{aligned} \tag{8.16}$$

with K_p in pu/pu and K_I in pu per second per pu. There is no derivative increment, and none is appropriate, because the measurement enters an inventory whose own dynamics already differentiate the imbalance.

The operator contribution is clipped to the reserve limits, the secondary contribution is added, and the demanded command passes through the reserve clamp and then the ramp limit, as shown in (8.17).

$$\begin{aligned} P_{op,k} &= \mathrm{clamp}\big(\Delta P_{c,k}^{man},\, \Delta P_{c,min},\, \Delta P_{c,max}\big) \\ S_k &= K_p\,e_k + I_{k-1} \\ c_k &= P_{op,k} + S_k \\ v_k &= \mathrm{clamp}(c_k,\, \Delta P_{c,min},\, \Delta P_{c,max}) \\ \delta_k &= \mathrm{clamp}\big(v_k - \Delta P_{c,k-1},\, -\dot{P}_{max}\Delta t_c,\, \dot{P}_{max}\Delta t_c\big) \\ \Delta P_{c,k} &= \Delta P_{c,k-1} + \delta_k \\ I_k &= I_{k-1} + \Delta u_k^I + \frac{\Delta t_c}{T_t}\left(\Delta P_{c,k} - c_k\right) \end{aligned} \tag{8.17}$$

where $\Delta P_{c,k}^{man}$ is the operator input, S_k is the secondary contribution, c_k is the command the algorithm demands, δ_k is the change in command that the ramp rate permits within one controller sample and $T_t = 1.0\,\mathrm{s}$ is the tracking time of the back-calculation. In the symbols of Equations (1.24) to (1.27), $P_{op,k}$ takes the place of the bias u_b, c_k is v_k, $\Delta P_{c,k}$ is u_k, K_I replaces K_c/τ_I, and the reserve clamp followed by the ramp limit is the saturation. The S_k of this chapter is the whole secondary contribution, the running sum of Chapter 1 plus I_{k-1}. The integral used at sample k was accumulated through sample $k-1$. For an unchanged operator input, subtracting the demand at sample $k-1$ from the demand at sample k gives the recursion actually executed, (8.18).

$$c_k = c_{k-1} + \Delta u_k^P + \Delta u_{k-1}^I + \frac{\Delta t_c}{T_t}\left(\Delta P_{c,k-1} - c_{k-1}\right) \tag{8.18}$$

Each demand is the previous demand plus one proportional increment, one integral increment computed from the previous error and a back-calculation correction. Away from the limits $\Delta P_{c,k} = c_k$, the last term vanishes, and the recursion is algebraically identical to the positional law $\Delta P_c = P_{op} + K_p e + K_I \sum e\,\Delta t_c$. At a limit the two differ, as described below.

When the secondary loop is out of service, whether because the application is in manual or because the secondary mode is droop only, the secondary contribution is zero and, as in the manual tracking of Section 1.6.2, the integral state is solved for the value that, with the proportional term of the next sample, reproduces the applied command, as shown in (8.19).

$$S_k = 0, \qquad I_k = \Delta P_{c,k} - P_{op,k} - K_p e_k \tag{8.19}$$

At the first sample after the secondary loop is switched on, the third line of (8.17) then gives $c_{k+1} = \Delta P_{c,k} + K_p(e_{k+1} - e_k)$ for an unchanged operator input, which is the previously applied command plus one proportional increment, and exactly that command at the default $K_p = 0$. The first integral increment reaches the command one sample later. Switching the secondary loop off is bumpless as well. At the first sample without it, the operator input is set once to the command last applied, so the operator takes over the whole command, as units keep their last basepoints when automatic generation control is suspended.

8.5.2 Saturation and Windup

Anti-windup is the back-calculation term alone. While the reserve limit or the ramp limit refuses part of a demand, each sample removes $\Delta t_c / T_t = 0.1$ of the refused part from the integral, so the demand is pulled back toward the command actually applied [17]. With $K_I = 10$ on the unit trip of scenario 3 the command rides the ramp limit for 1.6 s, from 2.72 to 4.32 s, and the frequency overshoots to 60.173 Hz. A counterfactual run with T_t made effectively infinite rides the limit for 1.8 s and overshoots to 60.207 Hz, and at $K_I = 15$ the two peaks are 60.245 and 60.317 Hz. The operator component stays at zero throughout, and the AGC component carries the whole 0.200 pu at the end of the run. A second unit trip on top of the scripted one makes a 40 % loss, which pins the command on its 0.30 pu reserve limit with the frequency held at 59.714 Hz. By $t = 20$ s the integral has settled at 0.329 pu, where the tracking term balances $K_I e$, against 0.729 pu and still rising with T_t effectively infinite. Removing the loss at $t = 20$ s drives the frequency to 61.467 Hz while the command ramps back to zero, against 61.479 Hz without the tracking term, so that peak is set by the ramp limit. The windup shows in the recovery instead: the frequency is back within 0.05 Hz at $t = 29.5$ s, against 33.7 s without the tracking term.

8.5.3 Tuning and Defaults

The integral gain K_I is adjustable over 0 to 40 pu/s per pu in steps of 0.5 with a default of 6. The optional proportional gain K_p is adjustable over 0 to 20 pu/pu with a default of 0. A nonzero K_p acts as additional droop routed through the command path, so it stiffens the arrested frequency in the same way a smaller R would and it still leaves an offset. The default secondary mode is droop only, so the offset is observed before any mechanism that removes it is introduced.

The operator command spans ± 30 % of base in steps of 0.1 %, and the frequency setpoint is adjustable over 59.8 to 60.2 Hz in steps of 0.005 Hz, which corresponds to the time-error correction bias that real interconnections apply [6].

8.5.4 Implementation

The app has the controls and the data export of Section 1.6.4. Manual and automatic modes are provided, with the aggregate generation command as the manipulated variable in manual and the integral law of (8.17) in automatic. The logged record covers the three states, both components of the command, the demanded command, the saturation and rate-limit flags, the achieved command rate, the borrowed kinetic energy, the running nadir and peak rate of change, both protection flags, the plant, dispatch and limit parameters and every disturbance channel, at 0.06 s intervals, which is what a reproduction of any table in this chapter requires.

8.6 Results

Eleven scenarios are supplied with the application and are listed in Table 8.3. Every one starts from an exact equilibrium at 60.000 Hz and applies its disturbance as a timed event at $t = 2$ s, apart from the variability case whose persistent component lands at $t = 10$ s and the handover case which switches secondary control on at $t = 10$ s.

8.6.1 Primary Response

Measured performance for the seven scenarios that test primary control alone is given in Table 8.4, each value recorded at the end of a 30 s run. Overshoot and settling time are omitted because the frequency setpoint never steps in these scenarios, which leaves both quantities undefined, and the application leaves both blank.

Three comparisons in that table establish the separation derived in Section 8.4. Scenarios 2, 6 and 7 differ only in inertia and give peak rates of change of -0.299, -0.150 and -0.746 Hz/s with nadirs of 59.779, 59.833 and 59.658 Hz, and an identical settled frequency of 59.857 Hz in all three. Scenarios 2, 8 and 9 differ only in droop and give settled frequencies of 59.857, 59.679 and

Table 8.3 Scenario Definitions

#	Scenario	Mode / secondary	Change from default
1	`load-step-2`	manual	$d_L = +2\%$
2	`load-step-5`	auto / droop	$d_L = +5\%$
3	`gen-trip`	auto / AGC	$d_G = 20\%$
4	`renewable-drop`	auto / droop	$d_R = -8\%$
5	`renewable-noise`	auto / AGC	$\sigma_R = 1.5\%$, $d_R = -6\%$
6	`high-inertia`	auto / droop	$H = 10\,\text{s}$, $d_L = +5\%$
7	`low-inertia`	auto / droop	$H = 2\,\text{s}$, $d_L = +5\%$
8	`weak-droop`	auto / droop	$R = 12\%$, $d_L = +5\%$
9	`strong-droop`	auto / droop	$R = 2\%$, $d_L = +5\%$
10	`agc-handover`	auto / both	$d_L = +5\%$, AGC at 10 s
11	`challenge`	auto / AGC	$K_I = 1$, $d_L = +5\%$

Table 8.4 Measured Primary Response at $t = 30\,\text{s}$. Settled Frequency Is the Measured Value Compared With (8.10)

#	IAE Hz s	$\lvert e\rvert_{max}$ Hz	nadir Hz	peak RoCoF Hz/s	settled f Hz
1	1.613	0.088	59.912	−0.120	59.943
2	4.032	0.221	59.779	−0.299	59.857
4	6.450	0.353	59.647	−0.479	59.771
6	3.963	0.167	59.833	−0.150	59.857
7	4.072	0.342	59.658	−0.746	59.857
8	8.866	0.356	59.644	−0.299	59.679
9	1.681	0.144	59.856	−0.299	59.941

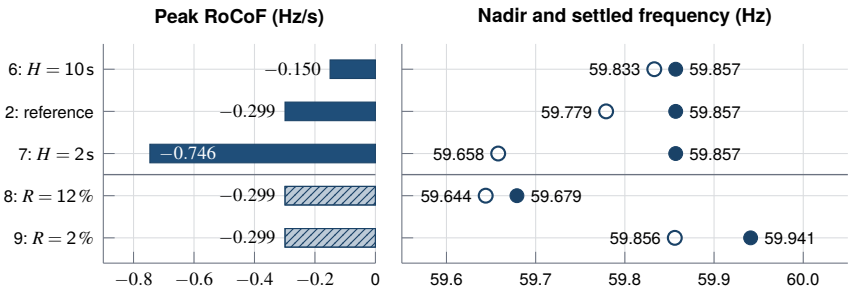


Figure 8.3 The droop-only load steps of Table 8.4: scenarios 6, 2 and 7 change the inertia (solid bars), scenarios 8 and 9 the droop (hatched bars), and each value is printed beside its mark. On the right the open circles are the nadirs and the filled circles the settled frequencies. Notice that inertia moves the peak rate of change and leaves the settled frequency at 59.857, while droop moves the settled frequency and leaves the peak rate at −0.299.

59.941 Hz against the predictions $-0.05/21$, $-0.05/9.33$ and $-0.05/51$ pu, with an identical peak rate of change of $-0.299\,\text{Hz/s}$ in all three. Inertia moves the transient and leaves the steady state unchanged, droop moves the steady state and leaves the transient unchanged, and the manipulated variable travel is 0 pu in every one of these runs because no command is issued. Scenario 4 confirms the disturbance equivalence, because an 8 % renewable loss settles at 59.771 Hz, which is $-0.08/21$ pu below nominal and identical to the result of an 8 % load rise. A load step of the same size added in the same run once the frequency has settled starts at the same −0.479 Hz/s and falls the same 0.353 Hz, from 59.771 to 59.418 Hz. Figure 8.3 sets the five droop-only load steps side by side.

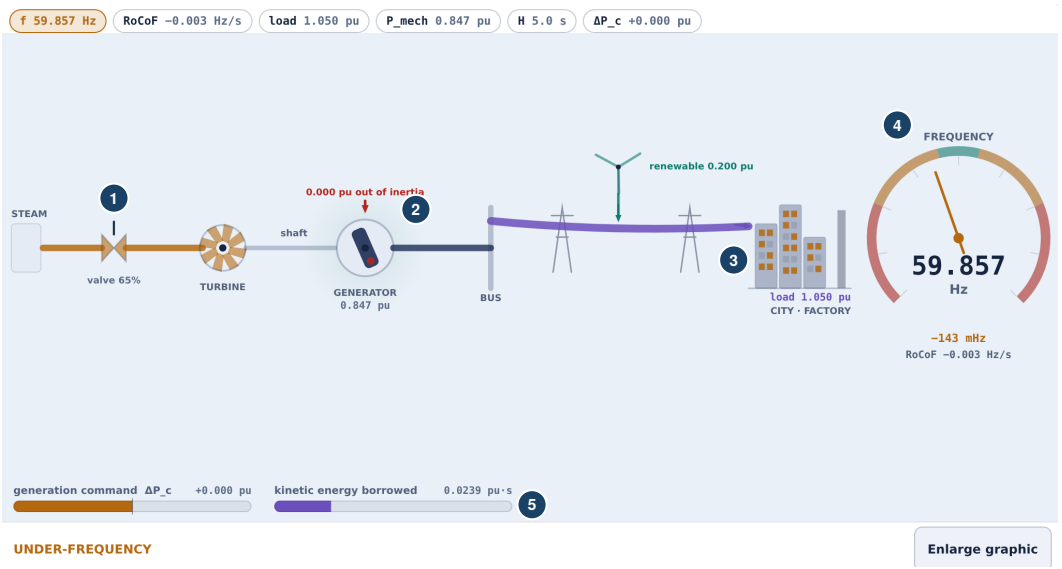


Figure 8.4 Reference scenario of a 5 % load step arrested by governor droop alone, on the Plant view cropped from the application window at $t = 9.7$ s (the clock is in the cut header). The step lands at $t = 2$ s, and the frequency falls at –0.299 Hz/s to a nadir of 59.779 Hz, values that this view does not print (Table 8.4, row 2). The governor valve (1) admits steam to the turbine, the generator (2) gives 0.847 pu and no longer draws on inertia, and the load (3) is the stepped one. The dial (4) reads 59.857 Hz, the 0.143 Hz offset below nominal that (8.10) predicts for an imbalance of 0.05 pu against a stiffness of 21 pu/pu. The borrowed kinetic energy (5) reads 0.0239 pu s. The generation command bar reads zero, inside the ±0.30 pu reserve limits of Table 8.2, and the under-frequency warning at the foot is shown because the frequency is outside the 0.10 Hz normal band.

The manual scenario states the requirement on secondary control as a measured quantity. A 2 % load step arrests at 59.943 Hz, and returning the frequency to 60.000 Hz requires a generation command of 0.020 pu, which is the imbalance and nothing else. A smaller command leaves part of the balance to the governors and leaves a smaller offset, and a larger command overshoots, because at nominal frequency the droop term contributes nothing and the balance closes on the command alone. Commands of 0.015 and 0.025 pu settle at 59.986 and 60.014 Hz.

8.6.2 The Reference Case

The reference scenario is a 5 % load step with droop only, shown in Figure 8.4, captured at $t = 9.7$ s. The frequency response passes through three phases in sequence. From $t = 2$ s the fall is almost straight at the inertial rate, the governors arrest it at $t = 3.22$ s at a nadir of 59.779 Hz, and the turbine lag carries the recovery past the eventual steady state to 59.884 Hz at $t = 4.98$ s, in a decaying oscillation that stays within 5 mHz of 59.857 Hz from $t = 7.3$ s. Mechanical power rises from 0.800 to 0.848 pu against a load that stepped from 1.000 to 1.050 pu, and the remaining 0.0024 pu of the imbalance is closed by the load itself, which draws that much less power at the depressed frequency. The stored energy borrowed from the rotors settles at 0.0238 pu s, which is $2H\,\Delta f$ at the settled offset, and it is held for as long as the offset lasts. The generation command stays at 0.000 pu throughout and the manipulated variable travel is 0 pu, because the secondary loop is out of service in this configuration.

Table 8.5 Measured Secondary Control. Values Are Recorded at the End of Each Run, at 30 s for Scenarios 3 and 10, 60 s for Scenario 5 and 16 s for Scenario 11

#	Secondary	IAE Hz s	MV travel pu	ramp lim. %	final f Hz
3	AGC, $K_I = 6$	2.001	0.200	0.0	60.000
3	AGC, $K_I = 10$	1.855	0.290	5.3	60.000
3	droop only	16.126	0.000	0.0	59.429
5	AGC, $K_I = 6$	3.264	0.326	0.0	60.038
5	AGC, $K_I = 2$	3.422	0.114	0.0	60.046
5	AGC, $K_I = 25$	13.005	4.729	59.8	59.900
5	droop only	9.316	0.000	0.0	59.862
10	AGC from 10 s	1.662	0.050	0.0	60.000
11	AGC, $K_I = 1$	1.505	0.025	0.0	59.927
11	AGC, $K_I = 8$	0.443	0.059	0.0	60.000

8.6.3 Secondary Response

Four scenarios engage automatic generation control, and the measured runs are compared in Table 8.5. Each row uses the scenario as loaded, with any change of mode or gain made before the run starts.

The unit trip of scenario 3 passes through all three tiers in one run. The frequency falls at −1.198 Hz/s to a nadir of 59.144 Hz at $t = 3.14$ s and spends 1.5 s below the 59.5 Hz load-shedding setting, so the flag picks up with secondary control in service. Secondary control then restores 60.000 Hz with the command at 0.200 pu, which is the mechanical power that was lost. The frequency is back within 0.05 Hz 6.4 s after the trip, and the command rate peaks at 0.0855 pu/s, 85.5 % of the ramp limit, without reaching it. With droop only the frequency settles at 59.429 Hz, the $-0.20/21$ pu offset, after a nadir of 59.118 Hz. Raising K_I to 10 puts the command on its ramp limit for 1.6 s, overshoots the frequency to 60.173 Hz and delays the return within 0.05 Hz to 7.5 s after the trip, so a faster integrator buys nothing once the ramp limit binds.

The handover of scenario 10 transfers with no step in the command. The command is still exactly zero after the first execution with secondary control in service at $t = 10.0$ s, and it takes its first increment of 0.001 438 pu at the execution at $t = 10.1$ s, which is one integral increment $K_I e \Delta t_c = 6 \times 2.397 \times 10^{-3} \times 0.1$ and confirms the holding rule of (8.19). The frequency is back within 0.05 Hz 3.1 s after the handover and within 0.02 Hz after 6.1 s, against a secondary time constant of 3.5 s. Halving K_I to 3 lengthens the time to 0.02 Hz to 13.1 s. Doubling it to 12 first reaches 0.02 Hz after 2.2 s, then rings to 60.016 and 59.979 Hz and stays inside only from 5.1 s.

> **TRY IT IN THE APP** **10 · Secondary Control Off → On**
>
> Load the case and press Run. The load step at 2 s settles with its droop offset, and at 10 s the scenario switches secondary control on. Watch the generation command trend at that instant: it does not jump, and it takes its first step of 0.001 438 pu at 10.1 s. The frequency is back within 0.05 Hz 3.1 s after the handover.

The variability case of scenario 5 separates the two parts of a renewable disturbance. Over one minute on the shipped seed, droop alone leaves the persistent 6 % loss as a 0.171 Hz offset and gives nearly three times the IAE of integral action without any command travel. Integral action removes that offset and can only add travel against the zero-mean part, so tripling K_I from 2 to 6 nearly triples the travel and lowers IAE by 5 %, while $K_I = 25$ spends 59.8 % of the run on the ramp limit, swings to 60.47 Hz and ends worse than droop alone. The ranking is not a property of the seed. Over five other seeds the IAE is 8.50 to 9.46 Hz s for droop alone, 2.99 to 3.70 Hz s at $K_I = 2$, 2.44 to 3.40 Hz s at $K_I = 6$ and 10.52 to 12.07 Hz s at $K_I = 25$.

DID YOU KNOW? A Stronger Integral Can Do Worse Than None

From the simulator: under renewable variability (scenario 5) an integral gain of 25 spends 59.8 % of the minute on the ramp limit and ends with an IAE of 13.005 Hz s, worse than the 9.316 Hz s of droop alone, while the default gain of 6 gives 3.264 Hz s (Table 8.5).

Three results for the load steps follow directly from the analysis of Section 8.4. The manipulated variable travel for the unit trip is 0.200 pu and for the handover is 0.050 pu, which are the imbalances themselves, so the integrator settles on the required correction and no travel is wasted. The challenge at $K_I = 1$ gives a secondary time constant of $21/1 = 21$ s and is still 0.073 Hz away at $t = 16$ s, against the $0.143 \exp(-14/21) = 0.073$ Hz that (8.14) predicts, while at the suggested $K_I = 8$ the time constant is 2.6 s and the residual is below the reported resolution. The nadir is 59.781 and 59.788 Hz in those two runs. At $K_I = 8$ the integrator has already commanded 0.018 pu by the nadir at $t = 3.1$ s, but that command reaches mechanical power through the two lags and raises the nadir by only 0.007 Hz. Integral gain sets the settling time and has almost no effect before the arrest.

The challenge asks for the frequency within ± 0.02 Hz of 60 Hz from $t = 10$ s to the end of its 16 s run, a nadir above 59.75 Hz and no time on the 0.10 pu/s ramp limit, with the model parameters and disturbances as loaded, so changing either voids the attempt. With the shipped $K_I = 1$ the worst error after $t = 10$ s is 0.099 Hz, five times the band, although the nadir and the ramp clause pass. With the suggested $K_I = 8$ the worst error is 0.008 Hz, a margin of 0.012 Hz, and the command never reaches the ramp limit. The gain applied before Run and the same gain carried through Reset give one run.

The scorecard leaves overshoot and settling time blank, because the frequency setpoint never steps, and it reports the time on the ramp limit and the time on the reserve limit as separate tiles. No run in Table 8.5 reaches the reserve limit.

8.7 Technology in Practice

The model of Section 8.3 reads one frequency and moves one aggregate generation command through two tiers of control. Figure 8.5 draws both tiers with plant and control-center devices in place of the schematic's symbols.

What Measures It

A real governor reads the speed of its own shaft, while the control center reads the frequency of the bus, and the two readings serve different time scales. The chapter's measurement is perfect, with no filter lag, dead band or telemetry delay, so its peak rate of change, read after the first 0.02 s step, is close to the theoretical value. A relay that averages over several hundred milliseconds may not be triggered by that initial rate if the rate is lower soon after [18]. Until recently the frequency nadir could not be measured reliably in the field and was studied only in simulation [5].

TECHNOLOGY Sensor: Frequency for the Governor and the Control Center

The flyball governors of early plants have largely been replaced by electronic sensors that measure frequency quickly and signal the unit to change its output [19]. Frequency is computed from the voltage over a measuring window. A European guideline asks for a window of 100 to 200 ms and an accuracy of 10 mHz for control at a generating unit, and for 500 ms and 1 mHz for central generation control [20]. Because of that filtering a relay will most likely not see the full initial rate of change, and in the first few cycles after an event different bus-bars read different rates before they converge [18]. Near noisy industrial load the computed rate jumps from sample to sample until it is averaged [20].

What Moves It

The governor of Equation (8.4) and the turbine of Equation (8.3) stand for the valves and steam paths of a whole fleet, with lags of 0.20 and 0.50 s and a ramp limit of 0.10 pu/s on the command, faster than most real units. Typical inertia constants run from 2 to 7 s, lowest in hydro plants and highest

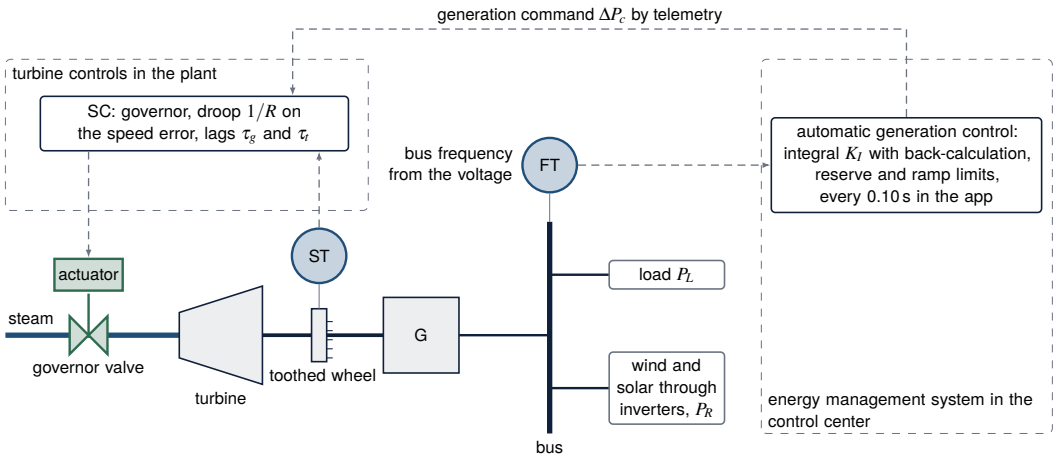


Figure 8.5 The two tiers of the chapter with plant and control-center devices in place of the schematic's symbols. In each plant the governor (SC) reads the shaft speed from a toothed wheel (ST) and moves the governor valve, while the control center reads the bus frequency (FT) and sends the generation command that shifts every governor's setting. The model replaces ST by the bus frequency and the many units by one equivalent machine. Notice that the fast proportional tier never leaves the plant, while the integral tier closes through the frequency of the whole grid.

per unit of capacity in gas plants, and one estimate for the Western Interconnection is about 3.9 s, so the default H of 5 s sits inside that range. A synchronized generator gives the same inertia at any output, but it can give primary response only from head room [19]. In the model the reserve limit of Table 8.2 bounds the generation command, the head room that secondary control can call on.

> **TECHNOLOGY** **Final Element: Turbine Valves That Need Headroom**
>
> When the frequency falls, a governor opens the valves that admit steam or water to its turbine, but the valves and the other mechanical parts take time to move. A typical primary response rises at about 0.3 % of rating per second in slower units and up to 2 % per second in fast units, including certain gas turbines, and in a worked example the output starts to climb a little over half a second after the loss [19]. With a 5 % governor setting a fall of 0.3 Hz raises output by 10 %, but only from a unit running at 90 % of its maximum or less, because a unit at full output has no head room [5].

Where the Controller Runs

Droop sits in each plant's turbine controls. A generating facility is run by a distributed control system that talks to the grid's supervisory system, and its loops run with cycle times from milliseconds to minutes [21]. Automatic generation control runs in the energy management system of the system operations center [15]. Its performance is judged on area control error averaged over one minute for the first control performance standard and over ten minutes for the second [5]. The app runs the integral law with back-calculation of Equation (8.17) every 0.10 s, faster than the multi-second cycle of a real secondary loop.

What Is Modernizing It

Fast frequency response replaces part of the mechanical response with electronics. In Texas, relays on volunteered industrial loads drop them in less than half a second, against the less than 2 % per second of most primary response, and since 2018 new utility-scale wind and solar plants in the United States must be able to respond to frequency. Grid-forming inverters, already used in zero-inertia microgrids typically smaller than 10 MW, can form a grid's voltage and frequency without

synchronous machines [19]. Measurement is changing as well: monitoring that now records the nadir in the field lets planners judge frequency response by the nadir rather than by the settled value [5].

> **IN PRACTICE** **A Governor Free to Respond**
>
> A North American operating policy stated that every generating unit should have an operating governor free to respond to frequency [15]. In service some owners disable their governors or run at maximum output, and others let plant controls restore the scheduled output within seconds, which overrides the governor. On an interconnection's frequency trace this shows as a slope that keeps falling, more slowly, after the frequency first seems arrested. A dead band keeps governors out of normal regulation, so that primary reserve is held for emergencies. The median frequency response of the Eastern and Western Interconnections fell by about a quarter from 2002 to 2008 [5].

8.8 Limitations

The model carries no tie lines, so it does not represent area control error, frequency bias or the inadvertent interchange accounting that occupies most of the practical literature on automatic generation control [6, 9]. The model is the single-area reduction, and the extension to tie-line bias control is not implemented.

A generator trip removes mechanical power but not inertia, which understates the severity of the contingency it is meant to represent. The under-frequency relay annunciates and does not shed, so a nadir below 59.5 Hz is a flag rather than an event and the model never shows the discontinuity that would follow in practice. Frequency measurement is perfect, which removes the filtering, dead band and telemetry delay that force real secondary control onto a multi-second cycle and that dominate its tuning.

The disturbance range extends to a 50 % generation loss, which is well outside the range in which the linear swing equation is accurate. The model integrates that case and reports values, and those values are not predictions of the response of a real system.

> **OPEN CHALLENGE** **How Little Inertia Is Enough**
>
> Inverter-based wind, solar and storage lower a grid's inertia, yet they answer a frequency drop far faster than a turbine valve, which lowers the inertia needed. How much wind and solar this arrangement can carry has not been examined in detail, and the cost, or the need, of a grid that runs with almost no inertia has not been analyzed. How to shape fast response from wind and solar, which need not copy a generator, also needs research [19]. The link between the share of inverters and the rate of change of frequency is not simple, so withstand limits must allow for a plant's whole life [18]. Progress would be studies that find the share at which new approaches are needed [19].

8.9 Discussion

Converter-interfaced generation changes exactly one term in (8.2). Wind and solar plant connected through power electronics supply energy without coupling any rotating mass to system frequency, so displacing synchronous plant reduces H while leaving ΔP_L and P_{loss} untouched [4, 12]. Equation (8.9) then says the rate of change of frequency after a given contingency rises in inverse proportion, and (8.10) says the settled frequency does not move at all. Scenarios 6 and 7 make that pair of statements in one experiment, with the peak rate rising by a factor of 4.98 from -0.150 to $-0.746\,\mathrm{Hz/s}$, against the factor of 5 that (8.9) predicts, and the settled frequency staying at 59.857 Hz to three decimal places.

The consequences are operational rather than theoretical. Relays that disconnect distributed generation on rate of change of frequency are commonly set between 0.5 and 1.0 Hz/s, and a contingency within those settings on a system with high inertia can exceed them on the same system with low inertia, which converts a contained event into a cascading one. Small and weakly interconnected systems reached that limit first, and inertia is now published as an operational quantity

and defended by a dispatch constraint [4, 13]. The stability analysis of systems in which converters supply most of the power is being reformulated, because the assumption under revision is that frequency is a slow, system-wide variable set by rotating mass [14, 16]. Equation (8.12) illustrates the same point on the model of this chapter, because with $1/R = 20$ and the default lags the primary loop itself loses stability below $H = 1.05\,\mathrm{s}$, which is a factor of five below the default value.

8.10 Conclusion

This chapter studied a single-area load frequency control simulator with three states, a governor whose droop feedback is a plant parameter rather than a controller setting, a first-order turbine, reserve and ramp limits on the aggregate generation command, and eleven scenarios. It separated three quantities that a frequency event produces and showed that each belongs to a different group of parameters, with the initial rate of change set by inertia alone at $-f_0(\Delta P_L + P_{loss})/2H$, the arrested offset set by the composite stiffness $1/R + D$ alone at 0.143 Hz for a 5 % step, and restoration to nominal available only from integral action.

Measured runs confirm the separation, with peak rates of change of -0.150, -0.299 and $-0.746\,\mathrm{Hz/s}$ at H of 10, 5 and 2 s against an identical settled frequency of 59.857 Hz, and settled frequencies of 59.941, 59.857 and 59.679 Hz at droop of 2, 5 and 12 % against an identical peak rate of $-0.299\,\mathrm{Hz/s}$. Integral action then restores the unit trip to 60.000 Hz with a command of exactly the 0.200 pu that was lost and without reaching the ramp limit, and the graded challenge is met at $K_I = 8$ with a worst error of 0.008 Hz after $t = 10\,\mathrm{s}$, while integral gains from 1 to 20 move the nadir by less than 0.02 Hz. The app gives a setting for studying proportional offset, integral restoration and two-time-scale control on an integrating process.

KEY POINTS

- Grid frequency is the level of the kinetic energy stored in the spinning machines, so a power imbalance makes it ramp until a control action arrests it.
- Inertia alone sets the initial rate of change, $-f_0(\Delta P_L + P_{loss})/2H$, and the composite stiffness $1/R + D$ alone sets the arrested offset, 0.143 Hz for a 5 % step.
- Droop is proportional action built into governor hardware, so no droop setting removes the offset. Only the integral action of automatic generation control restores 60 Hz.
- Integral gain sets the settling time and has almost no effect before the arrest: gains from 1 to 20 move the nadir by less than 0.02 Hz.

Where else this applies. Chapter 2 leaves the same proportional offset on a tank, Volume 3, Chapter 2 separates a fast inner loop from a slow outer one in the same way, and Volume 17, Chapter 7 restores an island's frequency with droop and an integral loop.

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] P. Tielens and D. Van Hertem, "The relevance of inertia in power systems," *Renewable and Sustainable Energy Reviews*, vol. 55, pp. 999–1009, 2016.

[5] J. H. Eto, J. Undrill, P. Mackin, R. Daschmans, B. Williams, B. Haney, R. Hunt, J. Ellis, H. Illian, C. Martinez, M. O'Malley, K. Coughlin, and K. H. LaCommare, "Use of frequency response metrics to assess the planning and operating requirements for reliable integration of variable renewable generation," Lawrence Berkeley National Laboratory, Berkeley, CA, Report LBNL-4142E, Dec. 2010.

[6] N. Cohn, "Some aspects of tie-line bias control on interconnected power systems," *Transactions of the American Institute of Electrical Engineers. Part III: Power Apparatus and Systems*, vol. 75, no. 3, 1956.

[7] C. Fosha and O. Elgerd, "The megawatt-frequency control problem: A new approach via optimal control theory," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no. 4, pp. 563–577, 1970.

[8] P. M. Anderson and M. Mirheydar, "A low-order system frequency response model," *IEEE Transactions on Power Systems*, vol. 5, no. 3, pp. 720–729, 1990.

[9] Ibraheem, P. Kumar, and D. P. Kothari, "Recent philosophies of automatic generation control strategies in power systems," *IEEE Transactions on Power Systems*, vol. 20, no. 1, pp. 346–357, 2005.

[10] S. K. Pandey, S. R. Mohanty, and N. Kishor, "A literature survey on load–frequency control for conventional and distribution generation power systems," *Renewable and Sustainable Energy Reviews*, vol. 25, pp. 318–334, 2013.

[11] J. W. Simpson-Porco, F. Dörfler, and F. Bullo, "Synchronization and power sharing for droop-controlled inverters in islanded microgrids," *Automatica*, vol. 49, no. 9, pp. 2603–2611, 2013.

[12] A. Ulbig, T. S. Borsche, and G. Andersson, "Impact of low rotational inertia on power system stability and operation," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 7290–7297, 2014.

[13] J. O'Sullivan, A. Rogers, D. Flynn, P. Smith, A. Mullane, and M. O'Malley, "Studying the maximum instantaneous non-synchronous generation in an island system—frequency stability challenges in ireland," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 2943–2951, 2014.

[14] F. Milano, F. Dörfler, G. Hug, D. J. Hill, and G. Verbič, "Foundations and challenges of low-inertia systems (invited paper)," in *2018 Power Systems Computation Conference (PSCC)*. IEEE, 2018, pp. 1–25.

[15] H. F. Illian, "Frequency control performance measurement and requirements," Lawrence Berkeley National Laboratory, Berkeley, CA, Report LBNL-4145E, Dec. 2010.

[16] H. Bevrani, *Robust Power System Frequency Control*, ser. Power Electronics and Power Systems. Cham, Switzerland: Springer International Publishing, 2014.

[17] K. J. Åström and T. Hägglund, *Advanced PID Control*. Research Triangle Park, NC: ISA – The Instrumentation, Systems, and Automation Society, 2006.

[18] ENTSO-E, "Rate of change of frequency (RoCoF) withstand capability: ENTSO-E guidance document for national implementation for network codes on grid connection," European Network of Transmission System Operators for Electricity, Brussels, Guidance document, Jan. 2018. [Online]. Available: https://eepublicdownloads.entsoe.eu/clean-documents/Network%20codes%20documents/NC%20RfG/IGD_RoCoF_withstand_capability_final.pdf

[19] P. Denholm, T. Mai, R. W. Kenyon, B. Kroposki, and M. O'Malley, "Inertia and the power grid: A guide without the spin," National Renewable Energy Laboratory, Golden, CO, Technical Report NREL/TP-6A20-73856, May 2020.

[20] ENTSO-E RG-CE System Protection & Dynamics Sub Group, "Frequency measurement requirements and usage," European Network of Transmission System Operators for Electricity, Brussels, Report Final Version 7, 2018. [Online]. Available: https://eepublicdownloads.entsoe.eu/clean-documents/SOC%20documents/Regional_Groups_Continental_Europe/2018/TF_Freq_Meas_v7.pdf

[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, Gaithersburg, MD, Special Publication 800-82r3, 2023.

8.11 Exercises

8.1 ○ Concept check · 2 min

The app holds the grid at 60 Hz with two control actions on the same frequency error, the governor droop and automatic generation control (AGC). Taken together, what kind of controller do they form, and where does each term act?

(a) A PID controller in the control center, in which the governor droop supplies the derivative term

(b) A split-range controller that hands the load from one group of turbines to the other as the error grows

(c) A Smith predictor in the AGC computer that compensates the delay of the telemetry from the governors

(d) A proportional term built into the governor hardware plus an integral term computed by AGC

8.2 Concept check · 1 min

Scenario 6 repeats the 5 % load step of scenario 2 with the inertia constant doubled from 5 to 10 s. Which of the three numbers of the frequency event stays exactly where it was?

(a) The settled frequency, because H drops out of the steady-state balance
(b) The initial rate of change, because the governors act before inertia matters
(c) The nadir, because the turbine lag alone sets how far the frequency falls
(d) None of them, because every number of the event scales with the stored energy

8.3 Concept check · 2 min

In scenario 1 the app is in Manual and a 2 % load step arrests at 59.943 Hz under droop alone. Which generation command returns the frequency to 60.000 Hz and holds it there?

(a) About 0.019 pu, the share the governors carry at the offset, since load damping supplies the rest
(b) Exactly 0.020 pu, because at 60.000 Hz neither droop nor load damping contributes
(c) About 0.021 pu, the imbalance plus the load that damping shed at 59.943 Hz
(d) About 0.040 pu, twice the step, to overcome the governors pulling back as the frequency rises

8.4 Theory and calculation · 6 min

Scenario 4 removes 8 % of base from the renewable infeed at $t = 2\,\mathrm{s}$ with droop only and the default parameters ($H = 5\,\mathrm{s}$, $D = 1$, $R = 0.05$, $f_0 = 60\,\mathrm{Hz}$).

(a) Find the initial rate of change of frequency and the settled frequency. *Checkpoint:* –0.480 Hz/s and 59.771 Hz.

(b) At the settled offset, how much of the 0.08 pu imbalance does the extra mechanical power carry, and how much does the load itself shed? *Checkpoint:* 0.0762 pu and 0.0038 pu.

(c) Find the kinetic energy borrowed from the rotors at the settled offset, taken as $2H\,|\Delta f|$, in pu s. *Checkpoint:* 0.0381 pu s.

8.5 Theory and calculation · 8 min

Use the characteristic polynomial of Equation (8.11) with $\tau_g = 0.20\,\mathrm{s}$, $\tau_t = 0.50\,\mathrm{s}$ and $D = 1$. A cubic $a_3 s^3 + a_2 s^2 + a_1 s + a_0$ with positive coefficients is stable when $a_2 a_1 > a_3 a_0$.

(a) Write the four coefficients in H, D and R for these lags, and evaluate both sides of the condition at $H = 5\,\mathrm{s}$, $R = 0.05$. *Checkpoint:* $a_2 a_1 = 75.97$ against $a_3 a_0 = 21$.

(b) Find the inertia constant below which the primary loop is unstable at $R = 2\,\%$. *Checkpoint:* $H = 3.21\,\mathrm{s}$.

(c) The completion rules of scenarios 7 and 9 ask the reader to lower H to at most 1.04 and 3.15 s. Say why these two values. *Checkpoint:* Each lies just below the threshold of its droop, 1.05 and 3.21 s.

8.6 Theory and calculation · 6 min

The graded challenge applies the 5 % load step at $t = 2\,\mathrm{s}$ with AGC in service from the start, and asks for the frequency within $\pm 0.02\,\mathrm{Hz}$ of 60 Hz from $t = 10\,\mathrm{s}$ on. Treat the restoration as the first-order response of Equation (8.14), decaying from the 0.143 Hz droop offset from $t = 2\,\mathrm{s}$.

(a) Find the secondary time constant at the shipped $K_I = 1$ and at the suggested $K_I = 8$. *Checkpoint:* 21 s and 2.6 s.

(b) Estimate the offset left at $t = 16\,\mathrm{s}$ at $K_I = 1$. *Checkpoint:* 0.073 Hz.

(c) Find the smallest K_I for which the estimate is inside 0.02 Hz at $t = 10\,\mathrm{s}$. *Checkpoint:* About 5.2.

8.7 App challenge · 7 min

Load scenario 2, +5% load step, droop only. The step lands at $t = 2\,\text{s}$ with Secondary control at Droop only. Use the default seed and run each case for 30 s.

(a) Run the scenario as shipped. Read the Peak RoCoF, Nadir and Offset now tiles at $t = 30\,\text{s}$, and compare the first and last with Equations (8.9) and (8.10). *Checkpoint:* −0.299 Hz/s, 59.779 Hz and 0.143 Hz, each within 0.002, against -0.300 and 0.143.

(b) In Show Model, Parameters, set Governor droop to 12 %, press Reset and run again. Read the same three tiles and the IAE tile. *Checkpoint:* −0.299 Hz/s, 59.644 Hz, 0.321 Hz and 8.87 Hz s, within 0.01 on the IAE.

(c) Which of the three numbers did the droop change, and by what factor did the offset grow? Compare that factor with the ratio of the two stiffnesses. *Checkpoint:* Nadir and offset moved, the rate did not. Offset 2.25 times, and $21/9.33 = 2.25$.

8.8 App challenge · 7 min

Load scenario 10, Secondary control off → on. A 5 % load step lands at $t = 2\,\text{s}$ with droop only, and a scripted event switches Secondary control to Droop + AGC at $t = 10\,\text{s}$. Use the default seed, run each case for 30 s and export the CSV file.

(a) Run the scenario as shipped. In column `total_command_pu`, read the command in the rows at 10.08 and 10.14 s, and find the first logged time after which the frequency stays within 0.02 Hz of 60 Hz. *Checkpoint:* 0 pu and 0.001 438 pu, and 16.1 s, within 0.1 s.

(b) Set AGC integral gain to 3, press Reset and run again. Find the same time. *Checkpoint:* 23.1 s, within 0.1 s.

(c) Express both times from the handover, and compare them with the time that Equation (8.14) predicts for the 0.143 Hz offset to fall to 0.02 Hz. *Checkpoint:* 6.1 and 13.1 s against 6.9 and 13.8 s.

8.9 App challenge · 8 min

Load scenario 3, Generator trip (−20 % of output). A unit carrying 0.20 pu trips at $t = 2\,\text{s}$ with Droop + AGC at $K_I = 6$. Use the default seed and run each case for 30 s.

(a) Run the scenario as shipped. Read the Nadir, Ramp limited and Offset now tiles and the MV travel on the scorecard. *Checkpoint:* 59.144 Hz, 0.0 %, 0.000 Hz and 0.200 pu.

(b) Press the reload button, set AGC integral gain to 10 and run again. Read the Ramp limited tile, the MV travel and the last value of column `peak_frequency_Hz` in the exported CSV file. *Checkpoint:* 5.3 %, 0.290 pu and 60.173 Hz, each within one in the last digit.

(c) The frequency is back within 0.05 Hz of 60 Hz 6.4 s after the trip at $K_I = 6$ and 7.5 s after it at $K_I = 10$. Explain why the faster integrator is slower here. *Checkpoint:* The command rides the 0.10 pu/s ramp limit, and the excess demand overshoots.

8.10 Application area · 4 min

Relays that disconnect distributed generation on the rate of change of frequency are set between 0.5 and 1.0 Hz/s. Use Equation (8.9).

(a) What smallest inertia constant keeps the initial rate of change of a 5 % load step within 0.5 Hz/s? What would the 20 % trip of scenario 3 need? *Checkpoint:* 3.0 s, and 12 s.

(b) The model keeps H when a unit trips (Section 8.8). Suppose the tripped unit also held 20 % of the stored energy, so H falls from 5 to 4 s on the same base. Find the initial rate of change, and compare it with the model's. *Checkpoint:* −1.50 Hz/s against −1.20 Hz/s.

8.11 Application area · 3 min

Two generating units of equal rating run in parallel on one bus, one with 4 % droop and one with 5 %.

(a) What share of a load increase does each unit take once the frequency has settled? *Checkpoint:* 55.6 % and 44.4 %.

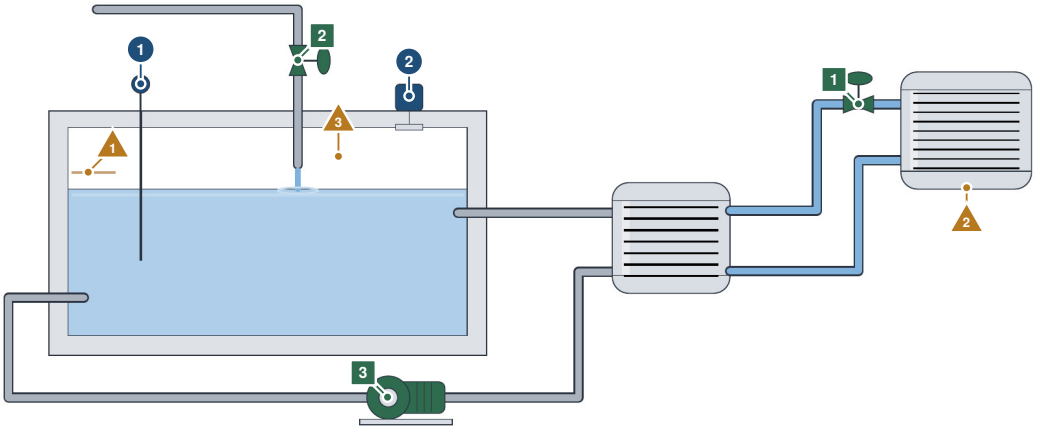
(b) What would happen if both governors were isochronous instead? *Checkpoint:* The division of load between them is indeterminate.

8.12 Review the agent · 5 min

A planner asks an AI assistant for an integral gain that passes the graded challenge with margin. It replies: "Use $K_I = 30$. The secondary time constant is $(1/R + D)/K_I = 21/30 = 0.7\,\mathrm{s}$, so the 0.143 Hz offset is gone in about 1.4 s, long before $t = 10\,\mathrm{s}$. Integral action cannot act before the arrest, so the nadir stays at about 59.78 Hz, and the gain only shortens restoration." Find the flaw and say how to check it.

CHAPTER 9

Spent-Fuel Pool Simulator for Integrating Inventories and Time-to-Limit Prediction



The controller moves 1 cooling water valve 2 makeup valve 3 cooling train pump

The controller reads 1 pool temperature 2 pool level

What disturbs it 1 water lost to evaporation or a leak 2 cooling water temperature 3 air over the pool

Notice the two separate paths: the cooling train carries heat out to the cooling water, while only the makeup line puts back the water the pool loses. Illustration, not to scale.

IN PLAIN TERMS A Warm Pool That Can Run Out of Time

The main idea. Used nuclear fuel keeps making heat for years, so it is stored under deep water in a pool. Pumps send the water through a cooler, and a valve adds water to replace what slowly evaporates.

The challenge. If the pumps stop, the pool warms so slowly that nothing seems wrong for hours. Warm water also swells, so the level can rise even while water is being lost.

Why it matters. The useful number is how many hours are left before the pool boils, and working it out step by step beats stretching a straight line.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the history of the dense racks and the fact on real pools (Section 9.2), the try-it of scenario 3 (Section 9.4), the fact on feedforward and the try-it of scenario 10 (Section 9.6) and the key points
Study	65 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	20 min	the first page, the results (Section 9.6), Technology in Practice (Section 9.7), Limitations (Section 9.8) and the open challenge

The control question The level of a spent-fuel pool is a pure integrator, and its temperature is a self-regulating first-order plant only while a cooling train runs. How does one step test tell which plant the pool is, and what should an operator compute when cooling is lost and the useful number is the time left to a limit?

Learning objectives

- Explain from the two balances why the level integrates and why the temperature turns into an integrator when both trains stop.
- Compute the temperature time constant and the heat-up rate from the thermal capacity and the feedback of the exchanger and the surface.
- Identify a first-order and an integrating model from one step test, and state the tuning each one commits the loop to.
- Design an averaging level loop, and predict its offset when the disturbance is a leak rather than evaporation.
- Compare a marched time to a limit with a present-rate extrapolation, and feedforward from computed evaporation with level feedback.

Prerequisites Sections 1.4.2 and 1.4.3 (balances and linearization), 1.4.4 (an integrator as a long lag), 1.4.6 (effective dead time) and 1.6.2 (velocity form). Chapter 2 shows a tank that regulates itself.

The app https://apmonitor.com/apps/spent-fuel-pool/

See also Volume 6, Chapter 4 for averaging level control and fouling that walks toward a limit, Volume 6, Chapter 10 for leak detection by mass balance. Volume 18, Chapter 10 for decay heat on a pool that boils, Volume 18, Chapter 2 for regulation against self-regulation.

9.1 Introduction

A slow process is usually assumed to be a forgiving one. A pool of 1231 m^3 of borated water over 448 discharged fuel assemblies rejects 2148 kW through two cooling trains while losing 39.5 kg/h of water to evaporation, which is 0.37 mm/h of level. If both trains stop the pool warms at 1.49 °C/h. Nothing on the trend display moves during the first shift. The assumption that follows from that appearance is wrong twice over. It is wrong because an integrating plant does not forgive an error, it accumulates one, so the deviation at any moment is the imbalance multiplied by however long nobody acted. It is wrong again because the two rates just quoted are the rates now, and a straight line drawn through the last hour of either trend predicts the next ten badly. The pool boils 1.1 h later than the temperature ramp says it will, and reaches its shielding limit 21.9 h into a loss of cooling whose first 8.8 h carry a level trend that is rising.

A simulator supplies the part a transfer function cannot. The plant in question is not one plant. With a train running the exchanger duty rises with pool temperature at 423 kW/K, which is real negative feedback, and the pool is first order with a time constant of 3.33 h. With both trains stopped only 4.20 kW/K of surface loss remains and the time constant is 339 h, which no controller and no operator can use as a time constant. A student handed either model separately learns a correct thing about a plant that does not exist by itself, and a student handed both learns the thing that matters, which is that the identification depends on which equipment happens to be in service and that operating experience built during normal operation is the wrong preparation for the day the pumps stop. Surveys of what control graduates are short of keep returning to the same gap, which is practice at getting a model out of a plant instead of being handed one in a problem statement [1, 2].

The general class is broader than a nuclear plant. It is an integrating plant whose apparent dynamics depend entirely on whether its dominant negative feedback path is in service, carrying two coupled inventories, and for which the operative quantity during an upset is a time to a limit rather than an error. Large tank farms and surge inventories behave this way whenever the discharge is set by a pump rather than by hydrostatic head, and level loops on them are detuned on purpose so that the inventory absorbs upsets instead of passing them downstream [3]. Building thermal mass behaves

this way when the plant is off, which is why predictive control of a building is written against a stored-energy state and a forecast rather than against a temperature error [4]. The state of charge of a battery is the same integrator with a different unit. Tuning rules for integrating plants with dead time make the dead time the whole of the denominator of the gain [5], and the identification that supplies that dead time is a slope and a delay rather than a gain and a settling point.

The second difficulty is that the two inventories are coupled by a mechanism that appears on neither loop's block diagram. Evaporation from the pool surface takes heat and water away together, it is driven by the vapor-pressure difference to air on the refueling floor that nobody controls, and it grows steeply with pool temperature, from 39.5 kg/h at 35 °C to 217 kg/h at 55 °C and 1864 kg/h as the surface approaches saturation. The same disturbance is negligible in the energy balance and dominant in the mass balance, so the loop that could reject it does not need to and the loop that needs to cannot. However, a level measurement on a pool of 108 m^2 of surface is a very poor flowmeter, because 39.5 kg/h of evaporation is 8.8 mm of level a day against 3 mm of transmitter noise.

This chapter works with an interactive spent-fuel pool in which the same temperature measurement identifies two different plants depending on what is running, and in which the inventory loop is answered by computing a disturbance rather than by tuning a controller. It measures how far a marched time-to-limit calculation beats a present-rate extrapolation on this plant, and why the first-day level deviation under feedback contains no controller gain at all. The pool model and the tuning arithmetic are standard. Section 9.2 traces how at-reactor pools came to hold what they hold and how their instruments were judged after 2011. Section 9.3 sets out the seven states, the 31 constants and the two instrument chains closely enough that the pool can be rebuilt from the page, with decay heat summed over a mixed-age inventory and evaporation from a published mass-transfer correlation. Section 9.4 linearizes both balances and identifies the pool twice from one step test. Section 9.5 states the four control architectures, the tuning rules the identification commits to, the evaporation feedforward and the two time-to-limit calculations. Section 9.6 measures the twelve scenarios, from a pool in which nothing appears to happen to a graded challenge. Section 9.7 sets the model beside the instruments and valves of real pools, Section 9.8 states the limitations, Section 9.9 weighs makeup water against shielding and Section 9.10 draws the results together.

9.2 Background

The pool was never intended to hold what it now holds: water storage at the reactor was designed as a short buffer and became a long-term store, as the box below traces, and original open racks were replaced by high-density racks analyzed as a criticality problem in their own right [6], and the hazard of the dense inventory was reviewed by the National Academies [7].

> ⧖ HISTORY **Sized for Reprocessing, Packed Instead (1974 to 1982)**
>
> Pools in the United States, like those in Europe and Japan, were first sized to hold spent fuel for only a few years, until it was cool enough to travel to a reprocessing plant that would recover its plutonium and uranium. After India tested a nuclear explosive made with plutonium recovered for peaceful purposes in 1974, the United States halted in 1977 the licensing of an almost completed reprocessing plant. A 1982 act promised a deep underground repository that was to begin accepting fuel by 1998. Operators packed their pools instead. Racks once spaced in a loose square array gave way to dense-pack racks at 23 cm, barely above the 21.4 cm of a reactor core, with each assembly in a box whose walls hold neutron-absorbing boron, and virtually every pool in the country was re-racked [8].

Thermal analysis of the pool followed the same path from short buffer to long-term store. The early treatment is a lumped energy balance on a well-mixed pool against decay heat, with evaporation and surface loss as the only passive sinks, extended to the local temperature distribution around the racks and to the natural circulation that carries heat out of them [9]. Decay heat itself is the forcing function and it is the one quantity the analysis cannot choose, being fixed by burnup and cooling time,

and modern practice validates decay heat power for light water reactor fuel against depletion codes rather than against a single correlation [10]. Evaporation received much less attention than decay heat until it mattered, and the correlations available for undisturbed water surfaces were assembled and tested against a wide range of pools, including spent-fuel pools, only recently [11, 12].

Cooling and makeup requirements evolved from the licensing basis rather than from control practice. Two trains, each a pump and a plate exchanger to a component cooling water system, remove decay heat and hold the bulk temperature below a technical-specification limit. Makeup replaces evaporation, is very often started by hand, and is sized generously because the case it has to cover is a boiling pool rather than an evaporating one. The depth of water above the fuel is not there to cool anything at all. Seven meters of water over the tops of the assemblies is shielding, and the technical-specification minimum in this application is 6.40 m, so the level constraint that governs the inventory loop is a dose-rate constraint and not a heat-transfer constraint.

> ✦ **DID YOU KNOW? Real Pools Are Built to This Scale**
>
> Pools in light water reactors are typically about 12 m deep, with approximately 7 m of water over the stored fuel. A 2006 French assessment found that during a refueling outage a total loss of cooling could bring a pool to boiling within five to eight hours [13].

March 2011 changed what was measured. Loss of cooling at Fukushima Daiichi left operators without a reliable indication of pool level or pool temperature at exactly the point where those were the only quantities that mattered, and the response in the years that followed was directed at instrumentation and at inventory rather than at cooling capacity. The National Academies review of the accident treats pool instrumentation and the ability to add water by alternate means as central findings [14]. Analysis of pool behavior during a station blackout became a question about the joint history of temperature and level rather than about a peak temperature [15], and the evaporation rate from the damaged pools was reconstructed from a mass-transfer model to check the decay heat estimates against observation [16]. That reordering of priorities is the reordering this application asks a student to make. The temperature trend goes flat at 100 °C and stops carrying information, at which point the level is the only measurement left and the useful number is how many hours remain rather than how many degrees the pool is off setpoint.

9.3 Pool Model

The pool is 11.40 m deep with 108 m^2 of free surface, holding 1231 m^3 of borated water at 35.00 °C over racks whose tops are at 4.40 m, so 7.00 m of water covers the fuel. The design inventory is 448 assemblies, six batches of 64 at 3.0, 4.5, 6.0, 7.5, 9.0 and 10.5 y producing 682.7 kW between them, plus a batch of 64 thirty days out of the reactor producing 1465.0 kW, for a total decay heat of 2147.7 kW. Two cooling trains take suction below the surface, pass the water through a plate exchanger to component cooling water at 30.0 °C and return it to the pool. A makeup line replaces evaporation. Figure 9.1 shows the arrangement and every symbol that appears in the equations.

Table 9.1 classifies what the two loops interact with. Four limits bound normal operation, a high pool temperature alarm at 52 °C, a technical-specification temperature limit at 60 °C, a low-level alarm at 6.70 m of shielding and the technical-specification shielding minimum at 6.40 m. None of them stops a run. The two alarms color the badges and annunciate on the status bar while the condition holds, and the two technical-specification limits also mark the trend when they are crossed and total the hours spent outside them. The plant is expressed in the units an operator reads, temperatures in °C, masses in kg, mass flows in kg/h, volumes in m^3, volume flows in m^3/h, duties in kW, elevations in m and valve positions in percent. **Time is in hours throughout.** Seven states are integrated.

🕮 **Pause.** Try Exercise 9.2 before reading on.

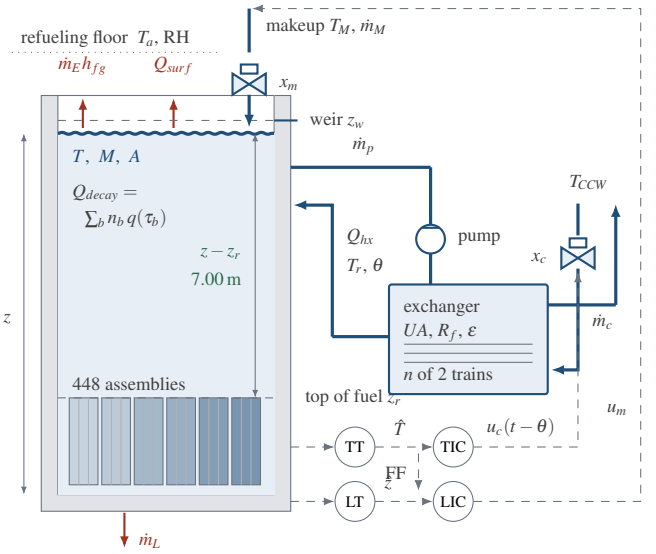


Figure 9.1 Spent-fuel pool with its cooling train and makeup line. Water of mass M at bulk temperature T stands at elevation z over racks whose tops are at z_r, so $z - z_r$ is the shielding depth the license is written around. Decay heat Q_{decay} is summed over batches of n_b assemblies at age τ_b. Rack shading darkens as batch age falls, with the freshest batch darkest and at the right. One of n trains draws $\dot{m}_p$ through a plate exchanger of conductance UA and fouling resistance R_f, returning water at T_r after a transport delay θ and rejecting Q_{hx} to component cooling water at T_{CCW} through butterfly valve x_c. Evaporation $\dot{m}_E$ and surface loss Q_{surf} leave through the free surface of area A against refueling floor air at T_a and relative humidity RH, makeup at T_M enters through equal-percentage valve x_m, and any liner leak $\dot{m}_L$ leaves at the floor. Water above the weir z_w spills. The makeup command u_m carries the evaporation feedforward computed from $\hat{T}$, T_a and RH.

9.3.1 Energy and Mass Balances

The two accumulations are Equations (9.1) and (9.2), written as Section 1.4.2 writes them for a heated tank.

$$Mc_p \frac{\mathrm{d}T}{\mathrm{d}t} = Q_{decay} - Q_{hx} - Q_{surf} - \dot{m}_E h_{fg} + \dot{m}_M c_p (T_M - T) \tag{9.1}$$

$$\frac{\mathrm{d}M}{\mathrm{d}t} = \dot{m}_M - \dot{m}_E - \dot{m}_L - \dot{m}_{boil} \tag{9.2}$$

where M is the pool water mass in kg, $c_p = 4.182\,\mathrm{kJ/(kg\,K)}$, T is the bulk temperature, Q_{decay} is decay heat, Q_{hx} is the duty of the running trains, Q_{surf} is convection plus radiation from the free surface, $\dot{m}_E$ is evaporation, h_{fg} is the latent heat, $\dot{m}_M$ is makeup at temperature T_M, $\dot{m}_L$ is a liner leak and $\dot{m}_{boil}$ is boil-off. Both sides of (9.1) carry kJ/h once the duties are multiplied by 3600 s/h, and both sides of (9.2) carry kg/h. At the design point $Mc_p = 5.120\,\mathrm{GJ/K}$.

Nothing on the right-hand side of (9.2) is a function of M, and that absence is the cleanest integrating structure in the plant. A makeup valve one percent from the balancing position produces a level ramp that never stops and never settles anywhere. The application verifies the property directly by evaluating $\mathrm{d}M/\mathrm{d}t$ at two elevations 1.9 m apart and requiring the two values to agree to within 1×10^{-9} kg/h. Equation (9.1) is different, because Q_{hx}, Q_{surf} and $\dot{m}_E$ all rise with T, and Section 9.4 quantifies how much of that self-regulation survives when the pumps stop.

Table 9.1 What the Two Loops Interact With. The Nominal Column Is the 35 °C Design Point the Pool Opens At.

Symbol	Role	Nominal	Range
T (TT)	CV, measured	35.00 °C	20–60 °C SP
$z - z_r$ (LT)	CV, measured	7.00 m	6.40–7.20 m SP
x_c (CCW)	MV	44.20 %	0–100 %
x_m (makeup)	MV	2.37 %	0–100 %
n trains	MV, discrete	2	0, 1, 2
offload	DV, measured	64 assemblies	0–193
trains available	DV, measured	2	0–2
T_{CCW}	DV, measured	30.0 °C	15–45 °C
T_a	DV, measured	30.0 °C	10–45 °C
RH	DV, measured	45 %	10–95 %
T_M	DV, measured	20.0 °C	5–35 °C
ventilation	DV, unmeasured	1×	0.4–2×
R_f fraction	DV, unmeasured	0.15×	0–1×
$\dot{m}_L$ leak	DV, unmeasured	0 m³/h	0–5 m³/h
T	state	35.00 °C	$\le T_{sat}$
M	state	1.2243×10^6 kg	$\ge \rho A z_r$
T_r	state	32.33 °C	—
x_c, x_m	states	44.20, 2.37 %	0–100 %
R_f	state	9.0×10^{-5} K/kW	$0–6 \times 10^{-4}$ K/kW
T_{cw}	state	30.0 °C	—

Level and shielding follow from the mass through the water density, as shown in Equation (9.3).

$$z = \frac{M}{\rho(T) A}, \qquad \rho(T) = 1001.6 - 0.0872\,T - 0.00337\,T^2 \tag{9.3}$$

where z is the surface elevation above the pool floor in m, $A = 108\,\text{m}^2$ is the free surface area and ρ is in kg/m³ from a quadratic fit good to 1.5 kg/m³ between 20 and 100 °C. The shielding depth is $z - z_r$ with $z_r = 4.40\,\text{m}$. Because ρ depends on T, differentiating (9.3) gives the level rate in Equation (9.4).

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \frac{1}{\rho A}\frac{\mathrm{d}M}{\mathrm{d}t} - \frac{z}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}T}\frac{\mathrm{d}T}{\mathrm{d}t} \tag{9.4}$$

The second term is positive for a warming pool because $\mathrm{d}\rho/\mathrm{d}T$ is negative, so a heating pool rises while it is losing water. That term is the reason a level transmitter cannot distinguish thermal expansion from inventory and the reason a present-rate extrapolation of the level gets the sign wrong early in a loss of cooling. A level that rises while the vessel loses water is the swell that Chapter 6 meets in a flash vessel, there from vapor held below the surface and here from the density of warming water.

Above the skimmer weir at $z_w = 11.60\,\text{m}$ the surface is pinned and the excess leaves as spill, as shown in Equation (9.5).

$$\dot{m}_{spill} = \max\left(0,\ \frac{\mathrm{d}M}{\mathrm{d}t} - A z_w \frac{\mathrm{d}\rho}{\mathrm{d}T}\frac{\mathrm{d}T}{\mathrm{d}t}\right) \quad \text{when } z \ge z_w \tag{9.5}$$

The subtracted term is the mass the pinned surface must shed as the water expands, so heating alone sends water over the weir while the inventory is falling. Equation (9.5) is also what makes feed and spill work as an emergency heat sink, because makeup can be run far above the loss rate and the surplus leaves over the weir.

At saturation the pool cannot get hotter, so any remaining surplus leaves as steam, as shown in Equation (9.6).

$$\dot{m}_{boil} = \frac{M c_p}{h_{fg}(T)} \left.\frac{\mathrm{d}T}{\mathrm{d}t}\right|_{\text{unclamped}}, \qquad \frac{\mathrm{d}T}{\mathrm{d}t} = 0 \quad \text{when } T \geq T_{sat} \tag{9.6}$$

with $T_{sat} = 100\,^{\circ}\mathrm{C}$. Past this point the temperature trend is flat and the level trend carries the whole of the information.

9.3.2 Decay Heat Over a Mixed-Age Inventory

Decay power per assembly is a three-term fit, as shown in Equation (9.7).

$$q(\tau) = 88.08\,\tau^{-0.5277} + 7.975\,\mathrm{e}^{-\tau/188.6} + 1.491\,\mathrm{e}^{-\tau/1160} \tag{9.7}$$

where q is in kW per 17×17 pressurized water reactor assembly and τ is days since shutdown. The power law carries the fission products in aggregate and the two exponentials carry the short-lived and medium-lived groups a single power law cannot. Equation (9.7) returns 40.71 kW at 7 d, 22.89 kW at 30 d, 6.15 kW at 1 y, 1.98 kW at 5 y and 0.81 kW at 20 y, and it is a least-squares fit to decay powers of the kind the ANSI/ANS-5.1 standard produces for a burnup of 45 GWd/tU, with a worst residual near 5 % over that range. On logarithmic axes it falls with a mean slope of -0.56 between one week and twenty years, steepening from -0.40 over the first month to -0.70 between one and five years, which is why a pool remains a heat source for decades. The same power-law clock drives the shut-down reactor of Volume 18, Chapter 10, whose only heat sink is a pool that boils to atmosphere. The total is summed over batches and each batch ages as the simulation advances, as shown in Equation (9.8).

$$Q_{decay}(t) = \sum_{b=0}^{n_{aged}-1} n_b\, q(\tau_b + t) + n_{fresh}\, q\big(\tau_{fresh} + t\big) \tag{9.8}$$

A full core of 157 assemblies eight days out of the reactor raises the total from 2148 kW to 6731 kW, and the same core twelve days out gives 5816 kW.

9.3.3 Evaporation and Surface Loss

Evaporation is the natural-convection correlation for undisturbed water surfaces, in which the driving force is the buoyancy of the humid layer standing over the pool and the potential is a humidity-ratio difference, as shown in Equation (9.9) [11, 12].

$$E'' = C_E\, \rho_w\, (\rho_a - \rho_w)^{1/3}\, (W_w - W_a), \qquad C_E = 35 \tag{9.9}$$

where E'' is in kg/(m^2 h), ρ_w is the air density at the water surface temperature and saturated, ρ_a is that of the room air, and W_w and W_a are the corresponding humidity ratios, $W = 0.622\, p_v/(p_{atm} - p_v)$. Both densities are the mass of dry air per unit volume of moist air, which is the density a psychrometric chart carries and the definition the published model uses, so each is evaluated as $\rho = (p_{atm} - p_v)/(0.287\, T_K)$ and the vapor partial density is not added to it. The coefficient $C_E = 35$ belongs to this set of units and to no other. Saturation pressure comes from Buck's correlation, accurate to 0.05 % below 100 °C [17], as shown in Equation (9.10).

$$p_{sat}(T) = 0.61121\, \exp\left[\left(18.678 - \frac{T}{234.5}\right)\frac{T}{257.14 + T}\right] \tag{9.10}$$

with p_{sat} in kPa and T in °C. Total evaporation is $\dot{m}_E = E'' A f_v$ with f_v the ventilation multiplier, and the correlation is evaluated at a surface temperature capped 5 °C below saturation because W_w diverges there. The underlying model has been tested against 24 studies covering pool areas from 0.022 to 525 m^2, water temperatures from 7 to 94 °C, air temperatures from 6 to 51 °C, relative humidities from 16 to 98 % and air velocities to 1.9 m/s, with a mean absolute deviation of 13.9 % [12], and it

has been applied to spent-fuel pools specifically [16]. The pool of this application sits inside that range on every variable except that the boiling branch of (9.6) takes over above 95 °C, one degree above the highest water temperature in the validation set.

One departure from the published model has to be stated, because the evaporation term drives the whole of the loss-of-cooling result. The published model returns the larger of (9.9) and a vapor-pressure floor of $5 \times 10^{-5}\,(p_w - p_a)$ in the same units with the pressures in Pa, and the application omits the floor. Evaluated over the whole operating envelope of this pool the floor is always the smaller of the two, at 0.19 kg/(m^2 h) against 0.37 kg/(m^2 h) at the design point, 0.10 kg/(m^2 h) against 0.15 kg/(m^2 h) in a hot and humid building and 4.13 kg/(m^2 h) against 17.26 kg/(m^2 h) at 95 °C, so omitting it changes no result in this chapter. The density definition, which is the one place an implementation of this correlation most easily goes wrong, was checked term by term against the published nomenclature and agrees with it, so the rates below are the rates the correlation gives for this pool rather than an implementation of something near it.

Equation (9.9) as implemented returns 39.5 kg/h at the design point and 104.2 kg/h at 45 °C, 217.2 kg/h at 55 °C, 705.1 kg/h at 75 °C and 1864 kg/h at 95 °C, which is a doubling for every 9.6 °C averaged between 35 and 75 °C. The latent heat is a linear fit, $h_{fg} = 2500.8 - 2.36\,T$ in kJ/kg, within 9 kJ/kg over 20 to 100 °C.

Sensible loss from the surface is convection plus gray-body radiation, as shown in Equation (9.11).

$$Q_{surf} = A\left[\frac{h_c(T - T_a)}{1000} + \varepsilon\sigma\left(T_K^4 - T_{a,K}^4\right)\right] \tag{9.11}$$

where $h_c = 3.5\,\mathrm{W/(m^2\,K)}$, $\varepsilon = 0.95$, $\sigma = 5.67 \times 10^{-11}\,\mathrm{kW/(m^2\,K^4)}$ and T_K denotes absolute temperature. At the design point (9.11) returns 5.21 kW against 26.56 kW of evaporative duty, so the surface removes 31.8 kW of the 2148 kW total.

9.3.4 Cooling Train and Makeup Valve

Each train is an effectiveness-NTU counter-flow exchanger with the fouling resistance in series and a turbulent film coefficient on the component cooling water side, as shown in Equations (9.12) and (9.13).

$$UA = \left(\frac{1}{UA_p} + \frac{1}{UA_c\,f^{0.8}} + R_f\right)^{-1}, \qquad f = \frac{x_c}{100} \tag{9.12}$$

$$k = \varepsilon\,C_{\min}, \qquad Q_{hx} = n\,\dot{m}_p c_p (T - T_r) \tag{9.13}$$

where $UA_p = UA_c = 1800\,\mathrm{kW/K}$, $C_{\min}$ and $C_{\max}$ are the smaller and larger of the pool-side capacity rate $\dot{m}_p c_p = 395.6\,\mathrm{kW/K}$ and the component cooling water capacity rate $\dot{m}_c c_p f$, $C_r = C_{\min}/C_{\max}$, $N = UA/C_{\min}$, and ε is the counter-flow effectiveness $(1 - \mathrm{e}^{-N(1-C_r)})/(1 - C_r\mathrm{e}^{-N(1-C_r)})$, reduced to $N/(1+N)$ when $C_r > 0.999$. The duty coefficient k per train is strongly nonlinear in valve position, giving 63.4 kW/K at 10 % open, 141.9 kW/K at 25 %, 227.1 kW/K at 50 % and 303.7 kW/K wide open at the design fouling of 9.0×10^{-5} K/kW, against 311.7 kW/K clean and 264.4 kW/K at the end-of-season resistance of 6.0×10^{-4} K/kW. The return temperature is a state with its own thermal lag, as shown in Equation (9.14).

$$\tau_{hx}\frac{\mathrm{d}T_r}{\mathrm{d}t} = \left[T - \frac{k\,(T - T_{cw})}{\dot{m}_p c_p}\right] - T_r \tag{9.14}$$

with $\tau_{hx} = 0.08\,\mathrm{h}$. Substituting the bracketed steady value into (9.13) gives $Q_{hx} = nk(T - T_{cw})$, so the exchanger duty is a conductance multiplied by a temperature difference and nk is exactly the negative feedback the pool has. With two trains at 44.20 % that conductance is 423.0 kW/K, and wide open it is 607.5 kW/K.

Pause. Try Exercise 9.1 before reading on.

The makeup valve has equal-percentage trim, as shown in Equation (9.15).

$$q_M = q_{M,\max}\,\frac{R^{x_m/100}-1}{R-1}, \qquad R = 50 \tag{9.15}$$

where $q_{M,\max} = 20\,\mathrm{m^3/h}$ and R is the rangeability. The same trim passes 0.0396 m³/h at 2.37 % of travel, which is the design evaporation rate, and 20 m³/h wide open, so its duty spans a factor of 505 and equal-percentage trim is the reason one valve covers it. The linear gain at the design position is 0.0175 m³/h per percent of travel, which by (9.3) is 1.622×10^{-4} m/h of level per percent.

9.3.5 Instruments

Neither controller sees the plant directly. A Gaussian sequence is passed through a first-order filter and rescaled so that what comes out carries the standard deviation the instrument is specified at, that is added to the true value, and the sum then goes through the transmitter's own filter, as shown in Equation (9.16).

$$\begin{aligned} \tau_{Tm}\frac{\mathrm{d}\hat{T}}{\mathrm{d}t} &= T + n_T(t) - \hat{T} \\ \tau_{Lm}\frac{\mathrm{d}\widehat{(z-z_r)}}{\mathrm{d}t} &= (z-z_r) + n_L(t) - \widehat{(z-z_r)} \end{aligned} \tag{9.16}$$

with $\tau_{Tm} = 0.05\,\mathrm{h}$, $\tau_{Lm} = 0.10\,\mathrm{h}$, noise correlation times of 0.1 and 0.3 h, and default standard deviations of 0.05 °C and 0.003 m. Both filters step forward on the closed-form hold solution $\hat{x}_k = \alpha\hat{x}_{k-1} + (1-\alpha)x_k$ with $\alpha = \exp(-\Delta t/\tau)$, so no choice of sample period can destabilize them. Filtered rates of change of both measurements are reported separately in °C/h and m/h using filters of 0.6 and 1.5 h, and those are the instruments a step test is read from.

9.3.6 Constants, Opening State and Integration

Table 9.2 carries the 31 constants. Twenty-four of them are on a slider somewhere in the interface and the remaining seven are never shown at all, and it is the seven invisible ones that fix how the pool answers a step.

No scenario starts from a guess. Bisection on the net heat of (9.1), monotone decreasing in T, returns the pool temperature that the requested train count and valve position actually hold. The mass is set to put the surface at z_r plus the shielding setpoint, the return temperature to its own steady value from (9.14), the fouling resistance to its commanded fraction of $R_{f,\max}$ and the makeup valve to the position that (9.15) makes equal to evaporation plus any leak. In automatic modes the valve position is instead found by a second bisection for the opening that holds the temperature setpoint. The one exception is on/off train control, which has a limit cycle rather than a steady state, and so starts at the setpoint with the train stopped. The application checks the resulting state by requiring every scaled derivative to be below 1×10^{-7} and the temperature to be within 0.01 °C of 35 °C.

States advance on fourth-order Runge-Kutta with the step held at 0.01 h, which is 36 s, and both controllers scan on that same 36 s boundary. The historian records ten points per simulated hour. The engine, its seeding and its Reset are those of Section 1.6.1. Noise is drawn from a generator seeded at 20 260 906, so a run with the RTD noise switched on repeats to the last digit, and the five speed settings of 0.5, 1, 2, 5 and 20 simulated hours per wall-clock second change only how long a reader waits. A finished run leaves the browser as a comma-separated file of 98 columns, holding the true temperature beside the measured one, the commanded valve beside the actual stem, the four heat-balance terms beside the four mass-balance terms, and both time-to-limit calculations beside each other.

Table 9.2 The 31 Model Constants. The Seven Marked Hidden Never Reach the Screen, and a Step Test Is the Only Route to Them.

Symbol	Meaning	Value	Unit	Visibility
A	free surface area	108	m^2	adjustable
z_r	top of the assemblies	4.40	m	adjustable
z_w	skimmer weir	11.60	m	adjustable
—	shielding minimum	6.40	m	adjustable
—	low-level alarm	6.70	m	adjustable
—	temperature alarm	52	°C	adjustable
—	temperature limit	60	°C	adjustable
T_{sat}	saturation temperature	100	°C	adjustable
n_b	assemblies per batch	64	—	adjustable
n_{aged}	aged batches	6	—	adjustable
τ_0	youngest aged batch	3.0	y	adjustable
—	years between discharges	1.5	y	adjustable
τ_{fresh}	fresh batch cooling time	30	d	adjustable
$\dot{m}_p$	pool-side flow per train	94.6	kg/s	adjustable
$\dot{m}_c$	CCW flow per train, open	158	kg/s	adjustable
UA_p	pool-side conductance	1800	kW/K	adjustable
UA_c	CCW-side conductance	1800	kW/K	adjustable
$q_{M,\max}$	makeup line capacity	20	m^3/h	adjustable
R	makeup rangeability	50	—	adjustable
C_E	evaporation coefficient	35	—	adjustable
h_c	surface convection	3.5	$W/(m^2\,K)$	adjustable
ε	surface emissivity	0.95	—	adjustable
τ_{Tm}	pool RTD filter	0.05	h	adjustable
τ_{Lm}	level transmitter filter	0.10	h	adjustable
$R_{f,\max}$	end-of-season fouling	6.0×10^{-4}	K/kW	hidden
τ_{hx}	exchanger thermal lag	0.08	h	hidden
τ_v	CCW valve lag	0.03	h	hidden
θ	return transport delay	0.06	h	hidden
τ_{foul}	fouling development	10	h	hidden
τ_{cw}	CCW supply lag	0.5	h	hidden
τ_{vm}	makeup valve lag	0.05	h	hidden

9.3.7 Assumptions

Two balances carry all of the detail and everything else in the pool is thin, and each of the simplifications below has a place where it gives out.

One well-mixed bulk temperature. A real pool is stratified, warmer over the freshest racks and cooler at the surface, and local water temperature beside a recently discharged assembly can be several degrees above the bulk. That distribution is exactly what a pool thermal-hydraulic analysis is built to compute [9, 15], and it is the reason real pools carry more than one temperature element. The assumption fails hardest immediately after a discharge and during natural circulation with the pumps stopped.

Decay heat is a curve, not a calculation. Equation (9.7) is a least-squares fit with a worst residual near 5 % between one week and twenty years, and it carries no dependence on enrichment, burnup history or actinide content. A licensing calculation uses a depletion code validated against measured decay heat [10].

Evaporation has no forced-convection term, and is extrapolated one degree past its validation range. Equation (9.9) is the natural-convection branch, and the ventilation multiplier is a blunt stand in for air movement that on some refueling floors doubles the rate. The full model carries a velocity term that this application omits, and it is that term which extends the correlation to moving

air [12]. The surface temperature is capped at 95 °C, one degree above the 94 °C top of the validated water-temperature range, and the 1864 kg/h that the capped correlation returns there sets the 1173 kW of latent absorption that fixes both clocks during a loss of cooling. That single evaluation therefore carries more weight than any other in Section 9.6, it sits outside the range the correlation was tested over, and the correlation's own mean absolute deviation of 13.9 % is the smallest uncertainty that can honestly be attached to it.

Boiling is a clamp. Equation (9.6) pins the pool at 100 °C and sends the surplus out as steam, with no building pressure rise, no venting resistance, no condensation on the structure and no local boiling on hot assemblies before the bulk saturates. All of those matter to an accident analysis and none of them changes the control lesson.

One exchanger per train. There is no tube-side maldistribution, no internal pinch, and no distinction between fouling on the pool side and the component cooling water side, and R_f develops as a first-order lag rather than from a growth model.

Dead time is one lumped constant. The delay applies to the CCW valve command and covers piping transport only, with the valve actuator and the exchanger inventory represented as the separate lags τ_v and τ_{hx}. A real transport delay falls as pool-side flow rises, and a step test cannot separate the three contributions anyway, which is the subject of Section 9.4.3.

Nothing below the top of the racks exists. When the surface reaches $z_r + 0.05\,\mathrm{m}$ the level is clamped and the run is flagged, because cladding temperature, zirconium oxidation, hydrogen generation and the dose rate on the refueling floor are not represented in any form.

No chemistry. Boron concentration, water chemistry, criticality and rack neutronics are absent, and makeup is assumed borated to the pool concentration. Dilution is a real and separate safety concern this application says nothing about.

The numerics are more repeatable than a control system. Hold the step and the scan and every machine produces the identical trajectory, which is not a property any distributed control system has.

9.4 Dynamic Analysis

9.4.1 Two Plants, One Measurement

Linearizing (9.2) and (9.15) about any operating point, as Section 1.4.3 linearizes a tank, gives the level plant of Equation (9.17), because no loss term contributes a derivative with respect to M.

$$\frac{Z(s)}{U_m(s)} = \frac{K_L\,\mathrm{e}^{-\theta_m s}}{s}, \qquad K_L = \frac{1}{\rho A}\frac{\partial \dot m_M}{\partial x_m} \tag{9.17}$$

where $K_L = 1.622 \times 10^{-4}\,\mathrm{m/h}$ of level per percent of makeup valve at the design position and $\theta_m = 0.05\,\mathrm{h}$ is the makeup valve lag, a first-order lag in the model that the integrating form approximates as a dead time. Because the steady-state gain of (9.17) is infinite, a step response has no plateau to measure a gain against, and the pool does not regulate itself at any elevation whatever. Linearizing (9.1) instead gives Equation (9.18).

$$\frac{T(s)}{U_c(s)} = \frac{K\,\mathrm{e}^{-\theta_{\mathrm{eff}} s}}{\tau s + 1}, \qquad \tau = \frac{M c_p}{3600\,k_{self}}, \qquad K = K_I \tau \tag{9.18}$$

where $k_{self} = nk + k_{surf}$ is the total negative feedback in kW/K and K_I is the short-horizon integrating gain in °C/h per percent of valve. The whole of the difficulty is in k_{self}. With two trains at the design point $nk = 423.0\,\mathrm{kW/K}$ and the surface contributes $k_{surf} = 4.20\,\mathrm{kW/K}$, so $\tau = 3.329\,\mathrm{h}$ and the pool is an easy self-regulating loop. With both trains stopped the same expression gives $\tau = 339\,\mathrm{h}$ against decay heat of 2148 kW, and a time constant a hundred times longer than any experiment anybody will run is an integrator, the limit of a long lag that Section 1.4.4 describes. The pool becomes an integrator exactly when it begins to matter. The tank of Chapter 2 turns into an integrator when its outlet valve closes, and this pool's temperature turns into one when its pumps stop.

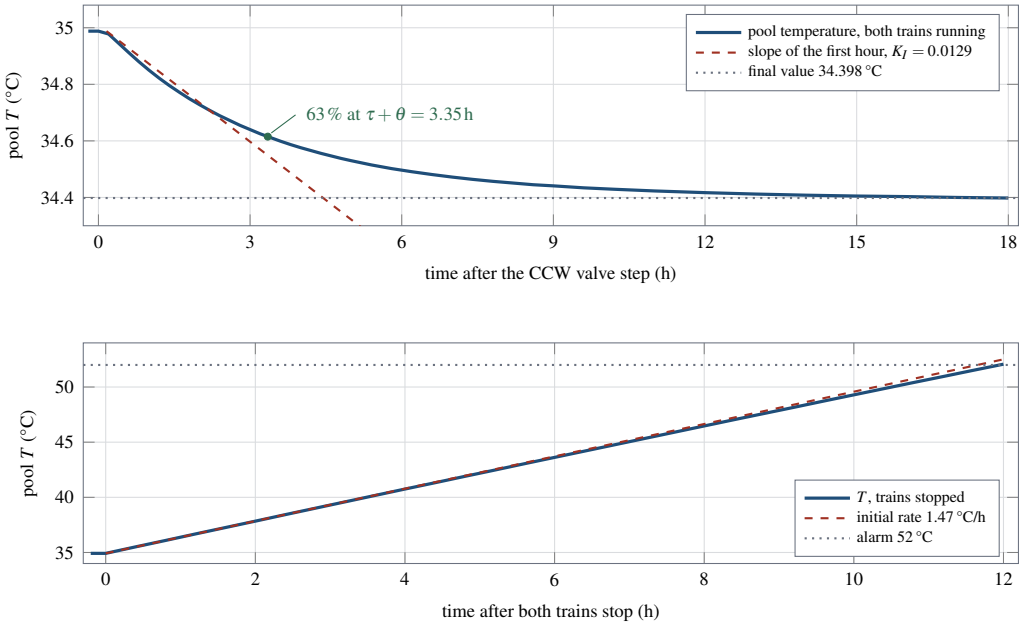


Figure 9.2 The same pool identified as two different plants. Above, a 10.70 % open step on the CCW valve with both trains running settles 0.590 °C lower, giving $K = -0.0551$ °C per percent, $\tau + \theta = 3.35$ h and $\theta = 0.16$ h. Reading only the first hour gives the integrating gain 0.0129 °C/h per percent shown as the dashed tangent, which is a correct measurement of the wrong model. Below, both trains stop 40 h into the same scenario and the pool ramps at 1.47 °C/h initially, averaging 1.43 °C/h over the twelve hours to 52.07 °C because evaporation grows on the way up.

9.4.2 Identification From One Step Test

The identification experiment steps the CCW valve from 44.3 % to 55.0 % with both trains running, waits, and then stops both trains, so that one scenario produces both models. The measured responses are shown in Figure 9.2.

With the trains running the step settles 0.590 °C lower, which gives $K = -0.0551$ °C per percent of valve. The response reaches 63 % of that change 3.35 h after the step and first departs from the pre-step value by 0.005 °C at 0.160 h, so $\theta_{\text{eff}} = 0.16$ h and $\tau = 3.19$ h, an easy loop with $\tau/\theta = 19.9$. Both times are read at the 0.01 h integration step, and the 0.1 h historian places them at its 0.2 and 3.4 h samples. The step test recovers 3.19 h where (9.18) at the design point gives 3.329 h, because the step traverses valve positions at which k is larger and the average gain over the traverse, 0.0551 °C per percent, is smaller than the incremental gain at 44.2 %, which is 0.0674 °C per percent. That difference is the nonlinearity of (9.12) and it is why the identified plant depends on where it is identified.

One variable does not hold still during the experiment. Decay heat falls from 2143 to 2129 kW over the eighteen hours of the test, and a second run in which the valve is returned to 44.3 % on the scan after the scripted step shows the pool cooling 0.032 °C on that account alone. Differencing the two runs gives 0.558 °C of step response and $K = -0.0522$ °C per percent, so the raw reading overstates the gain by 5.7 %. A student who wants the plant rather than the load has to run the unstepped case as well, which is the ordinary price of identifying anything on a plant with a decaying forcing function.

The instructive failure is to stop the test early. Read over the first hour alone, the same response has a slope of –0.1377 °C/h and gives $K_I = 0.0129$ °C/h per percent, a perfectly good integrating gain

and a perfectly good number for a rule that assumes the pool never settles. Every self-regulating plant looks like an integrator over a small fraction of its time constant, and one hour is 31 % of 3.19 h. The response reaches 98 % of its change only 13.65 h after the step, so a test long enough to distinguish the two models costs most of a day. Section 9.5 measures what the distinction is worth.

Stopping both trains produces the lower panel of Figure 9.2. The pool ramps at 1.466 °C/h at the instant of the stop and averages 1.429 °C/h over the twelve hours it takes to reach 52.07 °C, the difference being evaporation growing from 39 to 178 kg/h on the way. The stop occurs 40 h into the scenario, by which point decay heat has fallen to 2117 kW, and the same stop at the untouched design point of 2148 kW gives 1.487 °C/h. With a full core eight days out of the reactor the design-point rate is 4.710 °C/h. There is no gain, no time constant and no final value in any of these cases.

> **TRY IT IN THE APP** **3 · Step Test: One Pool, Two Plants (MANUAL)**
>
> Load the case and press Run. Watch the pool temperature trend after the CCW valve step: it bends over and settles 0.590 °C lower, a plant that regulates itself. Then, 40 h in, both trains stop. The same trend turns into a straight ramp at 1.47 °C/h, an integrator with no final value.

9.4.3 Why the Apparent Dead Time Exceeds the Transport Delay

The plant of (9.18) is not first order plus dead time. It carries a pure transport delay of 0.06 h on the valve command, a valve actuator lag of 0.03 h and the exchanger and train inventory as a further lag of 0.08 h, and a step test cannot separate the three, because a small lag in series with a dominant one displaces the response in the same way a pure delay does. The sum is approximately the quantity an identification recovers, as shown in Equation (9.19).

$$\theta_{\text{eff}} = \theta + \tau_v + \tau_{hx} = 0.06 + 0.03 + 0.08 = 0.17\,\text{h} \tag{9.19}$$

The value read from Figure 9.2 is 0.160 h, one integration step short of (9.19), and it depends on the threshold, 0.13 h for a departure of 0.002 °C and 0.27 h for 0.02 °C, because a chain of lags has no sharp onset to read. Section 1.4.6 makes the same point on a heater: an effective dead time depends on the method that reads it. The consequence for derivative action is the opposite of the one an integrating level loop reaches. On a pure integrator the controlled variable is already a ramp and derivative action amplifies measurement noise without adding anticipation, but here 0.08 h of the 0.17 h is a real first-order lag, so derivative time up to 0.6 h is offered as an adjustable and is worth something on a quiet measurement.

9.5 Controller Design

9.5.1 Four Architectures

The application implements the four structures in the order a plant grows them. Manual mode gives the operator both valve stations and nothing holds anything. On/off train control starts a train when the measured temperature reaches $T_{\text{SP}} + \Delta/2$ and stops it at $T_{\text{SP}} - \Delta/2$, running the CCW valve wide open while the train runs and shut while it does not, which is how most pools are actually operated and which costs nothing and needs no tuning. Proportional-integral control modulates the CCW valve on measured temperature and the makeup valve on measured shielding. The fourth adds an evaporation feedforward to the makeup valve. Both manual valve stations remain live whenever their own loop is not in automatic, which is not the same condition as the global mode, so selecting the manual architecture in automatic leaves both sliders active.

9.5.2 PID Algorithm as Implemented

One incremental, or velocity, algorithm serves both loops, differentiating the measurement rather than the error, as set out in Equations (9.20) and (9.21). Each scan forms the proportional and derivative

increments and accumulates them in a running sum.

$$
\begin{aligned}
e_k &= a\,(x_{\mathrm{SP}} - \hat{x}_k)\,, \qquad P_k = K_c e_k \\
D_k &= -a\,K_c \tau_D \,\frac{\hat{x}^f_k - \hat{x}^f_{k-1}}{\Delta t_c} \\
s_k &= s_{k-1} + (P_k - P_{k-1}) + (D_k - D_{k-1}) \\
c_k &= u_{bias} + s_k + I_{k-1}, \qquad u_k = \mathrm{clamp}(c_k,\, 0,\, 100\,\%)
\end{aligned}
\tag{9.20}
$$

$$
I_k = I_{k-1} + \left[\frac{K_c}{\tau_I} e_k + \frac{u_k - c_k}{T_t}\right] \Delta t_c \tag{9.21}
$$

These are Equations (1.24) to (1.27) of Section 1.6.2 in this chapter's symbols, with $s_k = S_k$, $c_k = v_k$, $u_{bias} = u_b$ and the measurement $\hat{x}$ as the PV. The derivative filter has the time constant $\max(\tau_{Tm}, 0.02\,\mathrm{h}) = 0.05\,\mathrm{h}$, the tracking time T_t defaults to τ_I, and $\Delta t_c = 0.01\,\mathrm{h}$. The temperature loop uses $a = -1$ and is therefore direct acting, so a pool above setpoint opens the CCW valve. The level loop uses $a = +1$ and is reverse acting, so a falling level opens the makeup valve. Anti-windup is back-calculation through the second term of (9.21) [18], and transfer between manual and automatic is bumpless in both directions, both as Section 1.6.2 describes. Setting $\tau_I = \infty$ disables integral action, and the top detent of the level integral-time slider does exactly that, because proportional-only averaging control on a pool this size is an ordinary deliberate choice. With no integral there is nothing to unwind, so the back-calculation is skipped and a proportional-only output is simply clipped. Both outputs are clamped to 0–100 % and neither carries a rate limit, the actuator dynamics being the lags $\tau_v = 0.03\,\mathrm{h}$ and $\tau_{vm} = 0.05\,\mathrm{h}$ of Table 9.2. The CCW valve command is delayed by θ before it reaches the plant rather than delaying the flow, which is equivalent and keeps the delay line out of a right-hand side the integrator evaluates four times per step. Execution is every 0.01 h, against $\theta_{\mathrm{eff}} \approx 0.17\,\mathrm{h}$ of effective dead time.

9.5.3 Which Tuning Rule the Identification Commits To

Three rules are offered and all three read only the numbers the student identified. Equation (9.22) is the SIMC setting for an integrator with dead time and Equation (9.23) the setting for a first-order plant with dead time [5].

$$
K_c = \frac{1}{K_I(\tau_c + \theta)}, \qquad \tau_I = 4(\tau_c + \theta) \tag{9.22}
$$

$$
K_c = \frac{\tau}{K(\tau_c + \theta)}, \qquad \tau_I = \min(\tau,\, 4(\tau_c + \theta)) \tag{9.23}
$$

Which numbers go into these rules matters, and the application's own tuning buttons use the model's incremental gains at the present operating point rather than a student's fit. At the design point those are $K = -0.0674\,^{\circ}\mathrm{C}$ per percent, $\tau = 3.329\,\mathrm{h}$ and $K_I = K/\tau = 0.0202\,^{\circ}\mathrm{C/h}$ per percent, and at $\tau_c = 2\,\mathrm{h}$ Equation (9.23) returns $K_c = 22.77\,\%/^{\circ}\mathrm{C}$ and $\tau_I = 3.33\,\mathrm{h}$, which the application ships rounded to 23 and 3.3 h. On those same gains Equation (9.22) returns the identical gain and $\tau_I = 8.68\,\mathrm{h}$, because $K = K_I\tau$ makes the two expressions for K_c algebraically the same and only the integral time differs, the integrating rule having no plant time constant to cap it with. The whole penalty of the shorter step test is therefore an integral time 2.6 times too long, and Section 9.6 measures what that costs on a setpoint move.

The student's own fit is not the same pair of numbers, and it does not carry that identity. Feeding the measured $K = -0.0551\,^{\circ}\mathrm{C}$ per percent, $\tau = 3.19\,\mathrm{h}$ and $\theta = 0.16\,\mathrm{h}$ into (9.23) gives $K_c = 26.8\,\%/^{\circ}\mathrm{C}$ and $\tau_I = 3.19\,\mathrm{h}$, while the measured $K_I = 0.0129\,^{\circ}\mathrm{C/h}$ per percent in (9.22) gives $K_c = 36.0\,\%/^{\circ}\mathrm{C}$ and $\tau_I = 8.64\,\mathrm{h}$. The gains differ by 34 % because $K_I\tau = 0.0411\,^{\circ}\mathrm{C}$ per percent against a measured K of 0.0551 °C per percent, an inconsistency the step test cannot see and the identity above conceals.

The classical reaction-curve settings for the same integrating model follow from $K_u = \pi/(2K_I\theta)$ and $P_u = 4\theta$, giving Equation (9.24).

$$K_c = \frac{0.94}{K_I\theta}, \qquad \tau_I = 2\theta, \qquad \tau_D = 0.5\theta \tag{9.24}$$

On the model's K_I and $\theta = 0.17\,\mathrm{h}$, Equation (9.24) returns $K_c = 273.2\,\%/^\circ\mathrm{C}$, $\tau_I = 0.34\,\mathrm{h}$ and $\tau_D = 0.085\,\mathrm{h}$, and on the measured pair it returns 456.5 %/°C, 0.32 h and 0.08 h. The two gains are twelve and seventeen times the SIMC gain from the same identification respectively, because the rule places the closed-loop time constant at the dead time instead of at two hours. Neither can be entered, because the temperature gain is bounded at 200 %/°C, which is itself a statement about how aggressive the rule is on this plant.

The level loop is tuned by the averaging rule instead, stated the way an operator states it. For an integrator under proportional-integral control the closed loop is exactly second order with damping ratio $\zeta = \frac{1}{2}\sqrt{K_c K_L \tau_I}$, so critical damping fixes the integral time once the gain is chosen, as shown in Equation (9.25).

$$K_c = \frac{\Delta u}{\Delta h}, \qquad \tau_I = \frac{4}{K_c K_L} \tag{9.25}$$

Choosing $\Delta u = 26\,\%$ of valve travel for $\Delta h = 0.10\,\mathrm{m}$ of level gives $K_c = 260\,\%/\mathrm{m}$ and $\tau_I = 94.8\,\mathrm{h}$, against the shipped 96 h on a slider whose detents are 4 h apart, which is a damping ratio of 1.006 rather than exactly critical. Equation (9.25) is (9.22) with $\tau_c + \theta = \Delta h/(\Delta u K_L)$, so stating how much deviation a disturbance may cause is the same act as choosing a closed-loop time constant. Four days of integral time is a design decision and not a lapse, because the nearest level constraint is 0.60 m away and costs nothing until it is reached, while makeup water is metered, treated and borated. The argument reverses the moment the disturbance is a leak rather than an evaporation change, and the graded challenge is built on exactly that reversal.

9.5.4 The Evaporation Feedforward

The feedforward computes the makeup flow that closes the mass balance now, from quantities a plant measures, and positions the valve there through the inverse of (9.15), as shown in Equation (9.26).

$$\begin{aligned} q_{FF} &= \frac{k_{FF}\,\dot{m}_E(\hat{T}, T_a, \mathrm{RH})}{\rho_M} \\ u_{bias} &= \frac{100}{\ln R}\ln\left[1 + (R-1)\,\frac{q_{FF}}{q_{M,\max}}\right] \end{aligned} \tag{9.26}$$

where k_{FF} is an adjustable trim defaulting to 1 and ρ_M is the density of the makeup water. Above 99.5 °C the correlation stops meaning anything and the branch instead returns the boil-off that the energy balance cannot put anywhere else, $\dot{m}_{boil} = (Q_{decay} - Q_{hx} - Q_{surf} - \dot{m}_E h_{fg})/h_{fg}$. Both branches are arithmetic an operator can do on paper, which is the point. Equation (9.26) enters the level loop as the bias term of (9.20), so the feedback loop is left trimming the error in the calculation and catching the leak the calculation cannot know about, which is the structure simple feedforward tuning rules assume [19]. A change of architecture moves the bias by a whole valve position, so the level integral is re-seeded against the new bias on the change, which keeps the transfer bumpless.

9.5.5 Time to the Limits

The quantity an operator needs during a loss of cooling is not an error but a time. The application reports both the marched answer and the present-rate extrapolation, as shown in Equations (9.27) and (9.28).

$$t_{boil} : \text{Euler march of (9.1) and (9.2) at } 0.25\,\mathrm{h} \tag{9.27}$$

$$\hat{t}_{boil} = \frac{T_{sat} - T}{\mathrm{d}T/\mathrm{d}t}, \qquad \hat{t}_{low} = \frac{z - z_r - 6.40\,\mathrm{m}}{-\mathrm{d}z/\mathrm{d}t} \tag{9.28}$$

Table 9.3 The Twelve Scenarios and What Each One Is For

#	Scenario	What it exercises
1	Normal operation	the scale of everything, no self-regulation on level
2	Hold it by hand for a week	exact flow matching over days
3	Step test, both plants	FOPDT and IPDT from one experiment
4	On/off train control	deadband against starts, capacity margin
5	PI from the identified model	SIMC, reset time, a load it cannot reject
6	Full core offload	actuator saturation, thermal expansion
7	A season of fouling	degradation visible only in valve travel
8	Hot, humid building	one disturbance, two very different balances
9	A liner leak	when averaging tuning becomes the wrong choice
10	Cooling lost during the outage	marched against extrapolated time to limit
11	Evaporation feedforward	computed disturbance against feedback
12	Graded challenge	one train, a fresh core and a leak

The march freezes the present cooling, makeup and leak, advances the two balances to a horizon of 120 h, and reports the first crossing of saturation and of the shielding minimum. It is refreshed every 0.50 h of simulated time, which is far more often than anybody would redo it by hand, and it costs 480 evaluations of the balance. The present-rate form of (9.28) is what an operator computes in their head and what most plant computers display, and it uses the level rate of (9.4) including the expansion term, so it returns no answer at all while the pool is warming faster than it is evaporating.

9.6 Results

Table 9.3 lists the twelve scenarios in the order a laboratory would work through them. Each one arrives with its mode, its architecture, both sets of tuning constants, every disturbance setting and every timed event already fixed, and its Learn card opens by asking for a written prediction before the run. Section 1.6.4 describes the briefing, the completion rules and the scorecard that every scenario shares. The app lists the twelfth as its Challenge, without a number. Figure 9.3 shows the tenth in progress.

9.6.1 The Scale of Normal Operation

The arithmetic of scenario 1 settles the question before a controller appears anywhere in it. The design inventory produces 2147.7 kW, two trains at 44.20 % remove 2115.2 kW of it, the surface takes 31.8 kW and the cold makeup water the remaining 0.7 kW, so the pool sits at 35.00 °C. Evaporation of 39.55 kg/h is 8.84 mm of level a day against 3 mm of transmitter noise, and the makeup valve balances it at 2.367 % of its travel. One cubic meter an hour of imbalance is 9.26 mm/h of level. Scenario 2 puts both valves in manual for a simulated week and asks for the level to be held, which cannot be done, because holding it requires matching a flow of 0.0396 m^3/h exactly and indefinitely on a valve whose smallest adjustment is 0.05 % of travel.

9.6.2 Reset Time and the Load the Loop Cannot Reject

Scenario 5 verifies the tuning of (9.23) and then finds the end of the plant. A setpoint step from 35 to 38 °C at $t = 6$ h with $K_c = 23\,\%/°\mathrm{C}$ and $\tau_I = 3.3$ h settles within ±0.15 °C in 2.72 h and rises 0.051 °C above setpoint 27.6 h after the step, as a late noise-driven excursion (2.66 to 2.95 h and 0.032 to 0.056 °C over five other noise seeds). That excursion is not a dynamic overshoot. Repeating the run with the RTD noise set to zero leaves the response monotone, settling inside ±0.15 °C in 2.76 h and inside ±0.06 °C in 3.34 h with no crossing of setpoint at all, so the 0.05 °C is the true temperature being pushed past setpoint by a controller acting on a measurement that carries 0.05 °C of noise, and not a property of the tuning. The integrating rule's $\tau_I = 8.7$ h at the same gain, run the same way with the RTD noise at zero, takes 13.9 h to settle inside ±0.15 °C, which is what the shorter step test costs.

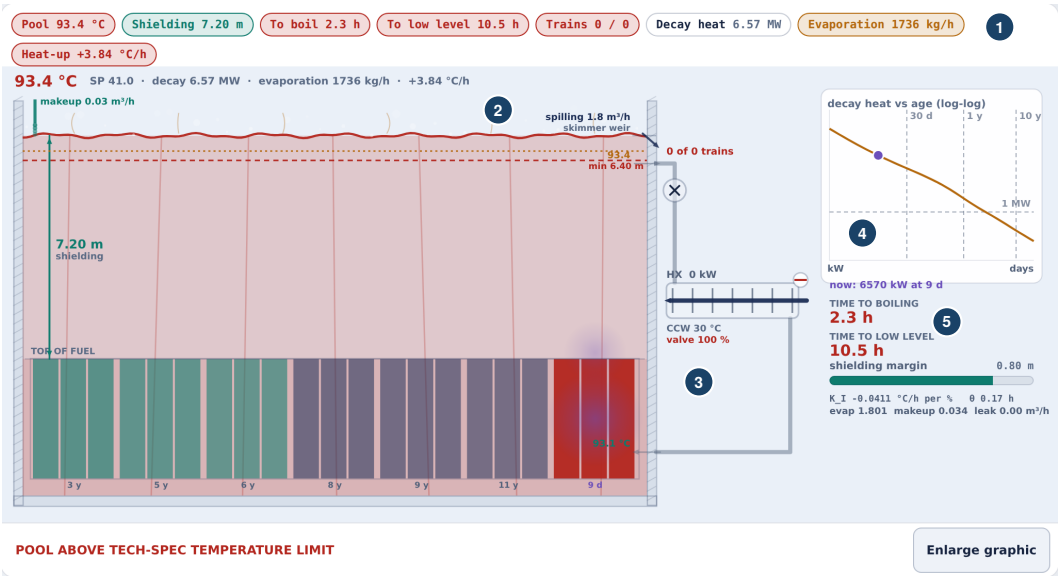


Figure 9.3 Loss of pool cooling in scenario 10 at $t = 13.00$ h, 12.00 h after both trains were lost with a full core in the pool and 1.74 h before the pool reaches saturation at $t = 14.74$ h, cropped to the plant view. (1) The badges read 93.4 °C, 7.20 m of shielding, 2.3 h to boiling, 10.5 h to the low-level limit, zero of zero trains, 6.57 MW of decay heat, 1736 kg/h of vapor and a heat-up rate of +3.84 °C/h. The vapor badge is total vapor, evaporation plus boil-off, and at this instant boil-off is still zero, so all 1736 kg/h of it is surface evaporation. (2) The surface is pinned on the skimmer weir with 1.8 m³/h going over it, and the line under the clocks reads 1.801 m³/h of evaporation and 0.034 m³/h of makeup. (3) The exchanger passes 0 kW, with the CCW valve wide open into a train that is not running. The pool stands more than thirty degrees over its 60 °C technical-specification limit while the shielding badge holds at the weir, which is the whole difficulty of the scenario: a level that has not moved at all. (4) The decay-heat inset places the freshest batch at nine days and 6570 kW on the log-log curve. (5) The two clocks disagree by design. The marched integral, last refreshed at 12.50 h and held until the next controller scan, is 2.25 h to boiling and 10.50 h to the low-level limit, shown rounded, while the straight-line extrapolation of the present rate carried in the CSV row at $t = 13.0$ h reads 1.71 h to boiling and no answer at all for the level, because the level is not falling. The status line reads POOL ABOVE TECH-SPEC TEMPERATURE LIMIT.

The component cooling water supply then rises from 30 to 38 °C and the loop answers by opening the valve until there is no valve left, settling at 41.31 °C with x_c on its stop. Two trains wide open have a duty coefficient of 607.5 kW/K, and 2148 kW across that conductance is a 3.5 °C approach to whatever the cooling water happens to be, so the entire disturbance-rejection range of this loop against T_{CCW} is about 1.5 °C. That is the plant and not the tuning, and no controller improves on it.

9.6.3 On/Off Control and the Cost of a Narrow Band

Scenario 4 runs one train on/off against the aged inventory alone. All five deadbands in Table 9.4 were measured over the same 168 h window, because a start count is meaningless without one. Figure 9.4 draws the starts and the swing side by side.

Narrowing the deadband from 3 to 1.5 °C nearly doubles the starts from 14 to 27 and nearly halves the swing from 3.19 to 1.72 °C, and buys nothing at all, because the temperature limit is more than 23 °C above the top of either band. Over five other noise seeds the narrow band gives 27 or 28 starts and every other count is unchanged, and removing the RTD noise leaves every count unchanged while narrowing each swing by 0.06 to 0.13 °C. Widening it has a hard end. One train wide open holds the

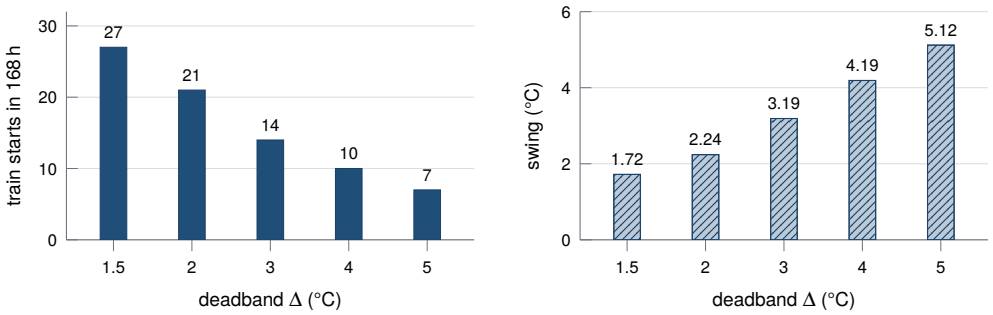


Figure 9.4 On/off train control against the aged inventory alone, from Table 9.4. Left, the train starts in 168 h (solid bars); right, the peak-to-peak swing of the true temperature (hatched bars). Notice that narrowing the band from 3 to 1.5 °C nearly doubles the starts and nearly halves the swing, while the temperature limit stays more than 23 °C above the top of either band, so the narrow band buys nothing.

Table 9.4 On/Off Train Control Against the Aged Inventory Alone, $T_{SP} = 35\,°C$ and the Shipped 0.05 °C RTD Noise, Measured Over 168 h With the Peak-to-Peak Swing of the True Temperature Read After $t = 24\,h$.

Δ (°C)	1.5	2	3	4	5
Starts	27	21	14	10	7
Swing (°C)	1.72	2.24	3.19	4.19	5.12
Mean period (h)	6.2	8.0	12.0	16.8	24.0
Duty (%)	45.9	46.1	46.3	49.5	53.0

aged inventory at 32.18 °C, so a stop temperature below that can never be reached, and any deadband wider than $2(35 - 32.18) = 5.64\,°C$ is a standing order to run. At $\Delta = 6\,°C$ the measured result is one start, 96.0 % duty and a swing of 0.15 °C, which is not on/off control at all. On/off therefore requires a capacity margin that nothing in the tuning displays.

9.6.4 Degradation That Nothing Alarms On

Scenario 7 fouls the exchanger over a season with the temperature loop in service, and the pool temperature does not move, staying within 0.05 °C of setpoint for 130 h (34.98 to 35.05 °C with the shipped RTD noise and 35.00 to 35.02 °C without it). What moves is the actuator. With the RTD noise removed so that the valve reports the plant rather than the instrument, the CCW valve walks from 42.83 % at $t = 4\,h$ to 48.83 % at $t = 130\,h$. The comparison that isolates fouling is the one the scenario asks for, a second run in which the exchanger fouling is set back to zero on the scan after the scripted event at $t = 4\,h$, and there the same valve walks from 42.83 to 40.03 % over the same window because decay heat falls from 2145 to 2055 kW. Fouling is therefore worth 8.81 % of valve travel at the same load and the same temperature. The fouled run on its own shows only 6.00 % of travel, because decay heat is closing the valve by 2.81 % at the same time, so the uncontrolled reading understates the degradation by a third. The duty coefficient measured at the operating point is not the number that reveals it, since it is 404.7 kW/K in both runs to four significant figures, both simply reporting the load the loop is carrying. The number that reveals it is the duty coefficient at a fixed valve position, which for two trains wide open is 532.4 kW/K fouled at $t = 130\,h$ against 623.5 kW/K clean, a loss of 14.6 % that reaches 15.2 % at the end-of-season resistance. When two batches are then discharged into the pool at $t = 130\,h$ both valves saturate and the fouled plant settles at 37.74 °C at $t = 180\,h$ where the clean plant gives 36.59 °C, a difference of 1.15 °C. Fouling is a margin problem rather than a temperature problem, and margin is only ever measured in actuator travel.

9.6.5 One Disturbance, Two Very Different Balances

Scenario 8 holds the pool at 42 °C with a full core twelve days out and moves the refueling floor. Decay heat is falling throughout, so every number below is read with the instrument noise removed and differenced against a companion run in which the building is set back to 30 °C and 45 % relative humidity on the scan after each scripted step and nothing else is changed. Taking the building to 40 °C and 85 % drops evaporation from 80.6 to 16.2 kg/h, and the latent duty with it from 53.8 to 10.8 kW. The CCW valve absorbs that in 0.98 % of travel, and the pool departs from the companion run by at most 0.026 °C, two hours after the step, and is back on it within a day rather than anywhere near its alarm. Taking the building to 22 °C and 20 % at $t = 72$ h raises evaporation to 108.9 kg/h and gives back 0.40 % of valve against the same reference, for at most 0.036 °C of pool temperature in the other direction. The whole factor of 6.7 in evaporation is therefore 1.38 % of valve travel, the pool temperature stays within 0.048 °C of its setpoint, and the level deviation never exceeds 10.3 mm. On the mass balance the second step is 0.099 m^3/h of unbalanced flow, which is 0.92 mm/h of level and clears 3 mm of transmitter noise in 3.3 h, so this disturbance is one the level can eventually see. The raw valve travel from the first step at $t = 8$ h to $t = 160$ h is 54.3 to 41.8 %, of which 12.2 of the 12.6 points is decay heat falling from 5761 to 4997 kW rather than the weather.

9.6.6 Saturation, and a Makeup Valve That Shuts

Scenario 6 discharges a full core eight days out of the reactor into a pool holding only the aged inventory, taking decay heat from 683 to 6680 kW, and starts the second train at the same instant. The CCW valve command reaches its 100 % stop at the scan at $t = 5.18$ h, 1.18 h after the discharge, and the temperature goes on rising for another 9.85 h, peaking at 40.68 °C at $t = 15.03$ h and falling only slowly afterward, to 40.49 °C a day after the discharge and 40.09 °C two days after. With the actuator pinned the pool is an integrator again, and an integrator does not care that its input has stopped moving. Watching the controlled variable reports the trouble about ten hours after watching the actuator would have.

The makeup valve does the opposite of what the mass balance asks. Evaporation nearly doubles from 39.5 to 71.5 kg/h, and with the instrument noise removed the level loop closes the valve to 0 % at the scan at $t = 4.98$ h and holds it there for 21.9 h, because 5.68 °C of warming on 1231 m^3 raises the surface by 18.1 mm through the second term of (9.4), and at the prevailing 0.66 mm/h of evaporation the surplus takes 29.4 h to leave, from its peak at $t = 9.81$ h to the return to setpoint at $t = 39.24$ h. With the shipped noise the valve chatters off its stop, but the level follows the same course, peaking 18.2 mm high at $t = 9.82$ h and back on setpoint at $t = 39.47$ h. A level transmitter cannot tell thermal expansion from inventory, and that is a property of every level measurement on a vessel whose temperature moves.

🕮 **Pause.** Try Exercise 9.3 before reading on.

9.6.7 A Leak Against an Averaging Loop

Scenario 9 opens a leak-chase drain at 0.05 m^3/h with the level loop deliberately loose at $K_c = 30\,\%/\mathrm{m}$ and proportional only. With the instrument noise removed the pool temperature does not move at all, staying between 34.996 and 35.000 °C over the first four days, because the whole makeup flow, 0.040 to 0.057 m^3/h of 20 °C water, is at most 1.0 kW against 2148 kW and the level is not in the energy balance in any case. On the level the deviation is 1.84 mm after four hours, 10.5 mm after a day and 33.0 mm after the eighty-eight hours before the leak grows, so it takes most of a day to clear 3 mm of transmitter noise. The mass balance is not so slow. At $t = 96$ h the metered makeup is 4.63 m^3 against an evaporation total of 3.82 m^3, and that difference plus the 3.57 m^3 the inventory has lost is within 0.02 m^3 of the 4.40 m^3 the drain has taken. When the drain then opens to 0.6 m^3/h, the same loop gives up 690 mm by $t = 400$ h and parks the pool at 6.310 m of shielding, below the 6.40 m minimum, with the makeup valve only 23.1 % open and still drifting downward at 0.36 mm/h. It is not diverging. Carried to $t = 900$ h the same run settles at 6.278 m, so the eventual offset is 722 mm

Table 9.5 Level Control Against a Swinging Evaporation Rate, Scenario 11. The Two Columns Differ Only in the Architecture. Worst Deviation Is Measured Over $6 \le t \le 240$ h With the Instrument Noise Removed.

	PI only	PI + evaporation FF
Worst level deviation (mm)	170.0	0.8
Time of the worst deviation (h)	132.0	10.7
Makeup valve travel (%)	5.36–10.46	4.52–14.19
Feedforward contribution (%)	0	4.52–14.18
Makeup used to $t = 240$ h (m^3)	38.5	46.8

and the 690 mm at 400 h is 96 % of it. An averaging design that is correct against an evaporation change is the wrong design against a leak, and no alarm distinguishes them.

9.6.8 Feedforward Against Feedback

Scenario 11 holds a hot pool at 50 °C with one train and swings the building twice. The two runs in Table 9.5 differ in the architecture and in nothing else, both starting from the same state with the same level tuning.

Evaporation goes from 94.8 to 301.8 and back to 78.9 kg/h, and the feedforward moves the makeup valve from 5.35 to 14.17 and back to 4.52 % without the level having moved at all, holding the worst deviation to 0.8 mm against 170.0 mm on feedback alone. The reason feedback loses is not the tuning, and the 170.0 mm figure needs its window stated, because the deviation is still growing when the disturbance reverses at $t = 132$ h. Holding the disturbance instead, by setting the building back to its hot, dry, ventilated state on the scan after the reversal, the proportional-only loop settles 294.2 mm low by $t = 900$ h, so the 170.0 mm observed at $t = 132$ h is 58 % of the eventual offset and the loop is neither diverging nor finished. What controls the first day is Equation (9.29).

$$\Delta z = \frac{\Delta q\, t}{A} \tag{9.29}$$

which contains no controller gain. An extra 0.21 m^3/h of evaporation over 108 m^2 is 1.9 mm/h whatever K_c is, and raising the gain shortens the tail while doing nothing to the beginning, because reading a change in evaporation off the level means waiting for its integral to clear the transmitter noise. The evaporation rate, by contrast, is computable now from a pool resistance temperature detector, a building thermometer, a hygrometer and Equation (9.9). The general criterion is that feedforward wins whenever the disturbance can be computed faster than its effect on the controlled variable can be measured, and on this pool the comparison is not close.

> **✦ DID YOU KNOW? Better Level Control Costs More Water**
>
> From the simulator: in scenario 11 the evaporation feedforward holds the worst level deviation to 0.8 mm against 170.0 mm for PI alone, yet it draws more makeup water, 46.8 m^3 against 38.5 m^3 by $t = 240$ h (Table 9.5).

9.6.9 Time to the Limits, Marched and Extrapolated

Scenario 10 is the memorable one. A full core eight days out of the reactor produces 6731 kW, both trains are lost at $t = 1$ h, and both valve stations are left in manual with the CCW valve wide open and the makeup valve at the 2.05 % the scenario ships, which passes 0.034 m^3/h against the 0.0396 m^3/h that balances the design point. The pool boils at $t = 14.74$ h and reaches the shielding minimum at $t = 22.90$ h. Figure 9.5 compares what the two calculations of (9.27) and (9.28) said along the way.

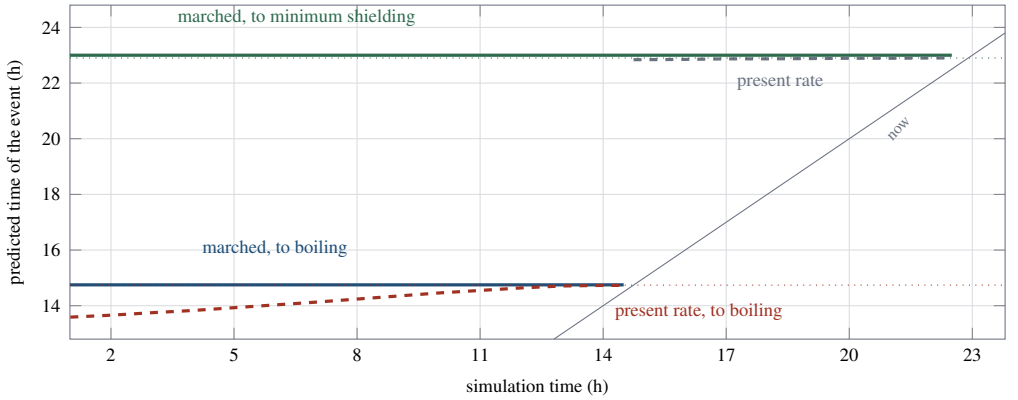


Figure 9.5 Predicted time of each limit against elapsed time during the loss of cooling of scenario 10, sampled at each half-hour refresh of the march, and hourly for the present-rate form. Solid lines are the marched predictions and dashed lines the present-rate ones, the lower pair for boiling and the upper pair for the shielding minimum. Dotted horizontals mark the outcomes, boiling at 14.74 h and the shielding minimum at 22.90 h, and the thin diagonal is the present instant. The marched prediction is available at $t = 1.00$ h, the controller scan on which the trains are lost, and is never more than 0.10 h from the outcome. The present-rate prediction of boiling starts 1.15 h early and reaches the truth only as the event arrives, and the present-rate prediction of the shielding minimum does not exist at all before $t = 14.74$ h because the level is rising until $t = 9.78$ h and pinned on the overflow weir from there until the pool boils.

> **TRY IT IN THE APP** **10 · Cooling Lost During the Outage**
>
> Load the case and press Run. Both trains are lost at 1 h. Watch the shielding badge: the level does not fall. Thermal expansion carries the surface onto the skimmer weir at 9.78 h and holds it there while the pool heats, and the pool boils at 14.74 h, from when the level is the only measurement that still says anything.

The marched calculation is available on the same controller scan the trains are lost on and it is right. At $t = 1.00$ h it reports 13.75 h to boiling and 22.00 h to the shielding minimum, which place the two events at 14.75 and 23.00 h against outcomes of 14.74 and 22.90 h, errors of 0.01 and 0.10 h on forecasts made thirteen and twenty-two hours ahead, and over the whole approach no refresh of either march is more than 0.10 h out. The present-rate extrapolation is wrong in both directions at once. On temperature it reports 12.11 h half an hour after the trip, placing boiling at 13.62 h and running 1.12 h early, because it takes no account of evaporation growing from 90 to 1864 kg/h and absorbing 1173 kW on the way up. On level it returns no answer at all, because the pool is warming at 4.68 °C/h and expanding upward at 20.4 mm/h through (9.4) while evaporation takes 0.85 mm/h away, a net rise of 19.9 mm/h, so the arithmetically correct present-rate answer is that the level is rising and will never reach the minimum.

Two consequences of (9.4) and (9.5) appear on the way. Thermal expansion carries the surface onto the skimmer weir at $t = 9.78$ h and holds it there until the pool boils at $t = 14.74$ h, spilling 10.15 m^3 of inventory at up to 2.48 m^3/h while the pool is losing water, so heating alone costs shielding. Then the pool saturates, and the vapor leaving the surface becomes 1864 kg/h of evaporation plus 8399 kg/h of boil-off, a total of 10.3 t/h, which is 99 mm/h of level against 0.40 mm/h before the trip. By $t = 30$ h the pool has boiled off 131.6 m^3. The temperature trend is flat from 14.74 h onward and the level is the only measurement that still says anything, which is the reordering of priorities that Section 9.2 describes.

9.6.10 Graded Challenge

The challenge puts a full core twelve days cooled, 5816 kW, into a pool with a fouling fraction of 0.6 and a makeup line rated at only 8 m³/h. One of the two trains trips at $t = 8$ h and returns forty hours later, and two hours after the trip a leak-chase drain opens at 2.5 m³/h. The target is the pool below 52 °C and the shielding above 6.60 m throughout, with the pool back on its 44 °C setpoint within ±0.6 °C by $t = 132$ h. Starting from the cautious default tuning the run fails all three tests, peaking at 52.03 °C at $t = 24.2$ h, falling to 5.165 m of shielding, spending 94.1 h below the minimum and ending 1.77 °C off setpoint. The suggested configuration, PI with the evaporation feedforward, $K_c = 23\,\%/°\mathrm{C}$ and $\tau_I = 3.3$ h on temperature and $K_c = 400\,\%/\mathrm{m}$ and $\tau_I = 32$ h on level, meets every part of it, with a peak of 51.18 °C, a minimum shielding of 6.886 m, no time below the minimum and a final error of 0.02 °C, at a cost of 317 m³ of makeup water against 121 m³. That leaves 0.286 m of shielding and 0.82 °C of temperature to spare. The verdicts are the same when the configuration is carried through Reset, and five other noise seeds give the same verdicts for both runs. The two bounds that must hold throughout keep their extremes from every physics step, including the starting state, so an excursion between logged rows still fails. The challenge is judged on the case as given, so changing a model parameter or a disturbance voids the attempt.

The two excursions look identical on a trend and have opposite causes. A slow ramp away from setpoint that eventually stops somewhere unhelpful describes both, and the actuator is what separates them. A CCW valve pinned at 100 %, as it is under the suggested tuning from the scan at $t = 8.91$ h until the train returns, is a plant that has run out of capacity, while a makeup valve sitting near 12 % at $t = 12$ h with the level already falling, as it does under the shipped tuning, is a controller that was never asked hard enough. The suggested configuration therefore does three different things. It tunes the temperature loop from the identified first-order model by (9.23), so the recovery after the train returns takes hours instead of days. It raises the level gain from 30 to 400 %/m and gives it a real integral time of 32 h, because this week the level loop is holding a shielding limit against a leak rather than doing averaging control against evaporation, and the right closed-loop time constant of (9.25) has changed. And it puts the evaporation feedforward in service, which does nothing about the leak but takes the evaporation term off the feedback loop, leaving all of the loop's authority for the disturbance nobody predicted. The makeup valve ends at 73.7 % of an eight-cubic-meter line, which is what balancing a 2.5 m³/h leak on equal-percentage trim requires.

9.7 Technology in Practice

The model of Section 9.3 reads one temperature and one level and moves a cooling water valve and a makeup valve. Figure 9.6 draws both loops with the devices of a real pool in place of the schematic's symbols.

What Measures It

Pools in light water reactors monitor level, temperature, liner leakage, pump discharge pressure and system flow. Most of it is shown on a local panel, only the important parameters reach the control room, and in general the instruments have no redundancy [13]. Temperature is read by thermocouples or resistance temperature detectors, mounted in contact with the water or in thermowells [20]. One pool design places them in thermal wells at the inlets and outlets of its cooling system and alarms at typically ±3 °C on the local panel, and a typical maximum pool temperature in normal operation is 60 °C [13], the technical-specification limit the chapter uses. The chapter's detector carries 0.05 °C of noise behind a 0.05 h filter, and its level transmitter 3 mm behind 0.10 h.

TECHNOLOGY Sensor: Pool Level on a Narrow Span

The commonest level measurement in nuclear plants reads differential pressure, the hydrostatic head between a tap at the bottom and one at the top [20]. Some pools use an air bubbler, a purge gas fed slowly down a dip tube whose back pressure equals the water head at its tip [13, 21]. Too much purge flow adds friction and reads high [21]. A pool level sensor typically spans only about 2 m, around the high and low alarms and the technical-specification minimum, so after a large loss of water it is off scale and useless. In the pools of the 2011 accident the level was read in the skimmer surge tank, which follows the pool only while the pool overflows into it [13].

What Moves It

A pool cooling system pumps the water through heat exchangers to an intermediate cooling system, which rejects the heat to the plant's ultimate heat sink. It takes suction through a skimmer or strainer at an elevation where a falling level makes the pumps lose suction, and its pipes enter the pool high on the wall, so that a broken pipe cannot drain the pool. Pipes that reach deep into the pool carry anti-siphon devices that let air in once the falling water uncovers them [13]. The chapter's two trains, each a pump and a plate exchanger to component cooling water, and its skimmer weir stand for this arrangement, with the valve x_c on the cooling water side as the final element of the temperature loop.

TECHNOLOGY Final Element: Makeup and Cooling Water Valves

Makeup comes from a high-purity source, usually the refueling water storage tank in a pressurized water reactor, through a connection to the suction of the pool cooling pumps, and local valve operations are needed to start it. Rates differ widely among plants, and service water or fire water can serve as alternates. The cooling water valves to the pool exchangers may be opened by hand [13]. Equal-percentage trim is built to offset the droop of a pressure drop that changes with flow, but its textbook formula never reaches zero flow and cannot be trusted at small openings [21], which is where this chapter's makeup valve balances evaporation, at 2.37 % of travel.

Where the Controller Runs

The safety systems of a pool are much simpler than the emergency core cooling systems, and manual operator actions are needed to cope with pool accidents [13]. The loops of a plant run on controllers with cycle times from milliseconds to minutes [22]. The application executes its two velocity-form PI blocks every 0.01 h, against an effective dead time near 0.17 h on temperature, and the evaporation feedforward enters the level block as its bias. The time-to-limit march is a calculation rather than a control block: it advances the two balances every 0.50 h of simulated time, beside the present-rate form that most plant computers display.

IN PRACTICE Valves, Not Pumps, Drain Pools

A 1997 regulatory study in the United States found that pools had lost more than one foot of water 10 times in 12 years, about once every 100 reactor-years, and had lost cooling with a rise of more than 11 °C 4 times. The main cause was human error, most often a mispositioned valve. Loss of cooling usually comes from an inoperable pump, and because these events are slow, operators have always restored cooling long before the pool reached its limit. The lessons drawn are reliable level, temperature and area radiation instruments, procedures and training, and checks of gates, seals and anti-siphon devices [13].

What Is Modernizing It

After the 2011 accident the regulator in the United States ordered reliable means of remotely monitoring wide-range pool level. European stress tests asked for better level, temperature and area radiation instruments, together with fixed and portable means to feed water to the pool [13]. Nuclear instrumentation and control is moving from analog to fully digital systems, apart from the transducer

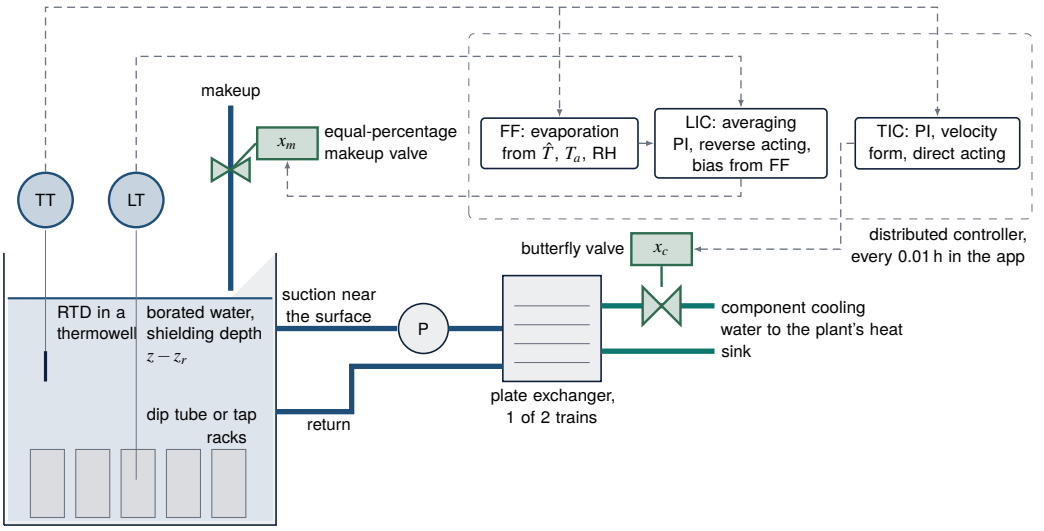


Figure 9.6 The chapter's two loops with the devices of a real pool in place of the schematic's symbols. A resistance temperature detector in a thermowell (TT) feeds the temperature controller (TIC), which moves the butterfly valve x_c on the component cooling water side of the running trains. A level transmitter on a dip tube or pressure taps (LT) feeds the level controller (LIC), whose bias is the evaporation feedforward (FF), and moves the equal-percentage makeup valve x_m. Notice that the only signal between the two loops is the measured temperature that the feedforward reads.

element, through individual upgrades constrained by licensing. Newer smart transmitters send digital serial bus signals, and fiber optics are being applied to temperature measurement [20].

9.8 Limitations

The model is not a spent-fuel pool safety analysis and cannot be made into one. With one well-mixed bulk temperature there is no stratification, no local temperature beside a fresh assembly and no natural-circulation pattern through the racks, all of which a pool thermal-hydraulic calculation exists to produce [9]. Decay heat is a fitted curve with a worst residual near 5 % rather than a depletion calculation [10]. Nothing below the top of the racks is represented in any form, so cladding temperature, oxidation, hydrogen generation and refueling floor dose rate are all absent, and the application clamps the level and raises a flag rather than integrating into nonsense. The boiling branch is a clamp with no building pressure, no venting resistance and no condensation. Evaporation carries no forced-convection term, so the ventilation multiplier stands in for air movement the correlation would otherwise compute [12], and the surface temperature it is evaluated at is capped at 95 °C, one degree past the top of the validated range, so the 1864 kg/h that fixes both clocks during a loss of cooling carries the correlation's 13.9 % mean absolute deviation at best. Chemistry is absent entirely, which puts boron dilution, criticality and rack neutronics out of reach. Disturbances step and events fire on a schedule, where a real refueling floor drifts, and a fixed step with a fixed scan repeats itself more faithfully than any plant does. Three of the scripted events are stylized on purpose. Both cooling trains disappearing at one instant is not how a plant loses cooling, but it isolates the change of plant character. A leak-chase drain that appears instantly at 2.5 m^3/h is not something a stainless liner does, but it separates a demand that exceeds the actuator from a controller that was tuned too loosely. The refueling floor moving from 40 °C and 85 % humidity to 22 °C and 20 % in one step is a caricature of weather, but it is the one building change large enough to leave a readable signature on a trend. The features that survived the simplification are the ones a controller has to be designed around, which are the pure integrator on level, the negative feedback path that can be removed, a transport delay in

series with two lags, the equal-percentage valve characteristic, the nonlinear exchanger conductance and the coupling of the two balances through evaporation. Where each individual assumption breaks is set out in Section 9.3.7.

> ⚠ **OPEN CHALLENGE** **What Happens Below the Top of the Racks**
>
> The chapter's model stops at the top of the racks, and the research does not stop there. A 2015 international status report finds the main phenomena of a pool cooling accident largely understood but names open questions: the flow and neutron behavior of partly uncovered assemblies in rack structures, hydrogen production, and the oxidation and bursting of zirconium alloy cladding in mixed air and steam. Most data come from tests made for reactor cores, and the codes were written for reactors. Progress would be dedicated pool experiments, benchmarks of the codes against them, and a ranking of the phenomena that are both important and uncertain [13].

9.9 Discussion

Pool inventory carries a direct cost and a direct constraint. Makeup water is treated and borated, and the challenge run in Section 9.6 consumes 317 m^3 of it against 121 m^3 for the run that fails, so the tuning that holds the shielding limit is also the tuning that buys the most water. That is the trade averaging level control exists to make, and it is why Table 9.5 reports valve travel and makeup volume beside the peak deviation rather than underneath it. What changes the trade is the identity of the disturbance. Against evaporation the deviation is free until the constraint is reached and a loose loop is correct, while against a leak the deviation ends at 6.310 m of shielding with the valve 23.1 % open, and the same loop is wrong. Distinguishing the two requires a mass balance rather than a level, and the arithmetic that makes the feedforward of (9.26) work is the same arithmetic that detects the leak, namely metered makeup plus reported inventory change against computed evaporation.

The monitoring calculations matter more here than the tuning does. Fouling consumed 15.2 % of the wide-open duty coefficient over a season while holding the pool at 35.00 °C, and the only evidence available at the time was 8.81 % of valve travel at the same load. No alarm exists for an actuator further open than it used to be, and on a plant whose load decays continuously nobody separates the two contributions by eye. The two calculations the exported record supports are the duty coefficient at a known valve position, which reveals exchanger condition, and metered makeup against computed evaporation, which reveals a leak within hours rather than the days a level transmitter needs. The cabin of Volume 6, Chapter 10 uses the same idea on a sealed inventory, an hourly nitrogen balance with a cumulative-sum test as its leak detector.

Methods the application does not itself implement are within reach of the model as stated in this chapter. The seven-state balance with two hard actuator limits and an explicit shielding constraint suits model predictive control, which would enforce that constraint directly and coordinate both valves as one two-input problem. Run as a moving horizon estimator, the same equations would back out the unmeasured leak and the ventilation multiplier from nothing but the two measurements, which promotes the mass-balance monitor to a soft sensor. The march of (9.27) is already an open-loop prediction over a 120 h horizon, so replacing it with an optimization that chooses the makeup profile and the train alignment to maximize the time to the shielding minimum is a small step from what is implemented.

9.10 Conclusion

This chapter studied a spent-fuel pool, a plant whose apparent dynamics depend on which equipment is running. The simulator integrates seven states, the pool temperature and mass, the exchanger return temperature, the two valve positions, the fouling resistance and the component cooling water supply temperature, with decay heat summed over a mixed-age inventory of 448 assemblies producing 2147.7 kW, evaporation from a natural-convection mass-transfer correlation and an effectiveness-NTU exchanger per train.

One step test identifies two plants. With both trains running the response settles and gives $K = -0.0551\,^{\circ}\mathrm{C}$ per percent, $\tau = 3.19\,\mathrm{h}$ and $\theta = 0.16\,\mathrm{h}$, while its first hour alone gives the integrating gain 0.0129 °C/h per percent and an integral time 2.6 times too long. With both trains stopped the same pool ramps at 1.49 °C/h on the design inventory and 4.71 °C/h with a full core in it, because the 423.0 kW/K the exchanger supplied has been replaced by 4.20 kW/K of surface loss.

Feeding the computed evaporation rate forward holds the level within 0.8 mm where the same feedback loop gives up 170.0 mm over 132 h and would eventually give up 294 mm, and the first-day deviation contains no controller gain at all. Marching the two balances forward predicts boiling and the shielding minimum within 0.10 h from thirteen hours out, where extrapolating the present rate is 1.12 h early on temperature and returns no answer on level. On/off control at a 3 °C deadband costs 14 pump starts a week, and widening the band past 5.64 °C stops the train from ever stopping.

A fuller model would add stratification as a second temperature node, a forced-convection evaporation term, boron concentration as a third inventory, and a model predictive controller that maximizes the time to the shielding minimum. As it stands, the pool is a place to identify, tune and predict an integrating process that is slow enough to be ignored.

KEY POINTS

- The pool level only adds up what flows in and out, so a makeup valve slightly off balance makes it ramp without end.
- With two trains running the pool temperature is first order with a time constant of 3.33 h. With both stopped it is 339 h, an integrator.
- Marching the two balances predicts boiling and the shielding minimum within 0.10 h from thirteen hours out, while the present rate is 1.12 h early on temperature and gives no answer on level.
- Feedforward from computed evaporation holds the level within 0.8 mm where feedback alone gives up 170.0 mm, because the first-day deviation contains no controller gain.

Where else this applies. The tank of Chapter 2 turns into an integrator when its outlet valve closes, Chapter 6 has a level that rises while the vessel loses water, and Volume 6, Chapter 10 finds a leak by mass balance.

References

[1] B. W. Bequette, “Process control practice and education: Past, present and future,” *Computers & Chemical Engineering*, vol. 128, pp. 538–556, 2019.

[2] 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.

[3] W. L. Luyben, “Liquid level control: Simplicity and complexity,” *Journal of Process Control*, vol. 86, pp. 57–64, 2020.

[4] F. Oldewurtel, A. Parisio, C. N. Jones, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, and M. Morari, “Use of model predictive control and weather forecasts for energy efficient building climate control,” *Energy and Buildings*, vol. 45, pp. 15–27, 2012.

[5] S. Skogestad, “Simple analytic rules for model reduction and PID controller tuning,” *Journal of Process Control*, vol. 13, no. 4, pp. 291–309, 2003.

[6] T.-K. Wang, S.-L. Chang, and S.-P. Teng, “Criticality safety evaluation for high-density spent-fuel storage racks,” *Nuclear Technology*, vol. 83, no. 1, pp. 5–15, 1988.

[7] *Safety and Security of Commercial Spent Nuclear Fuel Storage: Public Report*. Washington, DC: National Academies Press, 2006.

[8] R. Alvarez, J. Beyea, K. Janberg, J. Kang, E. Lyman, A. Macfarlane, G. Thompson, and F. N. von Hippel, “Reducing the hazards from stored spent power-reactor fuel in the United States,” *Science & Global Security*, vol. 11, no. 1, pp. 1–51, 2003.

[9] R. R. Gay, “Spent nuclear fuel storage pool thermal hydraulic analysis,” *Progress in Nuclear Energy*, vol. 14, no. 3, pp. 199–225, 1984.

[10] T. Yamamoto and D. Iwahashi, "Validation of decay heat calculation results of ORIGEN2.2 and CASMO5 for light water reactor fuel," *Journal of Nuclear Science and Technology*, vol. 53, no. 12, pp. 2108–2118, 2016.

[11] M. M. Shah, "Improved method for calculating evaporation from indoor water pools," *Energy and Buildings*, vol. 49, pp. 306–309, 2012.

[12] ——, "Improved model for calculation of evaporation from water pools," *Science and Technology for the Built Environment*, vol. 24, no. 10, pp. 1064–1074, 2018.

[13] OECD Nuclear Energy Agency, Committee on the Safety of Nuclear Installations, "Status report on spent fuel pools under loss-of-cooling and loss-of-coolant accident conditions: Final report," OECD Nuclear Energy Agency, Paris, Report NEA/CSNI/R(2015)2, 2015. [Online]. Available: https://www.oecd-nea.org/upload/docs/application/pdf/2021-02/csni-r2015-2.pdf

[14] *Lessons Learned from the Fukushima Nuclear Accident for Improving Safety of U.S. Nuclear Plants*. Washington, DC: National Academies Press, 2014.

[15] H. Mochizuki, "Evaluation of spent fuel pool temperature and water level during SBO," *Annals of Nuclear Energy*, vol. 109, pp. 548–556, 2017.

[16] M. M. Shah, "Calculation of evaporation from Fukushima nuclear power plant spent fuel pools," *Journal of Nuclear Engineering and Radiation Science*, vol. 5, no. 4, p. 041602, 2019.

[17] A. L. Buck, "New equations for computing vapor pressure and enhancement factor," *Journal of Applied Meteorology*, vol. 20, no. 12, pp. 1527–1532, 1981.

[18] K. J. Åström and T. Hägglund, *Advanced PID Control*. Research Triangle Park, NC: ISA – The Instrumentation, Systems, and Automation Society, 2006.

[19] J. L. Guzmán and T. Hägglund, "Simple tuning rules for feedforward compensators," *Journal of Process Control*, vol. 21, no. 1, pp. 92–102, 2011.

[20] International Atomic Energy Agency, *Core Knowledge on Instrumentation and Control Systems in Nuclear Power Plants*, ser. IAEA Nuclear Energy Series. Vienna: International Atomic Energy Agency, 2011, no. NP-T-3.12.

[21] 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.

[22] 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, Gaithersburg, MD, Special Publication 800-82r3, 2023.

9.11 Exercises

9.1 ⌕ Concept check · 2 min

Equation (9.1) is the energy balance on the pool water. Which of its terms supplies almost all of the pool's negative feedback while a cooling train runs, and disappears when both trains stop?

(a) Q_{decay}, the decay heat of the 448 assemblies in the racks
(b) $\dot{m}_E h_{fg}$, the latent heat that evaporation carries away
(c) Q_{hx}, the duty of the running trains through their exchangers
(d) $\dot{m}_M c_p (T_M - T)$, the heat taken up by the cold makeup water

9.2 ⌕ Concept check · 1 min

Table 9.1 classifies what the two loops interact with. Which of the following can a controller of the pool move?

(a) The shielding depth $z - z_r$ above the tops of the assemblies
(b) The component cooling water supply temperature T_{CCW}
(c) The liner leak $\dot{m}_L$ that drains from the pool floor
(d) The position x_m of the equal-percentage makeup valve

9.3 Concept check · 2 min

In scenario 6 a full core comes into the pool and evaporation nearly doubles, from 39.5 to 71.5 kg/h. With the instrument noise removed, the level loop closes the makeup valve to 0 % and holds it there for 21.9 h. Why?

(a) The warming water expands and lifts the surface above its setpoint, so the loop sees a high level

(b) The evaporation feedforward subtracts the extra evaporation from the bias of the makeup valve

(c) The level loop is direct acting, so a larger evaporation rate drives its output toward zero

(d) Back-calculation copies the saturated CCW valve into the integral of the level loop

9.4 Theory and calculation · 6 min

At the design point the pool holds $Mc_p = 5.120\,\mathrm{GJ/K}$. Two trains at 44.20 % give $nk = 423.0\,\mathrm{kW/K}$ and the surface gives $k_{surf} = 4.20\,\mathrm{kW/K}$.

(a) Compute the time constant of Equation (9.18) with both trains running and with both stopped. *Checkpoint:* 3.33 h and 339 h.

(b) Both trains stop at the design point. Decay heat is 2147.7 kW, the surface removes 31.8 kW and the cold makeup 0.7 kW. Compute the initial heat-up rate, and the time a straight line at that rate takes from 35 to the 60 °C limit. *Checkpoint:* 1.49 °C/h and 16.8 h.

(c) With a full core eight days out of the reactor the decay heat is 6731 kW. Repeat the initial rate with the same surface and makeup terms, and say why the real pool reaches boiling later than a straight line at that rate predicts. *Checkpoint:* 4.71 °C/h.

9.5 Theory and calculation · 6 min

The step test of scenario 3 steps the CCW valve 10.70 % open with both trains running. Read to the end it gives $K = -0.0551\,^\circ\mathrm{C}$ per percent, $\tau = 3.19\,\mathrm{h}$ and $\theta = 0.16\,\mathrm{h}$. Read over its first hour only, the pool falls at 0.1377 °C/h. Use magnitudes, since the controller action carries the sign, and $\tau_c = 2\,\mathrm{h}$.

(a) Compute the integrating gain K_I from the first hour, and the gain and integral time of the first-order rule, Equation (9.23). *Checkpoint:* $K_I = 0.0129\,^\circ\mathrm{C/h}$ per percent. $K_c = 26.8\,\%/^\circ\mathrm{C}$, $\tau_I = 3.19\,\mathrm{h}$.

(b) Compute the gain and integral time of the integrating rule, Equation (9.22), and the ratio of the two gains. Explain the ratio with $K_I\tau$. *Checkpoint:* $K_c = 36.0\,\%/^\circ\mathrm{C}$, $\tau_I = 8.64\,\mathrm{h}$, ratio 1.34.

(c) Compute the settings of Equation (9.24) on the same K_I and θ. Can they be entered in the app? *Checkpoint:* $K_c = 456.5\,\%/^\circ\mathrm{C}$, $\tau_I = 0.32\,\mathrm{h}$, $\tau_D = 0.08\,\mathrm{h}$. No.

9.6 Theory and calculation · 8 min

The level plant of Equation (9.17) has $K_L = 1.622 \times 10^{-4}\,\mathrm{m/h}$ of level per percent of makeup valve at the design position.

(a) Apply the averaging rule, Equation (9.25), for 26 % of valve travel on 0.10 m of level, and compute the damping ratio at the shipped integral time of 96 h. *Checkpoint:* 260 %/m, 94.8 h, $\zeta = 1.006$.

(b) Scenario 9 runs the level loop proportional only at 30 %/m, from a makeup valve at 2.367 % that passes the 0.0396 $\mathrm{m^3/h}$ of evaporation. A leak of 0.6 $\mathrm{m^3/h}$ opens and the pool temperature does not move. Use Equation (9.15) to find the valve position that passes the leak and the evaporation, and the level offset that holds it there. *Checkpoint:* 24.1 % and 0.72 m.

(c) Before the loop has acted, how fast does a leak of 0.05 $\mathrm{m^3/h}$ lower the level, and how many hours pass before the fall equals the 3 mm of transmitter noise? *Checkpoint:* 0.46 mm/h and 6.5 h.

9.7 App challenge · 8 min

Load scenario 3, *Step test: one pool, two plants (MANUAL)*, as shipped at the default seed. Run to 70 h at 20× and export the CSV file. Use the true pool temperature, column `pool_C`.

(a) Read `pool_C` at 6.0 and 24.0 h. Give the total change and the gain per percent of the 10.7 % CCW valve step. *Checkpoint:* −0.590 °C and −0.0551 °C per percent, within 0.003 °C and 0.0003 °C per percent.

(b) Find the first row after 6 h at which the pool has made 63.2 % of that change, and the first row at which it is 0.005 °C below its value at 6.0 h. Give both as times after the step, and estimate θ and τ. *Checkpoint:* 3.4 h and 0.2 h, so $\theta \approx 0.2$ h and $\tau \approx 3.2$ h.

(c) Both trains stop at 40 h. Compute the mean slope of `pool_C` from 40 to 52 h, and say why it is smaller than the 1.466 °C/h at the instant of the stop. *Checkpoint:* 1.429 °C/h, within 0.005 °C/h.

9.8 App challenge · 8 min

Load scenario 5, *PI from the identified model*, as shipped at the default seed, with $K_c = 23\,\%/°\text{C}$ and $\tau_I = 3.3$ h. The setpoint steps from 35 to 38 °C at 6 h and the cooling water supply rises from 30 to 38 °C at 40 h.

(a) Run to 100 h at 20× and export the CSV file. In column `pool_C`, find the first row after 6 h from which the pool stays within ±0.15 °C of 38 °C until 40 h, as a time after the step. Read the pool temperature and `ccw_pct` at 100 h. *Checkpoint:* 2.8 h within 0.1 h. 41.31 °C within 0.02 °C, valve at 100 %.

(b) Press Reset, open Advanced, set Pool RTD noise to 0 and Temperature integral time to 8.7 h, run to 40 h and export again. Make the same reading of the time to settle. *Checkpoint:* 13.9 h within 0.1 h.

9.9 App challenge · 8 min

Load scenario 11, *Evaporation feedforward, and why the level loop cannot do it*, at the default seed. Before Run, open Advanced and set Pool RTD noise and Level transmitter noise to 0. Run each case to 240 h at 20× and export the CSV file. The building turns hot and dry at 8 h and cool and damp at 132 h.

(a) On the PI architecture as loaded, read Makeup used on the scorecard, and the lowest value of `shield_m` with its time. *Checkpoint:* 38.5 m^3, and 6.830 m at 132.0 h, within 0.002 m.

(b) Press Reset, set Control architecture to PI + evap FF, and run again. Read the largest departure of `shield_m` from 7.000 m after 6 h, and Makeup used. *Checkpoint:* 0.8 mm within 0.3 mm, 46.8 m^3.

(c) From `q_evap_m3h` at 7.9 and 12 h, find the step in evaporation and the rate at which it lowers the level before any loop acts. Why can no level gain prevent this first fall? *Checkpoint:* about 0.21 m^3/h and 1.9 mm/h.

9.10 Application area · 4 min

Scenario 4 runs one train on/off at a 35 °C setpoint against the aged inventory, which one train wide open holds at 32.18 °C. Use Table 9.4.

(a) What is the widest deadband that still lets the train stop, and what happens at 6 °C? *Checkpoint:* 5.64 °C.

(b) A shift supervisor wants the band narrowed from 3 to 1.5 °C so the trend looks tidier. What does it cost and what does it buy? *Checkpoint:* 14 to 27 starts a week, swing 3.19 to 1.72 °C.

9.11 Application area · 4 min

In scenario 9 a leak-chase drain opens at 0.05 m^3/h at 8 h. At 96 h the metered makeup totals 4.63 m^3, the computed evaporation totals 3.82 m^3, and the inventory has fallen by 3.57 m^3.

(a) Estimate the volume lost to the leak and its mean rate since 8 h. *Checkpoint:* 4.38 m^3 and 0.050 m^3/h.

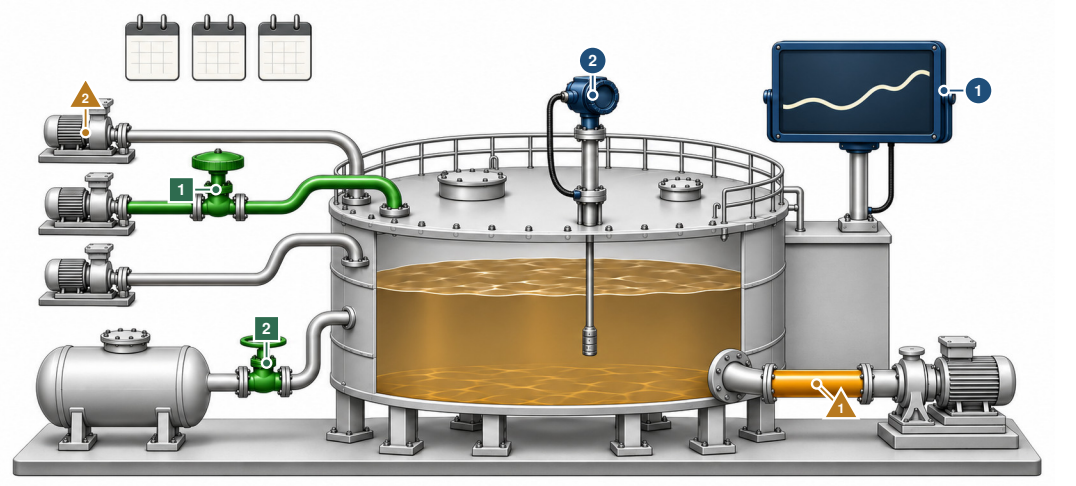
(b) The level deviation at 96 h is 33.0 mm. Which of the two, the level or the mass balance, is the better leak detector, and what does the mass balance need that the level does not? *Checkpoint:* The mass balance.

9.12 Review the agent · 5 min

An AI assistant drafts an operator aid for a loss of pool cooling: "Compute the time to boiling as $(100 - T)/(\mathrm{d}T/\mathrm{d}t)$ and the time to the shielding minimum as $(z - z_r - 6.40)/(-\mathrm{d}z/\mathrm{d}t)$ from the present trends. Both answers are conservative, because evaporation only grows as the pool heats, so the pool always reaches each limit later than the straight line says." Find the flaw, and say how you would check the claim in the app.

CHAPTER 10

Oil Price Versus Market Share in a Cartel's Quota Loop



The controller moves 1 the producer group's quota, set at its meetings 2 the consumers' reserve release

The controller reads 1 the oil price 2 the stocks, reported a month late

What disturbs it 1 a change in world demand 2 rival shale wells made cheaper by new technology

Notice that only one of the three supply lines into the tank passes the group's valve, so the rival lines are free to grow while a high price is held. Illustration, not to scale.

IN PLAIN TERMS Pumping Less Raises the Price, for a While

The main idea. A group of oil producers can lift the world price of oil within weeks by pumping less. The oil stored around the world works like one big tank: when less flows in than people burn, the tank drains, and a lower tank means dearer oil.

The challenge. A high price slowly pays rival drillers to pump more. Over years it also pays buyers to use less. So the group must keep cutting to hold its price, and in eight years it can lose about half of its share of the market.

Why it matters. A target that looks easy to hold this month can be out of reach once the slow responses catch up.

THREE WAYS TO READ THIS CHAPTER

Explore	13 min	the first page, the founding agreement (Section 10.2), the fact on the largest cut and the history of the consumers' reserve (Section 10.5), the try-it segments of cases 4 and 7 and the fact on the economic MPC (Section 10.6), and the key points
Study	50 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	15 min	the first page, the results (Section 10.6), Technology in Practice (Section 10.7), Limitations (Section 10.8) and the open challenge

The control question A cartel that pumps less lifts the oil price within weeks, and the price it then holds pays rivals to drill and consumers to buy efficiency for years. How far can a quota decided at two meetings a year hold the price before the market it protects has gone to the fringe?

Learning objectives

- Write the balance on the world's stocks with its price law, and compute the stocks for a price and the rate at which the stocks alone return to balance.
- Compute the market's short-run and long-run gains from its equations and compare them with a quota step in the app.
- Compute the long-run residual demand at a held price, and explain why no tuning holds both a high price and the cartel's share.
- Explain how the meeting calendar, the quantum and the cap shape a quota, and what a leaking actuator does to an integral loop.
- Compare a PI on price, a PI on stocks and an economic MPC on a ten-year challenge.

Prerequisites Sections 1.4.2, 1.4.3, 1.4.4, 1.6.2 and 1.6.4, and the tank balance of Chapter 2. Section 1.6.3 is needed only for the economic MPC.

The app https://apmonitor.com/apps/opec-oil-market/

See also Volume 6, Chapter 11 for stocks ordered on a calendar in a quantum with a cap, and Volume 6, Chapter 7 for the choice of which flow holds an inventory. Volume 9, Chapter 9 for a specification written as a band that a controller can spend, and Volume 9, Chapter 6 for a price set inside a predictive controller.

10.1 Introduction

A barrel that a cartel withholds is missed at once, because the world cannot stop burning oil quickly, so the price rises within weeks. The same price then pays shale drillers within a year, conventional producers within several years and consumers who buy efficiency over a decade, and the barrels they add or no longer need are the cartel's lost market. In the simulator of this chapter, one more million barrels a day (Mb/d) from the cartel lowers the price by \$8.1 within three months and by only \$1.6 seven years later. A cartel that holds \$100 through a shale revolution keeps the price there, at an average of \$98.3 over years 1 to 8, while its share of world supply falls from 30 to 15.27 %.

Oil economics has long described this market in the terms of dynamic optimization and feedback. The economics of exhaustible resources was set out in 1931 [1], a cartel dominating a competitive fringe was posed as a Nash–Cournot problem on the world oil market in 1976 [2], and producers' gains from cartelization were computed from optimal dynamic pricing on small models with lagged demand and supply [3], which is open-loop optimal control. When ten world oil models were compared in 1980 and 1981, most of them set the cartel's price by a rule driven by its capacity utilization, one treated the cartel as an intertemporal wealth maximizer, and every one made it the residual supplier [4]. The first is a feedback law and the second an economic optimizer, which economic model predictive control (MPC) now formalizes [5]. Inventories join the two sides of the market, because prices, production and stocks are set together through a market for storage [6], and inventory data identify the speculative part of the oil price [7]. The plant is a stock with four inflows and one outflow, the actuator is a quota and the controller is a rule or an optimizer, which makes the market a control problem in the ordinary sense of the word.

The belief this plant tests is that an integral controller can hold any setpoint its actuator is able to reach. A price held above the market's reference is itself an input to the plant, because it moves the slow states of the fringe and of the consumers, and those states take the cartel's barrels. In the application's own sweep of three price targets, four tunings and two calendars, every run that averages \$95 or more over years 1 to 8 loses at least 49 % of its share, so the infeasibility lies in the plant and not in the tuning. The same structure appears wherever a level is held by metering one of several inflows whose owners, or whose outflow, adapt to the level, such as a reservoir shared with other diverters, a buffer stock whose customers substitute, or a market that entrants join when the price

is high. In control terms it is a self-regulating plant (Section 1.4.4) whose long-run gain is about a ninth of its short-run gain, driven through an actuator that is sampled twice a year, quantized and rate limited. The level loop of Chapter 2 holds a tank against a fixed outflow law, and the central bank of Volume 2, Chapter 11 decides a policy rate through the same calendar module.

The work that explains the cartel's choices uses equilibrium and game models calibrated to history [8, 9], and the simulation models of the 1980s embedded a pricing rule as part of the market [4]. This chapter instead puts a chair, a PI loop, a stock target and an economic optimizer on one tank and one calendar, and measures how far integral control, a stock target and an economic optimizer can hold the oil price before the market they protect has gone to the fringe. The market structure, the PI loop, the inventory target and economic MPC are all standard. The analysis is a set of measurements on one deterministic simulator together with a linearization of the simulator's own equations. The simulator is a ten-state market whose barrel ledger closes to 7.4×10^{-13} Gb over ten years of every case, with a stock report published a month late beside its five-year average and an MPC that relinearizes the market five times at each meeting. Every equation, constant and limit is stated. Nine cases, a balanced reference and a ten-year challenge report price, share and revenue side by side.

Section 10.2 follows the domain from exhaustible-resource theory to the shale decade. Section 10.3 writes the balance on the stocks, the price law and the lagged fringe and demand, and Section 10.4 linearizes them to find the market's two speeds and its two gains. Section 10.5 sets out the meeting calendar, the two PI loops, the price war, the consumers' reserve and the economic MPC. Section 10.6 reports the measured cases and the challenge, Section 10.7 the surveys, targets and meetings of the real market, Section 10.8 the limitations and Section 10.9 the conclusions. The exercises in Section 10.10 close the chapter.

10.2 From Exhaustible Resources to the Shale Decade

The cartel's problem entered economics as a problem of time, through the economics of exhaustible resources [1]. In 1976 the cartel was modeled as a unified enterprise that dominates the other extractors because of its larger reserves, with equilibrium price and sales paths that give neither the cartel nor the competitive fringe an incentive to change its behavior. In that model a disproportionate share of the extra profit from forming the cartel went to non-members, and the cartel's restriction of sales eventually left it the sole supplier [2]. Two years later the gains from cartelizing the world petroleum market were calculated under optimal dynamic monopoly pricing, with small market models that carried short-term lags in demand and supply, by comparing the discounted profits of optimal price paths under competition and under cartelization [3]. The fringe, the lags and a discounted objective are all in the simulator, and the last becomes the objective of its economic MPC.

The forecasting models of the next decade wrote the cartel as a feedback rule. During 1980 and 1981 the Energy Modeling Forum at Stanford University compared ten prominent models of the international oil market. Most were simulation models in which the cartel set its price by a target capacity utilization rule, one let it act as an intertemporal wealth-maximizing price setter, one made its capacity endogenous while the rest held it constant at 34 million barrels a day, and all of them made it the residual supplier, producing the difference between world demand and non-OPEC supply [4]. The simulator's cartel has the same 34 Mb/d of capacity and the same residual role. Its controllers replace the capacity-utilization rule with a PI loop on the price or on the stocks, so that the rule becomes a design choice rather than a fixed part of the market.

The collapse of 1986 is the price war of the fifth case and the leaking valve of the seventh. From 1979 to 1985 the cartel's members restricted output, Saudi Arabia most of all, by more than 60 % of its 1979 level, and a shortfall in demand for the cartel's oil drove Saudi production below 3 million barrels a day in the summer of 1985 [4]. Early in 1986, after years of futile attempts to get its partners to observe their quotas and share the burden of restriction more equitably, Saudi Arabia and its allies produced their "fair share" and let the price fall. From August 1985 to mid-1986 the cartel's output

rose by about 4 million barrels a day, some 25 %, more than half of it Saudi, and in the first half of 1986 crude oil fell to about $12 a barrel, its level of 1974 [4]. A discussant of that analysis described the small members as cheating readily, knowing that the others would cover for them, while the agreement collapses if the largest member cheats [4].

Later episodes separated the causes of a high price. The run-up of 2007 and 2008 came from strong demand confronting stagnating world production, unlike earlier shocks that came from physical disruptions of supply [10]. A structural model that identifies the speculative component of the price from inventory data attributed the 2003 to 2008 surge to world consumption driven by the business cycle and not to speculation [7], and supply shocks, shocks to the demand for all industrial commodities and shocks specific to the oil market were shown to move the real price in different ways [11]. An overview of every major fluctuation from 1973 to 2014 also explains why the price remains hard to predict [12]. The simulator keeps one channel of all these, the price set by the stocks, and drives it with demand and supply shocks of declared size.

The shale decade made the fringe fast. In November 2014 OPEC announced a strategy aimed at its market share, and crude prices crashed over the following year. An equilibrium model calibrated to the market explains the switch from accommodating shale up to 2014 to defending share afterward by the growth of shale, slower demand, a less cohesive cartel and rising non-OPEC output, with cheaper shale working the other way [8]. Saudi policy in that fall was read as a trade-off between revenue now and market share later, in which uncertainty about the elasticity of shale supply made not cutting the safer choice [9]. Well-level data from North Dakota later put the supply elasticity of shale wells at 0.3 to 0.9 and found no such response in conventional wells [13]. The storage limit had its own episode in April 2020, when a collapse of demand during the pandemic took West Texas Intermediate (WTI) futures below zero [14]. The simulator's sixth case is that tank filling to its top.

> ⧖ HISTORY **A Founding Agreement That Regulates Production, 1960**
>
> Representatives of the governments of Iran, Iraq, Kuwait, Saudi Arabia and Venezuela met at Baghdad from 10 to 14 September 1960 and formed a permanent Organization of the Petroleum Exporting Countries. Their first resolution called on the members to study and formulate a system to ensure the stabilization of prices by, among other means, the regulation of production, with due regard to the interests of the producing and of the consuming nations. The organization was to hold meetings at least twice a year, and more often if necessary [15]. A loop that reads a price, moves production and decides at two meetings a year is the structure this chapter simulates, and twice a year is its shipped calendar.

10.3 One Tank for the World Market

The simulated market is one aggregate stock of commercial crude oil, drawn in Figure 10.1 as the application draws it, with the producers' inflows at the left, the world's demand leaving at the right and the price read off the level. Time is in years, flows in million barrels a day (Mb/d), stocks in billion barrels (Gb), prices in dollars a barrel and revenue in billions of dollars. A flow of 1 Mb/d held for a year is $K = 0.365$ Gb. Table 10.1 classifies the signals.

10.3.1 Stocks, Price and the Quota Valve

The stocks integrate the imbalance between supply and demand, as every inventory balance of Section 1.4.2 does, and the price is a steep, falling function of the stocks, as shown in Equations (10.1) and (10.2).

$$\dot{I} = K(S - D), \qquad S = q_c + q_{sh} + q_{cv} + q_o + r \tag{10.1}$$

$$P = P_0 \exp\left[\Pi_{[-8,5]}\left(-k(I - I_0)/I_0\right)\right](1 + d_r) \tag{10.2}$$

where $P_0 = \$75$ is the price at the normal stock $I_0 = 2.9$ Gb, $k = 9$, d_r is a risk premium and Π clips the exponent far outside the operating range. A 10 % build multiplies the price by $e^{-0.9} = 0.407$, so $100 sits at 2.807 Gb and $50 at 3.031 Gb. The shape follows the role that inventories play in the oil

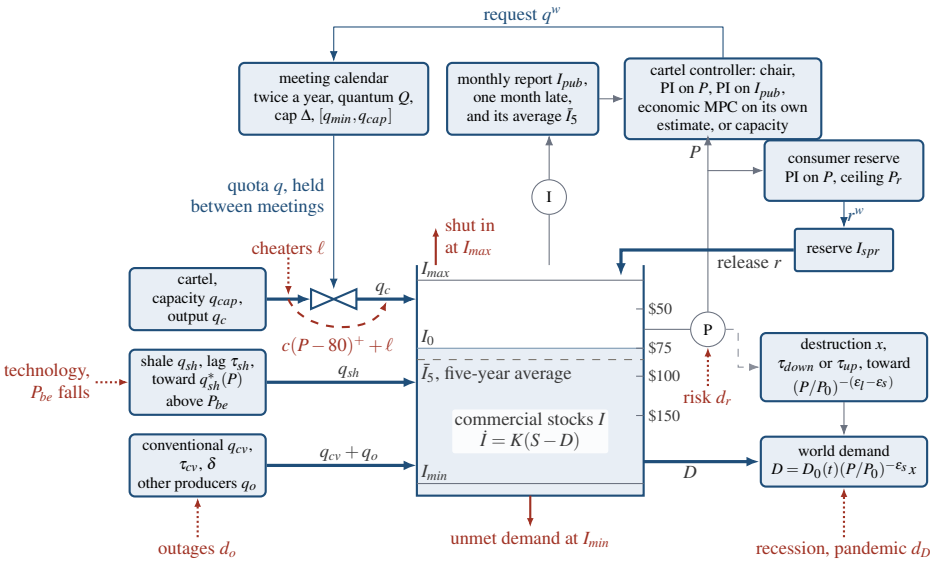


Figure 10.1 The world oil market as one tank, not to scale. Commercial stocks I integrate the cartel's output q_c, shale q_{sh}, conventional supply q_{cv}, other producers q_o and a consumers' release r, less world demand D. The ruler on the wall is the price law of Equation (10.2), cheapest when the tank is fullest, and oil above I_{max} is shut in. The quota q is a valve with a dashed bypass, the overproduction $c(P-80)^+$ and cheaters ℓ. Shale moves toward a price-dependent target above its breakeven P_{be}, and demand carries a destruction state x that falls faster than it recovers. The cartel's controller reads the price, or the monthly report I_{pub} and its five-year average $\bar{I}_5$, and its request q^w passes the meeting calendar's quantum Q, cap Δ and bounds. The consumers' reserve I_{spr} is a second controller on the same level. Densely dotted arrows (red on screen) are disturbances, solid red arrows are the oil shut in or not supplied, and the gray dashed line marks the price reaching the slow responses.

Table 10.1 Signals of the Market by Role

Symbol	Role	Unit
P	CV, oil price, SP P_{sp}	\$/bbl
I_{pub}	CV of the stocks loop, SP $\bar{I}_5 + \Delta I_{sp}$	Gb
q	MV, quota in force, q_{min} to q_{cap}	Mb/d
q^w	request of the chair or a controller	Mb/d
r	MV of the consumers' reserve, 0 to r_{max}	Mb/d
d_D, d_o, d_r	DVs, demand, other supply, risk premium	Mb/d, —
P_{be}, ℓ	DVs, shale breakeven, cheaters' output	\$/bbl, Mb/d
I, I_{spr}	states, commercial stocks and the reserve	Gb
q_{sh}, q_{cv}	states, shale and conventional output	Mb/d
x	state, demand factor, 1 less destruction	—
V	state, cartel's discounted revenue	\$ billion
q_c, S, D	outputs, cartel output, supply, demand	Mb/d
ϕ	output, cartel share of supply	%
$\bar{I}_5$	published average of 60 reports	Gb

market [6, 7], but the coefficients are a teaching closure and not an estimate. The cartel's output is the quota plus what the members add on their own, bounded by capacity, as shown in Equation (10.3).

$$q_c = \Pi_{[0,q_{cap}]}\left(q + c\,(P - P_c)^+ + \ell\right) \tag{10.3}$$

where $q_{cap} = 34\,\text{Mb/d}$, the members overproduce $c = 0.03\,\text{Mb/d}$ for each dollar above $P_c = \$80$, and every scripted cheater adds 1 Mb/d to ℓ. The quota therefore commands the valve, but not the whole flow through it. At \$100 the members add 0.6 Mb/d beyond any quota, enough to raise the CHEATING status, which reads more than 0.25 Mb/d above the quota, and in the fourth case the cartel produces 14.18 Mb/d against a quota of 13.6 at year 8.

10.3.2 A Fringe With Lags

Shale and conventional supply answer the price through first-order lags (Section 1.4.4), as shown in Equations (10.4) and (10.5).

$$\tau_{sh}\dot{q}_{sh} = \Pi_{[0,\,q_{max}]}\big(q_b + s\,(P - P_{be})\big) - q_{sh} \tag{10.4}$$

$$\dot{q}_{cv} = \frac{(1+\tau_{cv}\delta)\,q_{cv0} + s_{cv}(P-P_0) - q_{cv}}{\tau_{cv}} - \delta\,q_{cv} \tag{10.5}$$

where the breakeven P_{be} is the price at which new wells just replace the decline of old ones, so that shale holds at its base q_b, each dollar above it adds $s = 0.1\,\text{Mb/d}$ in the long run up to $q_{max} = 15\,\text{Mb/d}$, and $\tau_{sh} = 0.75\,\text{yr}$. Conventional supply has an investment response $s_{cv} = 0.2\,\text{Mb/d}$ per dollar, a lag $\tau_{cv} = 4\,\text{yr}$ and a decline $\delta = 3\,\%$ a year, so its steady state is $q_{cv0} + s_{cv}(P-P_0)/(1+\tau_{cv}\delta)$, 0.179 Mb/d for each dollar. Other producers supply a fixed $q_o = 10\,\text{Mb/d}$ plus any outage d_o. At \$75 shale gives 10 Mb/d, from $q_b = 7\,\text{Mb/d}$ and $P_{be} = \$45$ by default or from $q_b = 10.5\,\text{Mb/d}$ and $P_{be} = \$80$ in the shale-revolution cases, in which a technology shift lowers P_{be} by \$35 over four years. Read as an elasticity, s is 0.75 at \$75 and 10 Mb/d, inside the range estimated for shale wells [13], although that estimate is for individual wells.

10.3.3 Demand That Is Destroyed and Slowly Restored

Demand answers the price at once through the short-run elasticity and over years through a destruction state, as shown in Equations (10.6) and (10.7).

$$D = D_0(t)\,(P/P_0)^{-\varepsilon_s}\,x \tag{10.6}$$

$$\tau\dot{x} = x^*(P) - x, \qquad \tau = \begin{cases} \tau_{down} & x^* < x \\ \tau_{up} & x^* \geq x \end{cases}$$

$$x^*(P) = \big(\max(1,P)/P_0\big)^{-(\varepsilon_l - \varepsilon_s)} \tag{10.7}$$

where $D_0(t) = [D_0 \mathrm{e}^{gt}\,\sigma(t) + d_D(t)]^+$ is the demand at the reference price before any destruction, $D_0 = 100\,\text{Mb/d}$, the growth g is zero by default, $\sigma(t)$ is one unless a carbon policy turns the growth to $-1\,\%$ a year, d_D collects the scripted demand shocks, $\varepsilon_s = 0.06$, $\varepsilon_l = 0.35$, $\tau_{down} = 6\,\text{yr}$ and $\tau_{up} = 12\,\text{yr}$. On the model's own equation, a price held at \$100 cuts demand by 1.71 % at once and moves x 63 % of the way to its target of 0.920 in 6.0 yr, half in 4.16 yr, while a price then held at \$50 takes 8.32 yr to move it half of the way to its target of 1.125. Efficiency bought at a high price is therefore sold back at a low one only half as fast, the asymmetry that OECD demand shows between price increases and decreases [16].

10.3.4 Revenue, Share and the Monthly Report

The cartel's objective, its share and the consumers' reserve close the state equations, as shown in Equation (10.8).

$$\dot{V} = KPq_c\,\mathrm{e}^{-\rho t}, \qquad \phi = \frac{q_c}{\max(1,\,S-r)}, \qquad \dot{I}_{spr} = -Kr \tag{10.8}$$

where V is the discounted revenue at $\rho = 5\,\%$ a year in billions of dollars, the share leaves the reserve's release out of the denominator, and the reserve starts with 0.7 Gb. Four running totals, oil

Table 10.2 Scripted Shocks, With ς the Years Since the Event

Event	Effect
recession	$d_D = -3\varsigma$ for $\varsigma < 1$, then $-3(3-\varsigma)/2$ to $\varsigma = 3$
pandemic	$d_D = -20$ for $\varsigma < 0.25$, then $-20(1.75-\varsigma)/1.5$ to $\varsigma = 1.75$
outage	$d_o = -1.5$ for a year, or -5.7 for two weeks
technology	P_{be} falls by $35\min(\varsigma/4, 1)$
cheater	ℓ rises by 1, for good
carbon policy	$\sigma = \mathrm{e}^{(-0.01-g)\varsigma}$
risk spike	d_r rises by 0.2 for three months
reserve release	r rises by 1 for six months

Flows in Mb/d, prices in $/bbl.

supplied, oil demanded, oil shut in and demand unmet, keep the barrel balance auditable. At every month end the stocks are recorded, and the report of month m is published at the end of month $m + L_r$, with $L_r = 1$, so that

$$I_{\text{pub}} = I(t_m), \qquad \bar{I}_5 = \frac{1}{60}\sum_{i=0}^{59} I_{\text{pub},m-i} \tag{10.9}$$

where t_m is the end of the latest month reported and the average spans the last five years of reports. The price, by contrast, is read without delay.

10.3.5 Shocks, Initial State and Numerics

The scripted shocks of Table 10.2 are pure functions of the time since the event fired, so the derivatives stay pure, and each event takes effect on the earliest step that reaches its time. Three advanced sliders add constant shifts to demand (−20 Mb/d to 10), to other supply (−5 Mb/d to 5) and to the risk premium (0 to 0.5). The model draws no random number, so every run is deterministic.

Every case starts balanced at $75 unless it declares otherwise. Shale starts at its target, conventional supply at q_{cv0}, $x = 1$, and the quota at $D_0 - q_{sh} - q_{cv0} - q_o = 30\,\text{Mb/d}$, so that supply equals demand at I_0 and the sixty reports behind $t = 0$ all read I_0. The third case instead starts at 3.15 Gb, 0.25 Gb above I_0, after a glut that built linearly over three years, which the report history remembers as an average of 2.977 Gb. The physics advances by fourth-order Runge–Kutta on a step of a week, $\Delta t = 1/52$ year, and the controller runs every week before each step (Section 1.6.1). After each step, stocks above I_{max} are clipped and the excess booked as shut in, stocks below the working minimum $I_{min} = 0.1$ Gb are lifted and the deficit booked as unmet demand, the month-end record and the report queue are updated, and the status flags are set. The application's checks find every derivative at the balance below 1.8×10^{-15} and the reference unchanged after ten years, and over ten years of every case the stocks equal their start plus supply, less demand and shut-in oil, plus unmet demand, to 7.4×10^{-13} Gb. At half the step the challenge's suggested run keeps its verdict and its revenue to $0.1 billion. The application passes its own 23 checks and the 17 that the framework runs on every application.

10.3.6 Assumptions

One tank and one price. The world's commercial stocks are one vessel and the price is a function of their level alone. Real stocks and prices move together only loosely, through grades, regions, freight and a futures curve that the model leaves out, and a price shock can come from supply, from aggregate demand or from demand specific to oil [11].

A single decision maker. The cartel decides as one, its members do not bargain, and shale does not anticipate the cartel, so this is feedback against a price-taking fringe and not the game of [2, 8]. Cheating is an assumed slope above $80 and a set of scripted cheaters, and the loyal members' floor is a stated number.

Table 10.3 Teaching Values and Where Each Is Set: A Among the Advanced Model Parameters or Controls, C on a Controller Card, S in the Source. No Constant Is Hidden From the Reader

Symbol	Meaning	Default	Range	
D_0, g	demand (Mb/d), growth (/yr)	100, 0	70–130, −0.02–0.03	A
ε_s, ε_l	short-, long-run elasticity	0.06, 0.35	0.01–0.2, 0.2–0.7	A
τ_{down}, τ_{up}	destruction, recovery (yr)	6, 12	2–10, 6–20	A
I_0, I_{max}	normal stocks, storage (Gb)	2.9, 4.5	2–4, 2.6–6	A
P_0, k	reference price ($), sensitivity	75, 9	40–110, 3–15	A
L_r	report delay (months)	1	0–3	A
q_b, P_{be}	shale base (Mb/d), breakeven ($)	7, 45	3–12, 20–90	A
s, q_{max}	shale slope (Mb/d/$), ceiling	0.1, 15	0.02–0.3, 10–25	A
τ_{sh}	shale lag (yr)	0.75	0.25–2	A
q_{cv0}, s_{cv}	conventional (Mb/d), response	50, 0.2	35–60, 0–0.4	A
τ_{cv}, δ	investment lag (yr), decline (/yr)	4, 0.03	2–8, 0–0.1	A
q_o	other producers (Mb/d)	10	0–20	A
K	Gb per Mb/d for a year	0.365	—	S
P_c, I_{min}	overproduction price ($), working minimum (Gb)	80, 0.1	—	S
—	status limits: x, over quota, spare (Mb/d)	0.95, 0.25, 0.5	—	S
q_{cap}, q_{min}	capacity, quota floor (Mb/d)	34, 0	28–40, 0–30	A
c	overproduction (Mb/d/$)	0.03	0–0.1	A
ρ	discount rate (/yr)	0.05	0–0.15	A
$I_{spr}(0)$	strategic reserve (Gb)	0.7	0–1.5	A
P_{sp}	PI price target ($)	75	40–120	C
K_c, τ_I	price PI (Mb/d/$, yr)	0.05, 0.4	0.01–0.3, 0.1–5	C
ΔI_{sp}	stocks target above $\bar{I}_5$ (Gb)	0	±0.3	C
K_s, τ_{Is}	stocks PI (Mb/d/Gb, yr)	10, 1	1–40, 0.1–5	C
P_{lo}, P_{hi}	announced band ($)	60, 100	30–90, 70–150	C
ϕ_{min}, w, H	MPC share floor, move weight, horizon (yr)	0.27, 1, 8	0–0.4, 0.1–10, 2–10	C
P_r, K_r, τ_{Ir}	reserve ceiling ($), PI gain, time (yr)	90, 0.05, 0.5	50–150, 0.01–0.3, 0.1–3	C
Δ, Q, r_{max}	cap, quantum, largest release (Mb/d)	1.5, 0.1, 1	0.5–5, 0–0.5, 0.1–2	A
—	meeting calendar	twice a year	monthly, weekly	C
Δt, Δt_c	physics step, control period (yr)	1/52, 1/52	—	S
—	challenge: months in $60–100, share, V ($bn)	0.9, 0.3, 5950	—	S

First-order fringe. Shale and conventional supply relax toward price-dependent targets. There is no well-by-well decline, financing, bankruptcy or capacity lost to a shut-in, which matters most in the price war, where a real fringe would also lose firms.

No growth and no history. Demand growth is zero by default. With demand growing 1 % a year and fringe supply curves that do not grow, seven of the eleven shipped cases reach the cartel's capacity within the decade, at year 7.0 while a PI holds $75 and at 9.5 in the second case, so growth is an advanced parameter. No historical price series is drawn, and the cases echo 1986, 2008, 2014 and 2020 in shape and not in any figure.

Teaching values. Every price, elasticity, lag, capacity and threshold is chosen to expose a mechanism. Real values come from estimates of demand elasticities [17], of the asymmetric response of demand [16], of shale supply [13] and of the role of inventories [7], which the application anchors on but does not reproduce.

10.4 Two Speeds of the Market

The market regulates itself, because a fuller tank is cheaper oil and cheaper oil is burned faster. Linearized at the balance (Section 1.4.3), with the state (I, q_{sh}, q_{cv}, x), the stock row of the Jacobian is $(-6.80, 0.365, 0.365, -36.5)$ in the units of Table 10.1 per year, so the stocks alone would return to equilibrium at 6.80/yr. The price couples them to the fringe, because near the balance the price falls by $232.8 per Gb of extra stock and the shale and conventional targets fall with it. The four eigenvalues are a damped pair at $-3.85 \pm 3.47\,j$ per year and two slow modes at –0.651/yr and –0.225/yr when the price rises, and $-3.90 \pm 3.10\,j$, –0.560/yr and –0.142/yr when it falls. The two

sets differ because the destruction law of Equation (10.7) has a different time constant on each side of its target, so the linearization holds on one branch at a time.

Pause. Try Exercise 10.2 before reading on.

The quota's grip on the price fades between the two speeds. With the fringe and x frozen, an extra 1 Mb/d must be burned by consumers alone, a static short-run gain of $P_0/(\varepsilon_s D_0) = \$12.5$ per Mb/d. Once shale, conventional supply and demand have all adapted, the market absorbs 0.745 Mb/d for each dollar (s plus 0.179 Mb/d of conventional supply plus $\varepsilon_l D_0/P_0$), a long-run gain of \$1.34 per Mb/d, about a ninth of the short-run one. The measured step lies between the two, as shown in Figure 10.2(a). A quota of 31 Mb/d decided at the meeting of year 1 lowers the price to \$70.2 four weeks later and to its lowest, \$66.87, 0.27 year later, because shale has already begun to answer. The price is back to \$72.0 after a year, \$73.0 after three and \$73.4 after seven, with shale at 9.84 Mb/d, and it reaches \$73.67 after sixty. A cut to 29 Mb/d gives a peak of \$83.76 at 0.25 year, larger than the dip because the price law is convex, and \$76.40 after seven years.

What a held price leaves the cartel in the end does not depend on the controller. With every slow state at its target for a price P, the cartel supplies the long-run residual demand of Equation (10.10).

$$R_{lr}(P) = D_0\left(\frac{P}{P_0}\right)^{-\varepsilon_l} - q^*_{sh}(P) - q^*_{cv}(P) - q_o \tag{10.10}$$

where q^*_{sh} and q^*_{cv} are the steady states of Equations (10.4) and (10.5). After the technology shift of the fourth case it is 11.0 Mb/d at \$100 (13.5 before the shift), 16.1 at \$90 and 26.5 at \$75, a long-run share of 12.1 % at \$100 against 26.5 % at \$75. A PI controller that holds \$100 sees the fringe's growth and the destruction of demand as a disturbance that keeps rising like a ramp, and it can only follow the price by giving the barrels away. The 24 runs of Figure 10.2(b) lie along one frontier whatever their tuning, which runs close to the long-run curve, above it at high prices and below it at low ones, because eight years is not yet the long run.

10.5 Deciding at Meetings

The controller executes every week, $\Delta t_c = 1/52$ year, and an execution becomes a decision only when the cartel's calendar says so. At each execution the consumers' reserve sets its release, the selected controller forms a request q^w, and on a meeting the request is shaped into the quota that will hold until the next meeting. The chair and every automatic controller share that last step.

10.5.1 The Calendar and the Shape of a Quota

Both pieces come from the framework module `core/cadence.js`. The calendar meets twice a year, at every whole and half year from $t = 0$, as shipped, or monthly, or at every weekly execution, and a meeting falls at the first execution at or after its instant. A month is not a whole number of weeks, so monthly meetings fall four or five weeks apart, twelve a year, while the twice-yearly ones fall every 26 weeks. A meeting turns the request into the quota in force by Equation (10.11).

$$\begin{aligned} \tilde{q} &= Q\,\mathrm{round}(q^w/Q), \qquad \Delta_Q = Q\max\left(1, \lfloor \Delta/Q \rfloor\right) \\ q_k &= \Pi_{[q_{min}, q_{cap}]}\left(q_{k-1} + \mathrm{sat}_{\Delta_Q}(\tilde{q} - q_{k-1})\right) \end{aligned} \tag{10.11}$$

where $Q = 0.1\,\mathrm{Mb/d}$ and $\Delta = 1.5\,\mathrm{Mb/d}$, a tie is rounded away from zero, the bounds are first moved onto the grid, and a non-finite request leaves the quota where it was. In the application's check, which runs the challenge for six years on each of the three calendars and every case for ten years, all 916 decisions sit on the 0.1 grid, move by no more than the cap, stay within the bounds and fall on their calendar instants. No other rate limit acts on the quota, and neither the quantum nor the cap is fed back to a controller.

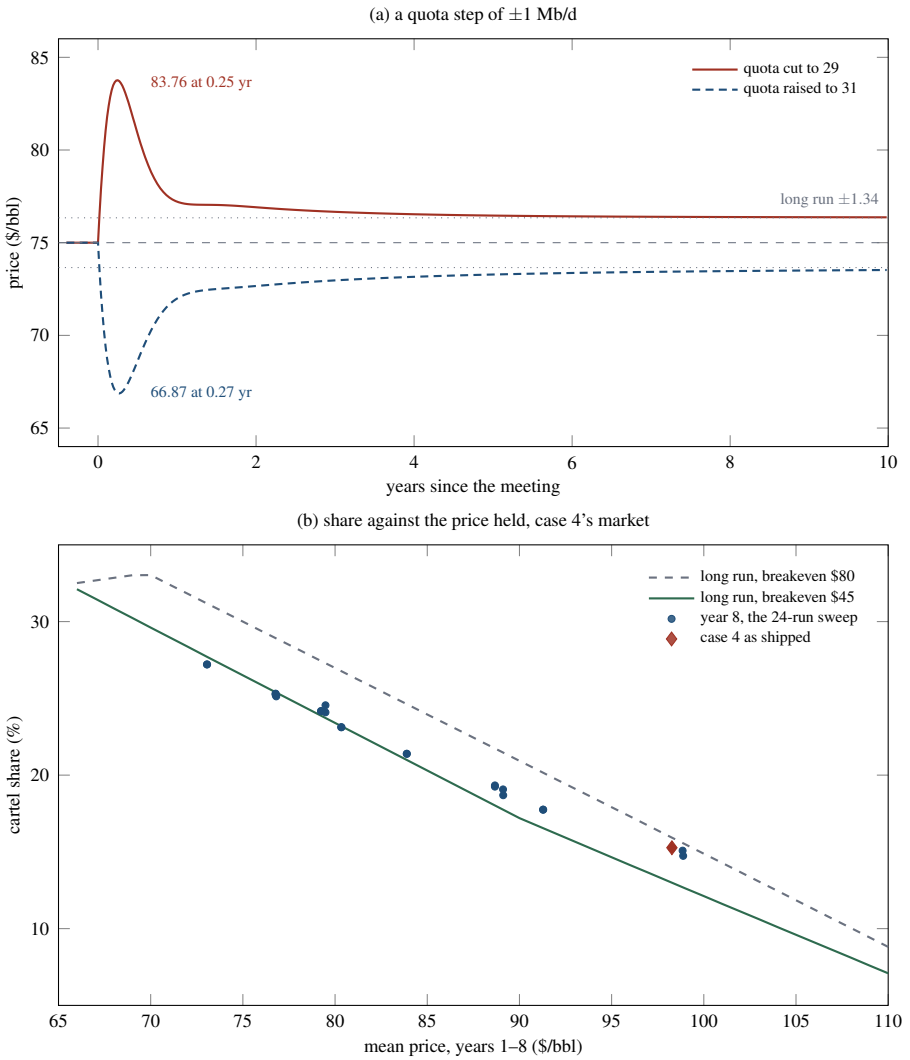


Figure 10.2 Two characteristics of the market from the application's own engine and functions. (a) The balanced reference with the quota cut (solid) or raised (dashed) by 1 Mb/d at the meeting of year 1, against the long-run gain of Section 10.4 (dotted). (b) Case 4's market, the share at year 8 against the mean price over years 1 to 8 for the 24 runs of the application's sweep (circles, the shipped case a diamond), with the long-run share, R_{lr} of Equation (10.10) over the long-run demand, before (dashed) and after (solid) the technology shift, capped where the cartel would need more than its capacity.

DID YOU KNOW? A Cut of 9.7 Million Barrels a Day, 2020

On April 15, 2020 the producer group and 10 partner countries agreed to cut crude oil output by an initial 9.7 million barrels a day from May. The group's production fell by 6.0 million barrels a day from April to May, its largest monthly decline since 1993 [18]. The simulator's cap allows 1.5 Mb/d a meeting.

10.5.2 Two PI Loops on One Valve

Each cartel loop is the framework's velocity-form PID of Section 1.6.2 with its derivative time set to zero, as Equation (10.12) writes it.

$$\begin{aligned} e_k &= a(y_k^{sp} - y_k), \qquad \Sigma_k = \Sigma_{k-1} + K_c(e_k - e_{k-1}) \\ \tilde{u}_k &= u_b + \Sigma_k + \mathscr{I}_{k-1}, \qquad u_k = \Pi_{[q_{min}, q_{cap}]}(\tilde{u}_k) \\ \mathscr{I}_k &= \mathscr{I}_{k-1} + \Delta t_c \left[K_c e_k / \tau_I + (u_k - \tilde{u}_k)/\tau_I \right] \end{aligned} \tag{10.12}$$

where Σ_k and $\mathscr{I}_k$ are the running sum S_k and the integral I_k of Equations (1.25) and (1.27), renamed because S is supply and I the stocks here, the tracking time equals τ_I, u_b is the starting quota and the request is $q^w = u_k$. The price loop has $a = -1$, $y = P$ and $y^{sp} = P_{sp}$, so a price above target raises the quota, with $K_c = 0.05\,\text{Mb/d}$ per dollar and $\tau_I = 0.4\,\text{yr}$. The stocks loop has $a = +1$, $y = I_{\text{pub}}$ and $y^{sp} = \bar{I}_5 + \Delta I_{sp}$, so a report above target cuts the quota, with $K_s = 10\,\text{Mb/d}$ per Gb and $\tau_{Is} = 1\,\text{yr}$. A stock target is a price target seen through Equation (10.2) and filtered by a sixty-month moving average of the controlled variable itself, which the application prints as the implied price on its card.

The loops run at every weekly execution, so their integrals keep collecting the error between meetings while the quota stands still, and a request is shaped only when a meeting falls. A loop that is not deciding tracks the quota in force, so a change of controller, or MANUAL to AUTO, is bumpless, and the first meeting after such a switch during a run holds the quota in force. A setting changed before the first step, a switch included, restarts the run with it, as Reset does (Section 1.6.1), so the meeting at $t = 0$ is decided with the learner's settings. In AUTO the chair's slider follows the request, not the shaped quota. SATURATED is read from the command, the loop's unclipped output beyond the floor or the capacity, or the cap binding at the last meeting.

10.5.3 The Chair, the Price War and the Consumers

In MANUAL the chair's slider, 0 Mb/d to 40 in steps of 0.1, is the request, shaped at the next meeting exactly as a controller's is. The price war is open loop, a request equal to q_{cap}. The consumers' strategic reserve is a second, independent controller on the same tank, off as shipped, a manual release or a PI on the price of Equation (10.12) with $a = -1$, $y^{sp} = P_r$, $K_r = 0.05\,\text{Mb/d}$ per dollar, $\tau_{Ir} = 0.5\,\text{yr}$ and its output clipped to $[0, r_{max}]$. The release is further limited to what the reserve holds, $r \le I_{spr}/(K\Delta t_c)$, and the reserve is never refilled.

> ⧖ HISTORY **A Reserve Born of an Embargo, 1973 to 1977**
>
> A national oil stockpile had been proposed in the United States since 1944, but the embargo of 1973 and 1974, which cut off the oil flowing to the country from the Arab members of the producer group, made the case. The Energy Policy and Conservation Act, signed on December 22, 1975, declared it national policy to establish a reserve of up to one billion barrels of petroleum. Existing salt caverns were acquired in April 1977, and on July 21, 1977 the first oil, about 412,000 barrels, was delivered [19]. The chapter's consumers' reserve is such a stock under a second controller on the same tank.

10.5.4 An Economic MPC by Successive Linearization

The economic MPC plans the quota at each meeting by maximizing discounted revenue over a horizon of $H = 8\,\text{yr}$, built with the framework's condensed linear MPC, `core/linmpc.js`, which Section 1.6.3 outlines and Volume 9 develops. Its prediction step is the twice-yearly meeting interval, $T_s = 0.5\,\text{yr}$, so $N = 16$ (on the monthly and weekly calendars T_s is a nominal month of 1/12 year), and the inputs are held over nine move blocks, four of one step, two of two, one each of three and four, and a last block of one step. At a nominal state $\bar{x}$ and quota $\bar{q}$ the four slow states of Equations (10.1) to (10.7) are linearized, by central differences in the states and a one-sided difference in the quota, backward at capacity and forward elsewhere, and discretized exactly for a zero-order hold, with the affine remainder and the previous quota carried as two more states. The price is linear in the stocks

with slope $p_I = -kP(\bar{x})/I_0$, and the share linear in the fringe and the quota. The quadratic program (QP) is Equation (10.13).

$$\begin{aligned}
\min_{u,\sigma} \; & \sum_{j=1}^{N} \Big[-KT_s \mathrm{e}^{-\rho t_j} \left(\bar{q}_c P_j + \bar{P} u_j \right) + 10^{-4} P_j^2 \Big] \\
& + R \sum_{j=1}^{N} (u_j - \bar{q})^2 + w \sum_{b} (u_b - u_{b-1})^2 \\
& + 30\sigma_1^2 + 10\sigma_1 + 10^5 \sigma_2^2 + 10^4 \sigma_2 \\
\text{s.t. } & \underline{u} \le u_j \le \bar{u}, \\
& |u_b - u_{b-1}| \le \Delta, \quad P_{lo} - \sigma_1 \le P_j \le P_{hi} + \sigma_1, \\
& \phi_N \ge \phi_{min} - \sigma_2, \quad \sigma \ge 0
\end{aligned} \tag{10.13}$$

where u_j is the quota over the step that ends at $t_j = t + jT_s$, the bounds are $\underline{u} = \max(q_{min}, \min(q, \bar{q} - 6))$ and $\bar{u} = \min(q_{cap}, \max(q, \bar{q} + 6))$, $\bar{q}_c$ and $\bar{P}$ are the cartel's output and the price at the nominal point, so the first term is revenue linearized in price and output, $w = 1$ billion dollars per $(\mathrm{Mb/d})^2$ suppresses moves between blocks, with $u_0 = q$ the quota in force, and $R = KT_s |p_I B_{d,I}| \mathrm{e}^{-\rho(t+NT_s/2)}$ restores the curvature that a linear revenue leaves out, one more barrel a day for a step lowering the price by $|p_I B_{d,I}|$. The bounds hold the quota within 6 Mb/d of the nominal, a trust region that always contains the quota in force, the price band is soft at every step and the share floor $\phi_{min} = 0.27$ soft at the last. The program has 11 variables and 71 constraint rows, and `core/qp.js` solves it by the numerically stable dual method [20] with a ridge of 10^{-9} times one plus the largest diagonal entry. After each solve the plan is rolled out on the nonlinear model in weekly Euler substeps, and the next linearization is taken at the rollout's mean state and the plan's mean quota. Five such solves make one meeting's plan, and the plan's first quota is the request that the meeting then rounds and caps. In the eighth case the 100 solves of ten years take 5.10 active-set iterations on average and 9 at most, with no failure, and across that case and the challenge no planned price leaves the band in the QP's prediction and the nonlinear rollouts leave it by at most \$0.83.

An earlier version of the application had one defect in this linearization, since corrected. A forward difference in the quota reads zero once the cartel's output sits at its capacity, where one more barrel meets the clip of Equation (10.3), so the QP saw no effect of the quota on the price and the MPC never left capacity. The difference is now taken from below there. The logged runs of all eleven scenarios over thirty years and the challenge's suggested run are byte-identical before and after that correction, and two learner experiments changed. With the eighth case run to year 9 and the floor then raised to \$80, the MPC used to stay at 34 Mb/d with the price at \$64.8 six years later, and it now cuts to 25.8 Mb/d and holds \$78.2. Selected at year 5 of the fifth case, it used to hold capacity at \$56.2 a year later, below its own \$60 floor, and it now holds \$60.9 at 32.5 Mb/d. A new check in the application runs both experiments.

10.5.5 Market, Prediction and Literature

The simulation model is Equations (10.1) to (10.9) with every clip, the storage limits, the monthly report and the shocks. The PI loops hold no model at all. The MPC's internal model is the same four slow states with every coefficient at its declared default, whatever the run's parameters, linearized afresh at each meeting, with no storage limit inside the QP and no report delay. Each plan starts from the cartel's own estimate. The declared model runs forward week by week with the quota and the reserve's release, its stocks are read from the price through Equation (10.2) whenever the two disagree, and the barrels a day that the model cannot account for are held over the horizon as one unexplained supply, with no shock to come foreseen. The destruction state x is never read and follows the declared equation. At the declared defaults with no shock the estimate equals the simulated market at every step. The models these stand for are the cartel-and-fringe models [2, 8] and the econometric

Table 10.4 The Cases and the Challenge

#	Change from the defaults	Teaching point
1	MANUAL, recession at year 2	the market regulates
2	$P_{sp} = \$80$	share erodes every year
3	PI on stocks, start after a three-year glut	a target that drifts
4	$P_{sp} = \$100$, $K_c = 0.2$, $\tau_I = 0.2$, monthly, shale shift from year 1	honestly infeasible
5	case 4 from year 0, valve at capacity from year 3	a war against a lag
6	$I_{max} = 3.5$ Gb, pandemic at year 1	the tank fills
7	$K_c = 0.12$, $\tau_I = 0.3$, monthly, band \$65–85, $q_{min} = 26$, a cheater a year from year 2	a leaking valve
8	economic MPC, shale shift from year 1	what a band costs
9	$P_{sp} = \$90$, reserve PI at \$80, outage at year 2	two controllers
ch	$P_{sp} = \$90$, recession and shale shift at year 2	choose the price
ref	overproduction off, nothing moves	the balance

Table 10.5 Measured Cases

Case	Result
1, hands off	low \$59.1 at yr 3.04, overshoot \$80.1 at 5.02
1, PI at \$75	low \$63.9 at yr 3.00
2	mean \$78.33 over yr 2–10, share 29.75, 28.81, 28.15 % at yr 1, 5, 10
3	implied target \$58.2, 59.5, 68.7 at yr 1, 4, 10, mean \$66.2 over yr 1–5
3, PI at \$75	mean \$73.7 over yr 1–5, share 30.24 % at yr 10
4	mean \$98.3 over yr 1–8, share 30 to 15.27 % at yr 8
5	\$23.5 at yr 3.94, shale 14.14 to 10.11 Mb/d at 4.58, V \$5603 billion
6	full yr 1.115–1.500, 0.687 Gb shut in, spike \$123.5 at 2.77
7	floor at yr 6.10, band broken at 8.12, \$61.1 at 9.25
8	mean \$62.1 over yr 1–10, share 32.64 %, V \$5973 billion
8, $P_{lo} = \$80$	mean \$78.0, share 24.70 %, V \$5951 billion
8, $P_{lo} = \$70$	mean \$68.9, share 29.18 %, V \$6047 billion
9	reserve empty at yr 4.33, \$86.3 at yr 10 against \$86.1 without
ref	largest derivative 1.8×10^{-15}, no drift in ten years

links between stocks, demand and price [6, 7, 17], whose coefficients are not copied here. Nothing in the application fits a model to data.

10.6 Ten Years at a Time

Table 10.4 lists the cases. Each starts from the defaults of Table 10.3 with the PI on price at \$75, the twice-yearly calendar, a 1.5 Mb/d cap and a 0.1 Mb/d quantum, and changes only what its row lists. The values quoted below were measured headless on each case as the dropdown loads it, with any variation applied before the first step, which gives the same run as the briefings' Reset, and Table 10.5 gathers the main ones. A mean over a span of years averages the weekly rows from its first year to its last, both included, and a quota decision is timed at its meeting. The scorecard's integrated absolute error is left out, since it grows with the length of the run.

10.6.1 A Chair, a Recession and a Held Price

In the first case the learner chairs the cartel while a recession takes 3 Mb/d of demand away over a year from year 2 and returns it over the next two. Left at 30 Mb/d, the quota sees \$65.2 at the meeting of year 2.5, the price bottoms at \$59.1 at year 3.04 and then recovers by itself, because consumers burn more at the lower price and shale falls to 9.01 Mb/d by year 3.33. When demand returns and shale has not, the price overshoots to \$80.1 at year 5.02. A cut to 28.5 Mb/d asked at year 2.1 takes effect at the meeting of year 2.5, keeps the price above \$65.2 and, if never given back, lifts it to \$82.6 after the recession, while asking for 30 Mb/d again at year 3.6, in force from the meeting of year 4, keeps it between \$67.9 and \$79.5 from then on. In AUTO the PI on price holds it above \$63.9.

The second case holds \$80 with the shipped PI. The price is \$76.8 at year 1 and averages \$78.33 over years 2 to 10, never more than \$2.85 from the target after year 2, while the share falls every year, 29.75, 28.81 and 28.15 % at years 1, 5 and 10. By year 10 the quota is 27.8 Mb/d, shale has gained 0.37 Mb/d, conventional supply 0.58 Mb/d, and 1.00 Mb/d of demand has been destroyed. The constant shortfall is a PI following a ramp. The calendar caps the gain, because the price settles within a season of a quota change. At $K_c = 0.06$ the quota reverses direction at 7 of the 20 meetings, at 0.07 at 14 and at 0.08 at 18, where the mean error grows to \$5.15.

Pause. Try Exercise 10.1 before reading on.

The third case starts after a three-year glut, with the stocks at 3.15 Gb, the price at \$34.5 and a five-year average of 2.977 Gb that implies \$59.0. The glut drains within months, to \$72.7 at year 0.25, but the PI on stocks aims the report at an average that still holds the glut. Its implied target is \$58.2 at year 1, \$59.5 at year 4, \$66.6 at year 6 and \$68.7 at year 10, when the price is \$68.7, against \$75.1 for the PI on price at \$75. The lower price buys share, 33.18 % against 30.24 % at year 10, for \$39 billion less revenue (6424 against 6463). The case's briefing says that the cartel's communiqués have stated their aim this way since 2017, a statement the application makes and this chapter does not test.

Pause. Try Exercise 10.3 before reading on.

10.6.2 The Shale Revolution and the Price War

The fourth case is the honestly infeasible one. A monthly PI with $K_c = 0.2$ and $\tau_I = 0.2$ yr holds \$100 while the breakeven falls from \$80 to \$45 between years 1 and 5. The price is \$96.3 at year 1, averages \$98.3 over years 1 to 8 and is \$99.5 at year 8, and the share falls from 30 to 15.27 %, a loss of 49 %, with the quota at 13.6 Mb/d, shale at its 15 Mb/d ceiling, conventional supply at 53.74 Mb/d and 5.49 Mb/d of demand destroyed. The demand factor first falls below 0.95 at year 6.85. In the application's sweep of targets of \$80, \$90 and \$100, four tunings and monthly and weekly calendars, every run that averages \$95 or more loses at least 49 % of its share, and the runs that lose less than a third average at most \$83.9. At a \$75 target the PI keeps 26.67 % at year 8. The state at year 8 is shown in Figure 10.3.

> **TRY IT IN THE APP** **4 · The Shale Revolution: Hold \$100**
>
> Load the case and press Run. Watch the share badge while the price sits near its \$100 target. Over eight years the price averages \$98.3, shale climbs to its 15 Mb/d ceiling, and the cartel's share falls from 30 to 15.27 %.

The fifth case opens the valve. After three years of holding \$100 while the breakeven falls, the price is \$97.2, the quota 19.3 Mb/d, the share 20.6 % and shale 14.14 Mb/d. The switch to the price war holds the quota at the meeting of year 3, and from the next meeting it climbs 1.5 Mb/d at each monthly meeting to capacity, reached at the meeting of year 3.85. The price halves by year 3.46 and bottoms at \$23.5 at year 3.94, when the stocks peak at 3.274 Gb. Shale falls half of its eventual fall by year 3.75 and bottoms at 10.11 Mb/d at year 4.58, then grows again, because the price, \$49.8 at year 5 and \$63.1 at year 10, is above the new breakeven. The war wins a year of shut-ins and not the fringe's exit. It still earns more over ten years, \$5603 billion, than case 4 earns by holding \$100, \$5269 billion, with the share back to 33.0 % at year 5.

10.6.3 A Full Tank and a Leaking Valve

The sixth case cuts storage to 3.5 Gb and takes 20 Mb/d of demand away for three months from year 1. Its default capacity follows normal inventory plus 0.6 Gb, unless the learner explicitly edits capacity. On the twice-yearly calendar the tank is full from year 1.115 to 1.500, the price sits at the full-tank \$11.7 and 0.687 Gb is shut in. Monthly decisions shut in 0.309 Gb and weekly ones 0.169 Gb, both with the 1.5 Mb/d cap, and only weekly decisions with a 5 Mb/d cap and $K_c = 0.3$,



Figure 10.3 Case 4, the shale revolution, at $t = 8.00$ yr in the Fringe and demand view, in AUTO as loaded, cropped to the plant view. (1) The badges read price \$99.5, share 15.3 %, stocks less the five-year average −0.002 Gb, spare capacity 19.8 Mb/d, discounted revenue \$4635 billion and 5.49 Mb/d of demand destroyed. (2) The shale panel shows today's supply curve above the \$45 breakeven beside the dashed curve it started on, with output at its 15 Mb/d ceiling. (3) The demand panel shows the steep short-run curve through the operating point and the flatter long-run curve. (4) The last panel shows what is left for the cartel at \$100, 14.2 Mb/d now and 11.0 Mb/d in the long run. (5) The status line reads DEMAND DESTRUCTION.

with τ_I left at 0.4 yr, keep the tank from filling, at a lowest share of 15.9 %. On every calendar the tank fills six to eight weeks after the shock, at year 1.115, 1.135 or 1.154. When demand returns, the cartel's cuts and shale's shut-ins, shale down to 6.17 Mb/d at year 1.85, are still in force, and the price spikes to \$123.5 at year 2.77 (\$107.7 monthly, \$107.4 weekly). A storage limit saturates the plant itself, and a controller too slow on the way down is too slow on the way up.

The seventh case holds \$75 inside an announced band of \$65 to \$85 with a monthly PI while one more member overproduces 1 Mb/d every year from year 2, and the loyal members will not go below a quota of 26 Mb/d. The first cheater takes the price to \$69.4 at year 2.10, and the integral has it back to \$74.6 by year 2.9. With four cheaters producing, the price is still \$74.2 at year 5.9 with the quota at 26.7 Mb/d. The quota reaches its floor at the meeting of year 6.10, after which each cheater is uncompensated, and the price leaves the band at year 8.12 and reaches \$61.1 at year 9.25, while the cartel's total output climbs to its 34 Mb/d and its share to 33.4 %. With $\tau_I = 5$ yr the band breaks at year 5.10, and on the twice-yearly calendar at year 3.12. Integral action is indifferent to why the output is wrong, which compensates the leak until the actuator runs out of travel.

> **TRY IT IN THE APP** **7 · The Cheater: An Actuator That Leaks**
>
> Load the case and run ten years. Watch the applied quota on the trend. Each new cheater pulls the price down and the integral cuts the quota, until the quota reaches its floor of 26 Mb/d at the meeting of year 6.10. From then on each cheater goes uncompensated, and the price leaves the band at year 8.12.

10.6.4 What a Band Costs

The eighth case gives case 4's market to the economic MPC with a \$60 to \$100 band and a 27 % share floor. It holds the price near the bottom of the band, at a mean of \$62.1 over years 1 to 10 and never below \$60.1, with the quota between 33.0 and 33.9 Mb/d from year 1 and at capacity from year 8, and ends with 32.64 % and \$5973 billion. Raising the floor to \$80 is the price of a band, a mean of \$78.0, 24.70 % and \$5951 billion, while case 4's PI holding \$100 in the same market keeps 14.39 % and earns \$5269 billion, case 4's own result. The MPC is a local optimizer and not the best policy. A floor of \$70 earns \$6047 billion, more than its own choice, and a PI at \$75 earns \$6044 billion and at \$70 \$6031 billion. The linearized revenue overrates one more barrel, and the internal model holds the extra supply it has inferred so far over eight years without knowing that the breakeven will keep falling. A share floor of 0.35 changes little (32.62 %, \$5978 billion).

> **DID YOU KNOW?** **The Optimizer's Own Choice Is Not Its Best**
>
> From the simulator: in case 8 the economic MPC settles near the bottom of its band, at a mean of \$62.1, and earns \$5973 billion. Raising its price floor to \$70 gives a mean of \$68.9 and earns \$6047 billion, more than its own choice, because the optimizer is local (Table 10.5).

10.6.5 Two Controllers on One Tank

In the ninth case the cartel's PI holds \$90 and the consumers' reserve, with 0.7 Gb, releases up to 1 Mb/d when the price is above \$80, while other producers lose 1.5 Mb/d for a year from year 2. The reserve takes the outage's peak from \$93.8 to \$90.7, but it fights the cartel the whole time. The cartel's integral cuts the quota for every barrel released, the price averages \$81.79 over years 0.5 to 4 against \$82.78 without the reserve, and the reserve is empty at year 4.33. At year 10 the price is \$86.3 with the reserve and \$86.1 without, and the cartel has earned \$6306 billion against \$6349 billion. Two integrators with different setpoints on one level fight until one saturates. With a ceiling of \$95 the reserve never releases and keeps its 0.7 Gb.

10.6.6 A Shale Decade

The challenge runs ten years with a recession at year 2, the breakeven falling from \$80 to \$45 between years 2 and 6, twice-yearly meetings and the 1.5 Mb/d cap. It asks for at least 90 % of month-end prices inside \$60 to \$100, a share of at least 30 % at year 10 and at least \$5950 billion of discounted revenue. Table 10.6 compares nine configurations. The shipped PI at \$90 never reaches its target, ends at \$84.9 and keeps every month in the band, but it cuts the quota to 20.8 Mb/d and ends with 21.61 % and \$5868 billion, short on share by 8.39 points and on revenue by \$82 billion. The suggested MPC with its floor raised to \$65 passes with 91.7 % of months, 31.77 % and \$5982 billion, margins of 1.77 points and \$32 billion, and so do a PI at \$65 and the PI on stocks 0.02 Gb above its average. At the shipped \$60 floor the unforeseen recession takes the price to \$50.7, and the MPC passes with no month to spare, 108 of 120 inside the band. The \$5 of margin in the suggested floor is back-off against a disturbance the MPC does not predict, and it buys two months. PI loops at \$70 and \$75 earn more than any passing run, \$6011 billion and \$6021 billion, and fail on share. The challenge is judged on the case as given, so a changed model parameter or disturbance, or an extra shock fired, voids an attempt. Figure 10.4 draws the nine settings against the three limits.

A setting chosen before the first step gives the same run as the setting carried through Reset, so the suggested run as Reveal applies it is the one in Table 10.6.

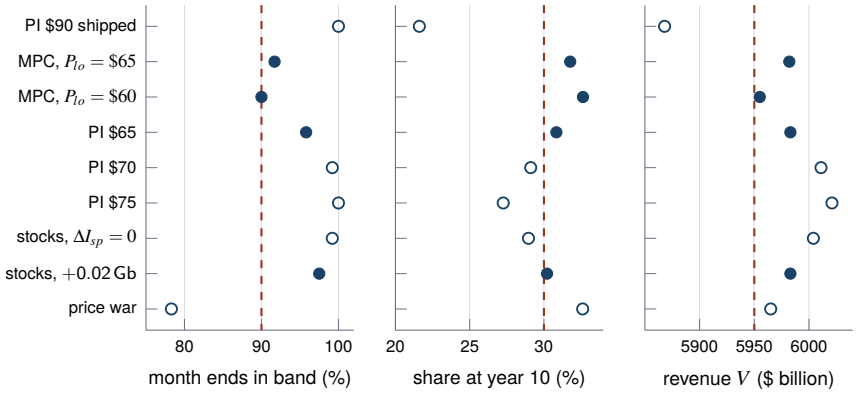


Figure 10.4 The ten-year challenge in nine settings, from Table 10.6: month ends inside $60–100, the share at year 10 and the discounted revenue, each against its limit (dashed line). Filled circles pass all three limits, open circles fail at least one. Notice that the PI loops at $70 and $75 earn the most and fail on share, and that every passing setting sits just above the share and revenue limits.

Table 10.6 The Ten-Year Challenge in Nine Configurations

Setting	band (%)	months in band	share (%)	V ($bn)	
PI $90, shipped	100.0	120	21.61	5868	fail
MPC, $P_{lo} = \$65$	91.7	110	31.77	5982	pass
MPC, $P_{lo} = \$60$	90.0	108	32.63	5955	pass
PI $65	95.8	115	30.84	5983	pass
PI $70	99.2	119	29.11	6011	fail
PI $75	100.0	120	27.26	6021	fail
stocks, $\Delta I_{sp} = 0$	99.2	119	28.96	6004	fail
stocks, +0.02 Gb	97.5	117	30.21	5983	pass
price war	78.3	94	32.61	5965	fail

Band: month ends in $60–100, of 120.

In the balanced reference the derivatives vanish to 1.8×10^{-15} and the price stays at $75 for ten years with the overproduction turned off.

10.7 Technology in Practice

The model of Section 10.3 reads a price and a stock report and moves a quota, signals made by markets, surveys and meetings. Figure 10.5 draws the loop with them in place of the schematic's symbols.

What Measures It

The price is a market price, and the spot price and the futures prices answer the stocks through the incentive to store oil [21]. The stocks are read by surveys. The industrialized countries usually publish stock statistics on a regular basis, while data for other countries, including fast-growing developing consumers and major producers, come late or not at all, and crude oil is also stored on ships at sea [21]. Even in the United States weekly report some figures are a model's output: weekly domestic crude oil production is estimated from the statistical agency's short-term outlook model and daily pipeline data, and most of the uncertainty of the estimate comes from that model, the more so when production changes quickly [22]. The chapter's price loop reads the true price at once, and its stocks loop reads a report published a month late beside the average of sixty reports.

TECHNOLOGY Sensor: Surveys of Commercial Stocks

No instrument reads the world's stocks. In the United States a statistical agency surveys operators each week, adding companies from the largest down until the sample covers about 90 % of the volume of each item and region. The week runs from Friday to Friday, and the figures appear the following Wednesday. Missing reports are filled with exponentially smoothed means of the company's recent values, and a company must resubmit an error greater than 5 %. The weekly figures are leading indicators of a complete monthly survey, the most accurate data, and stocks are the most accurate of the weekly estimates [22]. Its faults are delay and missing coverage rather than noise: outside the industrialized countries stock data come late or not at all [21].

What Moves It

The final element is a target for each member's production. The agreement of 2020 set an initial cut and then a taper through April 2022, and three members were exempt from it [18]. Production in the group is mainly in the hands of national oil companies, while producers outside it decide independently, as price takers that usually produce at or near full capacity [23]. The chapter's quota, shaped to a quantum of 0.1 Mb/d and a cap of 1.5 Mb/d a meeting, is this target, and its price-taking fringe is the rest of the market.

TECHNOLOGY Final Element: Production Targets and Spare Capacity

The producer group acts by setting production targets, limits on how much oil each member may produce. Members do not always adhere to them, and outages can cut their output further. The room to move up is spare capacity, defined by the United States statistical agency as production that can be brought online within 30 days and sustained for at least 90 days, and the group's members hold almost all of it. When spare capacity is low, prices tend to carry a rising risk premium, and the targets themselves follow a changing market with a lag [24]. The chapter's overproduction, cheaters, capacity and risk premium stand for these traits.

Where the Controller Runs

The controller is no plant computer but a conference of the members' governments, which the founding agreement asked to meet at least twice a year and more often if necessary [15]. The application executes its controller every week, $\Delta t_c = 1/52$ year, and turns a request into a quota only at a meeting through the framework's calendar module. Its PI loops are the velocity-form block of Section 1.6.2, and its MPC replans eight years at each meeting.

IN PRACTICE Stocks Are Read Against Past Years

Analysts do not read a stock level alone. Stocks are most usefully judged against the levels of prior years in the same calendar quarter, because demand for some products is seasonal [21], and the weekly report plots each stock against the highest and lowest levels of the same week over the prior five years [22]. They also watch the futures spread, the contract 12 months ahead less the one a month ahead. An expected shortage raises futures prices and pays companies to build stocks, while a sudden loss of supply lifts the spot price and draws stocks down [21]. The chapter's stocks loop holds its report against a five-year average in the same spirit, without the seasons.

What Is Modernizing It

Three changes speed up the measurement. Six international organizations launched a joint oil data exercise in April 2001, a monthly questionnaire on 42 key data points, and within a year more than 70 countries, with 90 % of world supply and demand, took part. It became a permanent initiative with a database open to all, whose stated challenge is wider coverage, shorter delays and better quality [25]. Administrative records now replace some surveys: weekly United States export figures come from

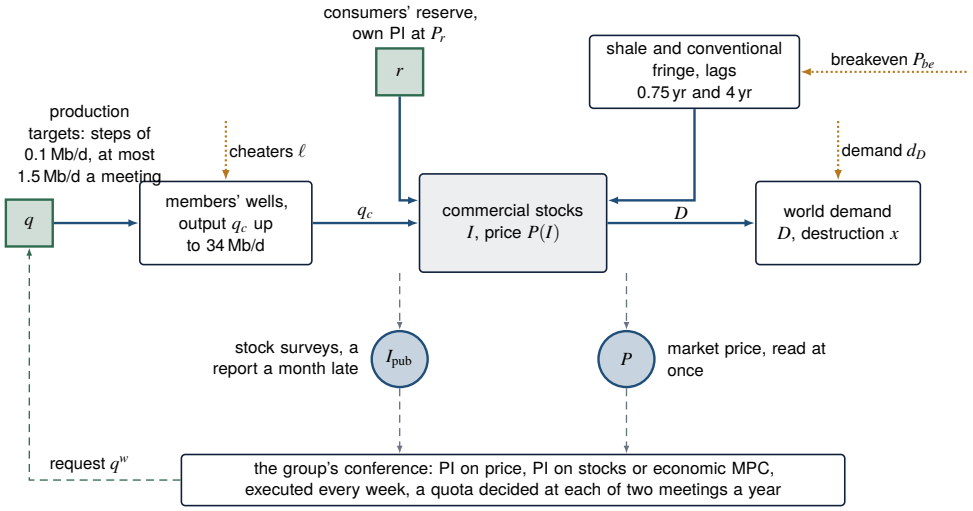


Figure 10.5 The quota loop with the institutions of the real market in place of the schematic's symbols. Surveys give the published stocks I_{pub} a month late and the market gives the price P at once, the conference forms a request each week, and only a meeting turns it into a step of the production targets q. The consumers' reserve r is a second controller on the same tank. Notice that the fringe answers the price with lags of years while the quota can move only 1.5 Mb/d a meeting, and that cheaters leak past the valve.

the electronic filings that exporters make with the customs agency, where the official monthly figures arrive about six weeks after the month ends. And model-based estimates fill the weeks between complete surveys, compared with the survey data and rebenchmarked when the two differ widely [22]. The chapter's MPC estimates the unexplained supply from the price in the same way.

10.8 Limitations

The market is one tank with one price, one decision maker and a first-order fringe, without grades, regions, futures or strategic play, and every coefficient is a teaching value. Demand growth is zero, the fringe's supply curves do not grow, and shut-in oil costs nothing later. The price controller reads the true price at once, the stocks loop a report a month old, and the MPC the price through its declared model, from which it estimates the rest of the market. The MPC is a local solution that a $70 band floor beats by $75 billion. The challenge's limits, like every other threshold here, were set for teaching.

> **OPEN CHALLENGE** **No Forecast Beats a Price That Stays Put**
>
> A plan for quotas needs the future price. A benchmark study with data through the collapse of 2008 found that at longer horizons the forecast that the real price stays where it is typically has the lowest mean squared error. Forecasts from futures prices could not be recommended, and the United States statistical agency's own forecasts had at best modest success in the short run and none at longer horizons. Results in the literature were often sensitive to the sample period [26]. The chapter's MPC plans eight years ahead and foresees no shock. Progress would be forecasts that beat the no-change forecast out of sample over the years that the fringe and the consumers take to answer.

10.9 Conclusion

The simulator writes a cartel's dilemma as a level loop whose setpoint pays the competing inflows and narrows the outflow. An extra 1 Mb/d moves the price $8.1 in a season and $1.6 in seven years,

and its static gains differ by a factor of 9.3. A PI holding \$80 loses share every year, from 29.75 to 28.15 %, and holding \$100 through a shale revolution costs 49 % of the share in eight years, with no tuning in the 24-run sweep doing better than 49 % at a mean of \$95 or more. A price war halves the price within half a year and wins a year of shut-ins, a full tank sets the price at \$11.7 on every calendar under the 1.5 Mb/d cap, and an integral compensates four cheaters before the floor binds. The economic MPC holds about \$62 for 32.64 % and \$5973 billion, and a \$65 floor passes the challenge that the shipped \$90 PI fails at 21.61 %. The market is one aggregate tank with teaching coefficients. A fuller model would add a strategic fringe, regional stocks and an estimator for the destruction state. The tank makes the price a cartel can hold a property of the market, measured in barrels and years.

KEY POINTS

- Oil held back lifts the price within weeks, but the high price pays rivals and thrifty buyers for years, and their barrels become the cartel's lost market.
- An extra 1 Mb/d moves the price \$8.1 in a season and \$1.6 in seven years, and the static gains differ by a factor of 9.3.
- Holding \$100 through a shale revolution costs 49 % of the share in eight years, whatever the tuning.
- An integral compensates four cheaters before the floor binds, and a \$65 floor on the economic MPC passes the challenge that the shipped \$90 PI fails.

Where else this applies. Volume 6, Chapter 11 orders stocks on a calendar in a quantum with a cap, and Volume 2, Chapter 11 decides a policy rate through the same calendar module.

References

[1] H. Hotelling, "The economics of exhaustible resources," *Journal of Political Economy*, vol. 39, no. 2, pp. 137–175, 1931.

[2] S. W. Salant, "Exhaustible resources and industrial structure: A Nash-Cournot approach to the world oil market," *Journal of Political Economy*, vol. 84, no. 5, pp. 1079–1093, 1976.

[3] R. S. Pindyck, "Gains to producers from the cartelization of exhaustible resources," *The Review of Economics and Statistics*, vol. 60, no. 2, pp. 238–251, 1978.

[4] D. Gately, "Lessons from the 1986 oil price collapse," *Brookings Papers on Economic Activity*, vol. 1986, no. 2, pp. 237–284, 1986, with comments by M. A. Adelman and J. M. Griffin.

[5] 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.

[6] R. S. Pindyck, "The dynamics of commodity spot and futures markets: A primer," *The Energy Journal*, vol. 22, no. 3, pp. 1–29, 2001.

[7] L. Kilian and D. P. Murphy, "The role of inventories and speculative trading in the global market for crude oil," *Journal of Applied Econometrics*, vol. 29, no. 3, pp. 454–478, 2014.

[8] A. Behar and R. A. Ritz, "OPEC vs US shale: Analyzing the shift to a market-share strategy," *Energy Economics*, vol. 63, pp. 185–198, 2017.

[9] B. Fattouh, R. Poudineh, and A. Sen, "The dynamics of the revenue maximization–market share trade-off: Saudi Arabia's oil policy in the 2014–15 price fall," *Oxford Review of Economic Policy*, vol. 32, no. 2, pp. 223–240, 2016.

[10] J. D. Hamilton, "Causes and consequences of the oil shock of 2007–08," *Brookings Papers on Economic Activity*, vol. 2009, no. 1, pp. 215–261, 2009.

[11] L. Kilian, "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market," *American Economic Review*, vol. 99, no. 3, pp. 1053–1069, 2009.

[12] C. Baumeister and L. Kilian, "Forty years of oil price fluctuations: Why the price of oil may still surprise us," *Journal of Economic Perspectives*, vol. 30, no. 1, pp. 139–160, 2016.

[13] H. C. Bjørnland, F. M. Nordvik, and M. Rohrer, "Supply flexibility in the shale patch: Evidence from North Dakota," *Journal of Applied Econometrics*, vol. 36, no. 3, pp. 273–292, 2021.

[14] S. Corbet, J. W. Goodell, and S. Günay, "Co-movements and spillovers of oil and renewable firms under extreme conditions: New evidence from negative WTI prices during COVID-19," *Energy Economics*, vol. 92, p. 104978, 2020.

[15] United Nations, "Agreement concerning the creation of the organization of the petroleum exporting countries (OPEC), done at Baghdad, on 14 september 1960," United Nations Treaty Series, vol. 443, No. 6363, 1962. [Online]. Available: https://treaties.un.org/doc/Publication/UNTS/Volume%20443/volume-443-I-6363-English.pdf

[16] D. Gately and H. G. Huntington, "The asymmetric effects of changes in price and income on energy and oil demand," *The Energy Journal*, vol. 23, no. 1, pp. 19–55, 2002.

[17] C. A. Dahl, "Measuring global gasoline and diesel price and income elasticities," *Energy Policy*, vol. 41, pp. 2–13, 2012.

[18] U.S. Energy Information Administration, "OPEC+ agreement to reduce production contributes to global oil market rebalancing," Today in Energy, 2020, september 23, 2020. [Online]. Available: https://www.eia.gov/todayinenergy/detail.php?id=45236

[19] U.S. Department of Energy, Office of Petroleum Reserves, "SPR origins," Web page, 2022, last updated September 9, 2022. [Online]. Available: https://www.energy.gov/hgeo/opr/spr-origins

[20] 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.

[21] U.S. Energy Information Administration, "What drives crude oil prices: Balance," Web page, undated, accessed November 21, 2026. [Online]. Available: https://www.eia.gov/finance/markets/crudeoil/balance.php

[22] ——, "Weekly petroleum status report, appendix B: Explanatory notes and detailed methods report," Frame table dated August 2026, accessed November 21, 2026. [Online]. Available: https://www.eia.gov/petroleum/supply/weekly/pdf/appendixb.pdf

[23] ——, "What drives crude oil prices: Supply non-OPEC," Web page, undated, accessed November 21, 2026. [Online]. Available: https://www.eia.gov/finance/markets/crudeoil/supply-nonopec.php

[24] ——, "What drives crude oil prices: Supply OPEC," Web page, undated, accessed November 21, 2026. [Online]. Available: https://www.eia.gov/finance/markets/crudeoil/supply-opec.php

[25] Joint Organizations Data Initiative, "History," Web page, undated, accessed November 21, 2026. [Online]. Available: https://www.jodidata.org/about-jodi/history.aspx

[26] R. Alquist, L. Kilian, and R. J. Vigfusson, "Forecasting the price of oil," Board of Governors of the Federal Reserve System, International Finance Discussion Papers 1022, Jul. 2011.

10.10 Exercises

10.1 Concept check · 2 min

In case 2 the PI holds the price within \$2.85 of its \$80 target after year 2, yet the cartel's share falls every year, from 29.75 to 28.15 % between years 1 and 10. Which explanation does the chapter give?

(a) The twice-yearly calendar caps the gain, so the PI cannot reject the steady disturbance that a price above \$75 creates

(b) The members overproduce above \$80, so the quota leaks barrels to cheaters while the price is held at the target

(c) The held price keeps moving shale, conventional supply and demand, a disturbance that grows like a ramp, and the PI answers by cutting the quota

(d) The monthly stock report is a month late, so the PI acts on old data and gives away barrels at every meeting

10.2 Concept check · 2 min

Linearized at the balance, the market has the eigenvalues $-3.85 \pm 3.47\,j$, –0.651 and –0.225/yr when the price rises, and $-3.90 \pm 3.10\,j$, –0.560 and –0.142/yr when it falls. Why are there two sets?

(a) The price law is exponential, so its slope at the balance is steeper for a rise than for a fall

(b) Overproduction begins above \$80, so the cartel's output answers a rising price and not a falling one

(c) Shale is clipped at its 15 Mb/d ceiling when the price rises, which removes one of its states

(d) The demand destruction state moves with $\tau_{down} = 6\,\text{yr}$ toward a lower target and with $\tau_{up} = 12\,\text{yr}$ toward a higher one

10.3 Concept check · 2 min

In case 3 the PI on stocks ends year 10 with the price at \$68.7, while the PI on price, run on the same start, ends at \$75.1. Why does the stocks loop settle lower?

(a) Its setpoint is the five-year average of its own reports, which carries the glut forward and then follows wherever the stocks have been

(b) The report is a month late, and a delayed measurement leaves an integral loop with a steady offset

(c) The stocks loop is direct acting, so its integral pushes the price down where the price loop pushes it up

(d) Its one-year integral time is too long for the integral to remove the error within ten years

10.4 Theory and calculation · 6 min

Use the price law of Equation (10.2) with $P_0 = \$75$, $I_0 = 2.9\,\text{Gb}$, $k = 9$ and no risk premium, and the demand law of Equation (10.6) with $x = 1$ and $D_0 = 100\,\text{Mb/d}$.

(a) Find the stocks at which the price is \$100 and \$50. *Checkpoint:* 2.807 Gb and 3.031 Gb.

(b) Find the slope $\mathrm{d}P/\mathrm{d}I$ at the balance. *Checkpoint:* −\$232.8 per Gb.

(c) With the fringe and x frozen, find the diagonal entry of the stock row of the Jacobian of Equation (10.1), and the time constant of the stocks alone in weeks. *Checkpoint:* −6.80/yr, a time constant of 0.147 yr, about 7.7 weeks.

10.5 Theory and calculation · 6 min

Use the defaults of Table 10.3 at the balance, $P_0 = \$75$ and $D_0 = 100\,\text{Mb/d}$.

(a) With shale, conventional supply and x frozen, find the static gain from the cartel's output to the price. *Checkpoint:* \$12.5 per Mb/d.

(b) With every slow state at its target, find how many Mb/d the market absorbs for each dollar, the long-run gain and the ratio of the two gains. *Checkpoint:* 0.745 Mb/d per dollar, \$1.34 per Mb/d, a ratio of 9.3.

(c) A quota raised by 1 Mb/d lowers the price by \$8.1 in three months and by \$1.6 after seven years. Use the slowest eigenvalue for a falling price to explain why the seven-year reading still exceeds the long-run gain. *Checkpoint:* −0.142/yr, a time constant of 7.0 yr.

10.6 Theory and calculation · 7 min

Case 4's market has $q_b = 10.5\,\text{Mb/d}$, $q_{max} = 15\,\text{Mb/d}$ and a breakeven that falls from \$80 to \$45. Use Equation (10.10) with the other defaults of Table 10.3.

(a) After the shift, find the long-run demand, the shale and conventional steady states and the residual demand R_{lr} at \$100. *Checkpoint:* 90.42 Mb/d, 15 Mb/d (at the ceiling), 54.46 Mb/d and 11.0 Mb/d.

(b) Find the long-run share R_{lr} over the long-run demand, and compare it with the 15.27 % that case 4 reaches at year 8. *Checkpoint:* 12.1 %, below the year-8 share.

(c) Find R_{lr} at \$75 after the shift, and the barrels a day that holding \$100 instead of \$75 costs the cartel in the long run. *Checkpoint:* 26.5 Mb/d, a loss of 15.5 Mb/d.

10.7 App challenge · 8 min

Load Reference · The balanced market. Before Run, switch to Manual and set Quota request to 31 Mb/d, so the meeting at $t = 0$ applies it. Run eight years at the default seed and export the CSV.

(a) Read the price, column `P`, at 0.25, 1 and 7 years, and the lowest price and its time. *Checkpoint:* \$66.88, \$71.99 and \$73.43, lowest \$66.87 at year 0.27, within \$0.05.

(b) Reload the case, switch to Manual, set Quota request to 29 Mb/d before Run and repeat the run. Read the price at 0.25 and 7 years. *Checkpoint:* \$83.76 and \$76.40, within \$0.05.

(c) From the two runs, give the price change per Mb/d at three months and at seven years for each direction, and explain why the cut moves the price more than the rise. *Checkpoint:* \$8.12 and \$8.76 at three months, \$1.58 and \$1.40 at seven years.

10.8 App challenge · 8 min

Load 4 · The shale revolution: hold \$100 and run eight years at the default seed.

(a) Read the cartel share badge, the quota in force on the quota card, and on the Fringe and demand view the long-run figure in the note of the LEFT FOR THE CARTEL panel. *Checkpoint:* 15.27 %, 13.6 Mb/d and 11.0 Mb/d, the share within 0.05 points.

(b) Set Price target to 80, press Reset and run eight years. Read the same three values. *Checkpoint:* 24.18 %, 23.8 Mb/d and 22.9 Mb/d, the share within 0.05 points.

(c) The long-run shares at \$100 and \$80 are 12.1 % and 23.4 %. Say which run is further from its long run at year 8 and why. *Checkpoint:* The \$100 run, by about 3.2 points against 0.8.

10.9 App challenge · 8 min

Load 7 · The cheater: an actuator that leaks. Run ten years at the default seed and export the CSV.

(a) Find the first row at which the quota, column `q`, reaches its floor of 26 Mb/d, the first row with the price, column `P`, below the \$65 band floor, and the lowest price after that. *Checkpoint:* Year 6.12 (the meeting of year 6.10), year 8.12, and \$61.06 at year 9.25, the times within 0.02 years.

(b) Set Price PI integral time to 5 yr, press Reset and run six years. Find the first row with the price below \$65. *Checkpoint:* Year 5.10, within 0.02 years.

(c) Explain why the shipped loop holds \$75 through four cheaters and then loses the band, and why the slower integral loses it three years sooner. *Checkpoint:* The integral compensates the leak until the quota floor saturates the actuator.

10.10 Application area · 4 min

In case 6, demand falls by 20 Mb/d for three months from year 1, the storage limit is 3.5 Gb, 0.6 Gb above the normal stocks, and the PI on price decides at its twice-yearly meetings with a 1.5 Mb/d cap.

(a) At a net inflow of 20 Mb/d and nothing else changing, how long does the headroom last? The chapter reports the tank full at year 1.115. Give the reason for the difference from the price law and the short-run elasticity. *Checkpoint:* 0.082 yr, about 4.3 weeks, against six weeks.

(b) Which meeting is the first that can cut, and by how much? What does this case say about sizing a buffer and choosing a sample period on a process plant? *Checkpoint:* The meeting of year 1.5, a cut of at most 1.5 Mb/d, after the tank has been full for 20 weeks.

10.11 Application area · 4 min

In case 9 the cartel's PI holds \$90 while the consumers' reserve, with its own PI, releases oil whenever the price is above \$80.

(a) Give the year the reserve runs empty, the price at year 10 with and without the reserve, and what the reserve cost the cartel in discounted revenue. *Checkpoint:* Year 4.33, $86.3 against $86.1, $43 billion (6306 against 6349).

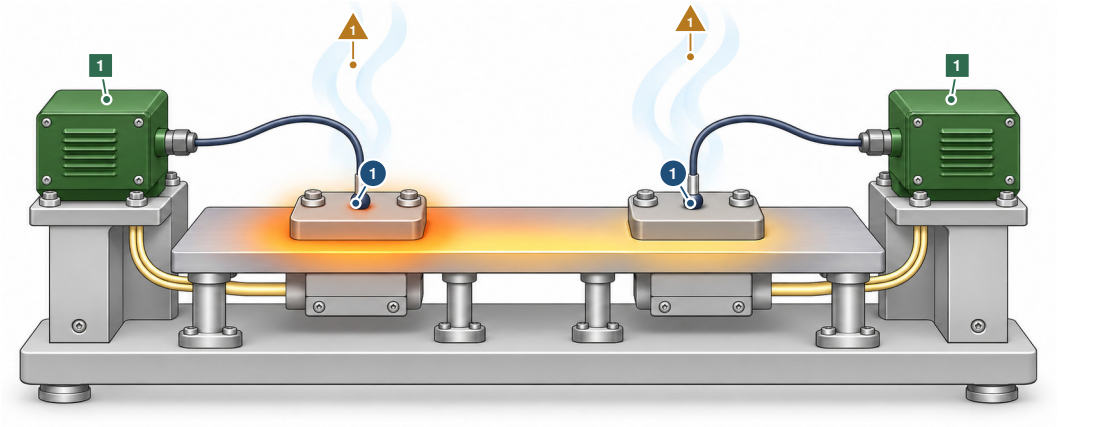
(b) Name two ways a plant avoids two integrating controllers with different setpoints on one inventory, and the one this case offers. *Checkpoint:* One controller per inventory, or a split of the range with a gap between setpoints. Here a reserve ceiling of $95.

10.12 Review the agent · 5 min

An AI assistant reviews the challenge: “The shipped PI fails because it is too sluggish on twice-yearly meetings and never reaches $90. Set Price PI gain to 0.2 and Price PI integral time to 0.2 yr. The price will then reach the $90 target, every barrel will earn more, and both the share and the revenue will clear their limits.” Find the flaw, and say how to check it in the app and with the chapter’s model.

CHAPTER 11

Model Structure Selection and Identifiability on a Temperature Control Lab



The controller moves 1 heater power, set by a stored program, not by feedback

The controller reads 1 temperature sensor beside each heater

What disturbs it 1 heat lost to the room air

Notice that the air carries heat away from both tabs while the plate carries it from one tab to the other: the chapter asks which of these paths one recorded temperature can tell apart. The drive boxes stand for the heater switching that sits on the board. Illustration, not to scale.

IN PLAIN TERMS Which Model Do the Data Support?

The main idea. A small board warms up after a heater turns on. The board, or a computer stand-in for it, records its temperature once a second. Three models of how the board loses heat are fitted to that record, each with more detail than the last.

The challenge. A model with an extra term always fits its own record at least as well. It does so even when the heat loss it adds is not there, and only a second test that the fit never saw can catch this.

Why it matters. Every controller and computer copy of a device built on a model inherits its mistakes, so the choice of model matters.

THREE WAYS TO READ THIS CHAPTER

Explore	11 min	the first page, the history of the digital twin (Section 11.2), the try-it on the lumped groups and the try-it on Best fit (Section 11.4), the two facts in the same section and the key points
Study	55 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	25 min	the first page, the results (Section 11.4), Technology in Practice (Section 11.5), Limitations (Section 11.6) and the open challenge

The control question A heater step test can be explained by more than one energy balance. Which terms of a lumped balance can one record identify, and how does a modeler reject a term that lowers the fitting objective but describes a mechanism the data do not contain?

Learning objectives

- Show that the first-order and convective balances are one two-parameter model, and name the parameter groups a step test identifies.
- Compute the steady-state gain and mean lag of the four-state data generator and use them to judge a fitted model.
- Explain why emissivity and the convective coefficient trade along a flat valley over the temperatures a TCLab reaches.
- Test a nested model by hold-out cross-validation on a second step test and read a structured residual as a missing state.
- Explain why an online moving horizon estimate settles away from the whole-record fit when the model form omits the sensor lag.

Prerequisites Sections 1.4.2 (energy balance), 1.4.4 (first-order response), 1.4.8 (least squares and the simplex search) and 1.4.9 (validation, identifiability and the information criterion), and Chapter 4 for the board and its four-state model. Section 1.6.3 is needed only for the online estimate.

The app https://apmonitor.com/apps/tclab-sim/

See also Volume 5, Chapter 11 for a soft sensor of a temperature that no instrument reaches, Volume 5, Chapter 10 for a soft sensor whose bias a delayed laboratory corrects, Volume 5, Chapter 12 for an observer of an unmeasured stored inventory and Volume 5, Chapter 13 for an extended Kalman filter that applies late measurements.



11.1 Introduction

The Temperature Control Lab (TCLab) is an Arduino shield that heats two transistors and reads two thermistors, published as a hardware benchmark on the argument that a real device imposes obligations a simulation study never meets, among them a sampling interval that has to be met, traffic on a serial link and a model that is never exactly right [1]. Most of what has been written about the device since then concerns control, from undergraduate tuning exercises to feedforward and cascade structures and the passage from a continuous design to a sampled one, and kits of this kind sit among the hands-on resources that control education had taken up by 2018 [2]. The step response that comes before all of that is usually treated as a preliminary, and the model fitted to it is accepted once the curves look close.

Closeness is the wrong test. Two energy balances that differ in physical content can produce prediction curves that lie within the width of a plotted line over one step test, and the objective function will prefer the one with more parameters whether or not the extra physics is present. The question a modeler actually faces is which structure the data can support, and that question has an answer only when a candidate set is written down, fitted under one objective, and then tested somewhere the fit did not see. Model discrimination in this sense is old work in chemical engineering [3] and is a routine step in estimation practice [4], but it rarely reaches an introductory course because it needs several models, several data sets and a way to compare them quickly.

Two pieces of undergraduate theory carry the whole chapter, and Chapter 1 develops both on a heated tank (Sections 1.4.2 and 1.4.4). A lumped-capacitance energy balance equates the rate of change of stored internal energy to the electrical power dissipated in the transistor minus the heat leaving its surface, and the driving force for that loss is a temperature difference for convection and a difference of fourth powers for radiation [5]. The dynamics that follow are first order, so the response carries exactly two independent pieces of information, a steady-state gain and a time constant, and any parameterization with more than two adjustable symbols is redundant by construction [6]. The principles at work are therefore lumped-capacitance modeling, first-order dynamics and step-test identification, structural identifiability and parameter correlation, the choice of objective function,

input design, cross-validation, and the distinction between fitting a record and choosing a mechanism. The industrial warrant is direct. Every gain error in an identified model is inherited by the controller tuned from it, and a structure chosen for the wrong reason is inherited by everything built on top of it.

This chapter works with a browser application in which a TCLab step test is recorded once and then fitted by hand against three nested energy balances under a single sum-of-absolute-errors objective. It measures what the second and third terms of those balances buy on data whose true structure is known. Lumped thermal modeling, absolute-deviation regression and identifiability analysis are all settled subjects. The chapter writes down the recorder, the three model forms, a re-prediction across the entire horizon on every parameter change, the loss breakdown and a bounded simplex search in enough detail to be rebuilt elsewhere. It then reports objective values, fitted parameters and cross-validation results read off the running application, which locate the boundary between a term the data can identify and a term they cannot. The experiments end with a model that fits better and is rejected.

Section 11.2 places the problem between empirical and physics-based modeling. Section 11.3 carries every constant needed to reproduce the application: the data generator, which is the four-state model that Chapter 4 uses as its twin, the three nested balances (Section 11.3.3), the objective and the search (Section 11.3.5), an online moving horizon estimate of the same parameters (Section 11.3.7) and the hardware path. Section 11.4 measures which parameter groups the data identify, what the radiative term does to the fit and to a second step test (Section 11.4.3), how flat the emissivity valley is, how input design and noise bound the comparison, and how the online estimate compares with the regression (Section 11.4.8). Section 11.5 describes the devices and platforms behind such a record, and Sections 11.6 and 11.7 state what the simulator cannot show and what carries over to plant models.

11.2 Background

Process engineering has two traditions of building dynamic models and they answer different questions. The empirical tradition treats the plant as a black box and identifies a transfer function or a difference equation from an input that the engineer designs, which is fast, requires no knowledge of the mechanism, and produces a model that is trustworthy only inside the operating band that the input visited. Four decades of that tradition are surveyed in the system identification literature, along with the recurring difficulty that an identified model is a compromise between what the data contain and what the model set can express [7]. The industrial evidence is that identification, not control design, is where the effort goes, because model development has been reported as the dominant cost of a model predictive control project [8].

> ⧖ HISTORY **From Cybernetics to the Digital Twin, 1948 to 2000**
>
> In 1948 Norbert Wiener defined cybernetics as the scientific study of control and communication in the animal and the machine, a definition that a 2019 review of digital twins links directly to the idea. The term digital twin itself can be traced back only to 2000, when it was mentioned in the context of manufacturing. Several industries had used the idea long before, without the name: online monitoring of process plants, weather forecasting that assimilates measurements into a running model, leak detection in oil and water pipelines, and the remote control and maintenance of satellites and space stations. Data assimilation, the review adds, is in other words the process of fitting models to data, which engineers call filtering and estimation [9].

The physics-based tradition writes conservation equations first and estimates whatever coefficients remain. Its advantage is extrapolation, since a balance that is correct in form stays correct outside the range where its coefficients were measured. Its cost is that the coefficients are often correlated, so the data determine a combination of them rather than each one separately, as Section 1.4.9 shows for the area and level of a tank. Identifiability in this practical sense, meaning whether a particular data set can separate the parameters of a particular model, is normally assessed through the sensitivity

Table 11.1 Recorded and Derived Signals

Symbol	Role	Units	Range
T_1, T_2	measured output, fitted	°C	unbounded
Q_1, Q_2	input, programmed or manual	%	0 to 100
$\hat{T}_1$, $\hat{T}_2$	model prediction	°C	unbounded
T_a	model parameter and generator setting	°C	10 to 40
T_{H1}, T_{H2}	generator state, unmeasured	°C	unbounded
T_{C1}, T_{C2}	generator state, measured	°C	unbounded
J	objective, absolute deviation	°C	≥ 0
T_{cut}	hardware cutoff	°C	50 to 90

matrix and its collinearity, and parameter subsets are chosen so that the remaining directions are distinguishable [10]. The complementary activity is to design the experiment so that the awkward directions become observable, which is the subject of model-based experimental design [11].

The objective function is a modeling decision of the same weight as the model structure. A sum of squared errors, Equation (1.19), is the maximum-likelihood choice for Gaussian noise and gives smooth derivatives, while a sum of absolute errors is far less sensitive to an isolated bad measurement, which is the classical argument for robust estimation [12]. The absolute-deviation form is available as a model option in nonlinear estimation packages and is often preferred for dynamic data reconciliation on plant measurements that contain outliers [13]. The application of this chapter uses the absolute-deviation form throughout, partly for that robustness and partly because a sum of absolute errors in degrees Celsius is a quantity a student can read and reason about without converting units in their head.

Selecting among candidate structures closes the loop. When two models are nested, the larger one cannot fit worse, so the comparison must be made against a penalty for the extra parameter or against data the fit did not see. Information criteria such as Equation (1.23) supply the first route [14], and cross-validation on a second record, as in Section 1.4.9, supplies the second. The measurements in Section 11.4 take the second route, because a step test at a different heater level is available for free and because the failure it exposes is a physical one rather than a statistical score.

11.3 Process Models and Data Generator

The physical board holds a pair of TIP31C power transistors and a pair of thermistors, one of each at either end of a small shield. Electrical power enters each transistor as a duty cycle between 0 and 100%, heat leaves both of them to the room, and the board itself carries heat from one end to the other. No thermistor is bonded to the case of the transistor it watches, so heat reaches the reported temperature only after crossing a thermal resistance. Figure 11.1 draws the arrangement and carries every symbol used in the balances below.

11.3.1 Signal Classification

The application records two temperatures and two heater commands once per second. Both temperatures are outputs to be matched, and only the first enters the objective unless the second heater is switched into it. Both heater commands are inputs supplied by a stored program or by the operator, and neither is computed from a measurement, because there is no feedback anywhere in the application. The ambient temperature is a parameter of the fitted model and also a setting of the data generator, and those two roles are separate. Table 11.1 lists the classification.

11.3.2 The Data Generator

Measurements come either from a physical board or from a four-state simulator that stands in for it. The simulator is deliberately not one of the three models being fitted, so no parameter set can recover

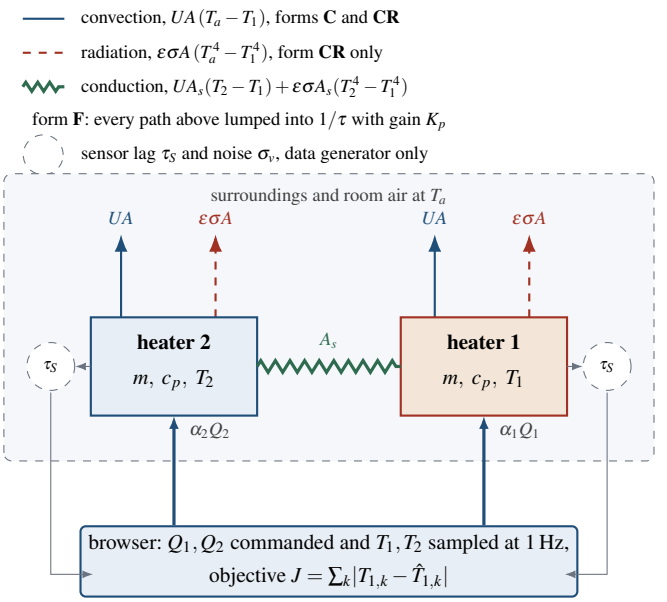


Figure 11.1 Two heaters on the Temperature Control Lab shield and the three loss paths that the nested energy balances resolve one at a time. The first-order form **F** collapses all of them into a single time constant, the convective form **C** resolves the solid convection path, and the convective plus radiative form **CR** adds the dashed radiation path. Conduction across the board is drawn as a zigzag and the data generator's sensor lags as dashed circles.

the structure that generated the data. It is the twin of Chapter 4 at its default settings. It carries a separate state for each transistor and for each thermistor, as given by Equations (11.1) to (11.4).

$$\frac{dT_{H1}}{dt} = \frac{P_1 Q_1}{C_H} + \frac{T_a - T_{H1}}{\tau_{HA}} - \frac{T_{H1} - T_{H2}}{\tau_{HH}} \tag{11.1}$$

$$\frac{dT_{H2}}{dt} = \frac{P_2 Q_2}{C_H} + \frac{T_a - T_{H2}}{\tau_{HA}} + \frac{T_{H1} - T_{H2}}{\tau_{HH}} \tag{11.2}$$

$$\frac{dT_{C1}}{dt} = \frac{T_{H1} - T_{C1}}{\tau_S}, \qquad \frac{dT_{C2}}{dt} = \frac{T_{H2} - T_{C2}}{\tau_S} \tag{11.3}$$

$$T_i = T_{Ci} + \sigma_v z, \qquad z \sim \mathcal{N}(0,1), \qquad i = 1,2 \tag{11.4}$$

where T_{H1} and T_{H2} hold the two transistor temperatures, T_{C1} and T_{C2} hold the two thermistor temperatures, P_1 and P_2 scale the available heater power and carry no units, C_H turns that power into a rate of temperature change and is fixed at 5720 % s/°C, τ_{HA} governs the loss to the room, τ_{HH} governs conduction from one transistor to the other, τ_S is the lag of the thermistor behind its transistor and σ_v sets the width of the measurement noise. The four states are advanced together with explicit Euler steps of 0.1 s and sampled once per simulated second. A speed control from 1× to 10× takes more substeps per timer tick and leaves the sampled record unchanged, so a ten-minute experiment takes one minute, which no board allows. As in the framework apps of Section 1.6.1, the speed does not change what a run computes. Unlike them, this standalone app draws its noise from the browser's unseeded generator, so two noisy runs differ.

Three properties of this generator matter for everything that follows. Its loss to the room is linear in temperature difference, so radiation is provably absent from the data. Its thermistor lag of 140 s is a state that none of the three fitted models contains, so no parameter set can drive the objective

to zero. Its steady-state gain from Q_1 to T_1 is 0.599 °C/% at every heater level, obtained by solving Equations (11.1) and (11.2) at steady state with the default parameters, and the mean lag of the same response is 157.6 s, obtained as the sum of the thermistor lag and the 17.6 s first moment of the transistor pair. Those two numbers are the ground truth against which every fit below is judged.

11.3.3 Three Nested Energy Balances

The first form is the one an introductory course writes down, a single lumped capacity relaxing toward ambient with a linear gain on heater power, the first-order model of Section 1.4.4, as shown in Equation (11.5).

$$\tau \frac{\mathrm{d}\hat{T}_1}{\mathrm{d}t} = (T_a - \hat{T}_1) + K_p Q_1 \tag{11.5}$$

where τ is the thermal time constant in seconds, K_p is the steady-state gain in °C/% and T_a is the ambient temperature. Every transport mechanism is inside τ, and nothing in the equation says how the heat leaves.

The second form is an energy balance on a lumped mass whose only loss is convection to the room, as shown in Equation (11.6).

$$mc_p \frac{\mathrm{d}\hat{T}_1}{\mathrm{d}t} = UA\,(T_a - \hat{T}_1) + \alpha_1 Q_1 \tag{11.6}$$

where m is the heated mass in kg, c_p is its specific heat capacity in J/(kg K), U is the overall heat transfer coefficient in W/(m^2 K), A is the exposed surface area in m^2 and α_1 is the electrical power delivered per percent of heater command in W/%. The lumped-capacity assumption behind this form is the standard one, valid when internal conduction is fast compared with the surface resistance [15], and the coefficient U stands for natural convection from a small package in still air [16].

Equation (11.6) is algebraically the same model as Equation (11.5) with $\tau = mc_p/(UA)$ and $K_p = \alpha_1/(UA)$, and the application reports both of those groups as derived quantities so the identity stays visible while the parameters are adjusted. What changes is not the dynamics but the bookkeeping, because the four separate symbols invite a student to give each one a physical value and then discover that only two combinations of them matter.

The third form adds radiation exchange with the surroundings through the Stefan-Boltzmann law, as shown in Equation (11.7).

$$mc_p \frac{\mathrm{d}\hat{T}_1}{\mathrm{d}t} = UA\,(T_a - \hat{T}_1) + \varepsilon\sigma A\,(T_a^4 - \hat{T}_1^4) + \alpha_1 Q_1 \tag{11.7}$$

where ε is the surface emissivity, σ is the Stefan-Boltzmann constant of 5.67×10^{-8} W/(m^2 K^4) and the fourth-power terms are evaluated with absolute temperatures in kelvin. Setting $\varepsilon = 0$ recovers Equation (11.6) exactly, so the three forms are strictly nested, and the reported objective at $\varepsilon = 0$ is a direct check of that nesting. The radiative term is the mechanism usually offered for the observation that the measured steady-state gain of a TCLab falls at elevated temperature [1].

11.3.4 The Two-Heater Extension

An option adds the second heater and the heat transferred between the two heaters, and adds $\sum_k |T_{2,k} - \hat{T}_{2,k}|$ to the objective. For the convective plus radiative form the extended pair is given by Equations (11.8) and (11.9).

$$mc_p \frac{\mathrm{d}\hat{T}_1}{\mathrm{d}t} = UA\,(T_a - \hat{T}_1) + \varepsilon\sigma A\,(T_a^4 - \hat{T}_1^4) + UA_s(\hat{T}_2 - \hat{T}_1) + \varepsilon\sigma A_s(\hat{T}_2^4 - \hat{T}_1^4) + \alpha_1 Q_1 \tag{11.8}$$

$$mc_p \frac{\mathrm{d}\hat{T}_2}{\mathrm{d}t} = UA\,(T_a - \hat{T}_2) + \varepsilon\sigma A\,(T_a^4 - \hat{T}_2^4) - UA_s(\hat{T}_2 - \hat{T}_1) - \varepsilon\sigma A_s(\hat{T}_2^4 - \hat{T}_1^4) + \alpha_2 Q_2 \tag{11.9}$$

where A_s is the area that couples the two heaters and α_2 is the power factor of the second heater. The coupling terms enter the two balances with opposite signs, so heat moved between the heaters is conserved. The two heaters share U, A, m and c_p, which is a structural constraint rather than a measurement, and Section 11.4 shows what that constraint costs. The convective form drops the two fourth-power terms, and the first-order form couples the two temperatures through a single time constant τ_c instead, with $\tau(\hat{T}_2 - \hat{T}_1)/\tau_c$ added to the first balance and subtracted from the second, and with a separate gain K_{p2} on the second heater.

11.3.5 Objective Function and Numerical Evaluation

The objective is the sum of absolute errors between the measurement and the prediction at every recorded sample, the objective of Equation (1.19) with absolute values in place of squares, as shown in Equation (11.10).

$$J = \sum_{k=1}^{N} |T_{1,k} - \hat{T}_{1,k}| + \delta \sum_{k=1}^{N} |T_{2,k} - \hat{T}_{2,k}| \tag{11.10}$$

where N is the number of recorded samples and δ is one when the second heater is included in the objective and zero otherwise. The objective is reported in degrees Celsius together with the mean absolute error J/N, so a run of 303 samples with an average deviation of 0.47 °C reads as 141.9 °C. No weighting, no normalization and no penalty on the parameters appear anywhere in Equation (11.10).

The prediction is regenerated over the entire recorded horizon whenever any parameter changes, rather than only from the point of the change onward. Each one-second measurement interval is integrated with the classical fourth-order Runge-Kutta method subdivided so that no step exceeds 0.5 s, giving two substeps per interval for the reported prediction and one substep inside the search, where the coarser step halves the cost of an objective evaluation. The heater command is held at the value recorded at the start of the interval, which matches how the firmware and the simulator apply a new setting. The initial condition is the first measured sample rather than the ambient parameter, so a noisy first reading displaces the whole predicted trajectory.

An optional bounded simplex search, the Best fit button, is available as a reference answer. It runs the Nelder-Mead method of Section 1.4.8 with reflection, expansion, contraction and shrink coefficients of 1, 2, 0.5 and 0.5 on parameters rescaled to the unit interval spanned by their adjustable range, which keeps quantities as differently scaled as A at 1.2×10^{-3} m^2 and c_p at 500 J/(kg K) comparable inside one simplex. The search stops at 2500 objective evaluations, at a simplex smaller than 10^{-6}, or when the best and worst vertices agree to 10^{-9}, and it is restarted twice with initial spreads of 0.12, 0.06 and 0.04. It moves two parameters for a single-heater fit, $\{\tau, K_p\}$ for the first-order form and $\{U, \alpha_1\}$ for the other two, and four when the second heater is included. Geometry, material properties, emissivity and ambient temperature are deliberately held, and any fitted parameter that finishes within 10^{-4} of the end of its range is reported as a likely sign that the run cannot separate it from another parameter.

Table 11.2 Parameters of the Fitted Models and of the Data Generator

Symbol	Meaning	Default	Range	Status
fitted model, first-order form ***F***				
τ	time constant, s	120	10 to 600	fitted
K_p	gain, °C/%	0.8	0 to 3	fitted
K_{p2}	gain 2, °C/%	0.6	0 to 3	T_2 only
τ_c	coupling time, s	1000	60 to 4000	T_2 only
fitted model, forms ***C*** *and* ***CR***				
U	coefficient, W/m^{2}K	10, 5	1 to 40	fitted
A	area, m^2	0.0012	0.0002 to 0.006	held
A_s	coupled area, m^2	0.0002	0 to 0.002	T_2 only
m	mass, kg	0.004	0.0005 to 0.02	held
c_p	heat capacity, J/kg K	500	100 to 1500	held
α_1	power factor, W/%	0.01	0 to 0.05	fitted
α_2	power factor 2, W/%	0.0075	0 to 0.05	T_2 only
ε	emissivity	0.9	0 to 1	held
T_a	ambient, °C	23	10 to 40	seeded
σ	S.-B. constant, W/m^{2}K^4	5.67×10^{-8}	—	hidden
data generator				
T_a	ambient, °C	21	0 to 45	setting
P_1	heater 1 capacity	200	50 to 255	setting
P_2	heater 2 capacity	100	25 to 255	setting
τ_{HA}	loss to the room, s	20	5 to 100	setting
τ_{HH}	heater to heater, s	100	20 to 300	setting
τ_S	thermistor lag, s	140	20 to 300	setting
σ_v	sensor noise, °C	0.2	0 to 1	setting
C_H	capacity constant, % s/°C	5720	—	hidden
h	integration step, s	0.1	—	hidden

11.3.6 Parameters, Ranges and Visibility

Table 11.2 lists every parameter of the fitted models and of the data generator, with the value each takes when the application opens. All thirteen model parameters are operator adjustable, and only those that appear in the selected balance are exposed. Each form keeps its own copy of the whole parameter record, so switching forms never discards a fit in progress.

The default U is 10 W/(m^2 K) for the first-order and convective forms and 5 W/(m^2 K) for the radiative form, which are the published starting values of the two course assignments the forms follow and already encode the expectation that radiation will carry about half the loss. Ambient temperature is not fitted. At the start of every simulator run the first reading is rounded to 0.1 °C and written into T_a for all three forms, because the board begins at room temperature and a direct reading of a measurable quantity is preferable to an estimate that can trade against the heater gain.

11.3.7 Online Moving Horizon Estimate

The simplex search answers once, after the record is complete. A second view, Online estimate, runs the same fit while the record is still arriving, as a moving horizon estimate of exactly the parameters the search moves [17], and draws each of them against time beside the search's answer. Section 1.6.3 describes the estimator options of the framework apps in general. This one is specific to the fit. It never writes a parameter, so the regression remains the reference. Each estimated parameter enters the state in the unit coordinate of its adjustable range, as shown in Equation (11.11).

$$x_k = \left[\hat{T}_{1,k},\ (\hat{T}_{2,k}),\ \tilde{\theta}_k^{\mathsf{T}}\right]^{\mathsf{T}}, \qquad \tilde{\theta}_i = \frac{\theta_i - \theta_i^{\min}}{\theta_i^{\max} - \theta_i^{\min}} \tag{11.11}$$

where θ holds $\{\tau, K_p\}$ for the first-order form or $\{U, \alpha_1\}$ for the other two, with $\{K_{p2}, \tau_c\}$ or $\{\alpha_2, A_s\}$ added when T_2 is included, and $\theta^{\min}$ and $\theta^{\max}$ are the ranges of Table 11.2. Those ranges become the hard bounds $0 \le \tilde{\theta} \le 1$, the box the search itself uses, while the temperatures are unbounded and A, m, c_p, ε and T_a stay at their set values. Each parameter is a random walk. The temperature rows of the transition f integrate one measured interval of the selected balance with the two Runge-Kutta substeps of the displayed prediction and the heater command held, and f is linearized once per step at the previous estimate $\hat{x}_{k-1}$ by central differences of 10^{-5}, which gives $A_k = \partial f/\partial x$ and $c_k = f(\hat{x}_{k-1}) - A_k \hat{x}_{k-1}$. At the linearization point the affine step reproduces the integrator to 1.4×10^{-14} °C. At every sample the estimator solves Equation (11.12) over the last $L = 30$ intervals, $s = \max(0, k - L)$.

$$\begin{aligned} \min\ & \|x_s - \bar{x}_s\|^2_{\bar{P}_s^{-1}} + \sum_{j=s}^{k-1} \|w_j\|^2_{W^{-1}} + \sum_{j=s}^{k} \frac{\|y_j - C x_j\|^2}{\sigma_y^2} \\ \text{s.t. } & x_{j+1} = A_{j+1} x_j + c_{j+1} + G w_j, \qquad 0 \le \tilde{\theta}_j \le 1 \end{aligned} \tag{11.12}$$

In it y_j holds the recorded T_1 (and T_2), C selects the modeled temperatures and G drives the parameters alone, so the estimate is an output-error fit like the search. The weights are $\sigma_y = 0.2\,^{\circ}\mathrm{C}$ and $W = (1/6)^2/3600$ per normalized parameter per step. The arrival term anchors the window at $\bar{x}_s$, the one-step prediction from the estimate made at $s-1$, weighted by a covariance $\bar{P}_s$ that a Kalman recursion propagates through each stored linearization. Both start at the ambient temperature with a spread of 2 °C and at the published parameter values with a sixth of each range. These weights were fixed before any estimate was computed. The value of σ_y is the generator's default noise held as a constant of the estimator, never read from the noise setting, and the walk adds the prior's variance again over 3600 steps, the longest run the application records. The window of 30 was the longest of 10, 30 and 60 whose 99th percentile cost per update on the costliest form stayed under 10 ms, a tenth of the sample period at 10×, which was also a rule set beforehand. Equation (11.12) is a dense quadratic program of 61 decisions with one heater and 122 with two, solved by a dual active-set method. It departs from the search in two ways, because it sums squared rather than absolute errors and because its parameters may drift along the record while the search holds one set over the whole of it.

While no bound is active, Equation (11.12) with each step's model fixed on entry has the same solution as the extended Kalman filter on the same linearization, for every window length [17]. A separately written filter matched the unbounded window to 9.6×10^{-9} of the range over every update examined, so the unbounded window serves as the Kalman filter in Section 11.4.8, and the window length changes the answer only while a bound is active. A window the solver rejects is replaced by a Kalman update clipped to the bounds and counted, and no measured run rejected one. The estimate consumes the recorded samples in order and nothing else, so it is causal and identical at 1× and 10×, and a record finished before the view was opened is replayed from its first sample. It reads the recorded temperatures and heater commands, the held parameters, the published values and the ranges, and seven altered copies of the source that route the generator's settings or state into it each fail the application's tests.

11.3.8 Absence of Feedback and the Overtemperature Comparator

Nothing in the application computes an actuator value from a measurement. There is no setpoint, no automatic mode, no proportional-integral-derivative algorithm and no tuning. Heater commands come from one of four stored programs or from two operator-set inputs, and the stored programs are piecewise constant schedules evaluated as functions of elapsed time alone.

One automatic action does exist and applies only to a physical board. An overtemperature limit, 70 °C by default and settable anywhere between 50 °C and 90 °C while nothing is plugged in, stops the run the moment either board reading reaches it, zeroes both heater commands, writes `X` to the

firmware and latches the alarm. That is a comparator with a latch rather than a controller, and it has no dynamics of its own.

11.3.9 Assumptions

The three fitted forms rest on six assumptions, each of which fails somewhere.

The heated object is isothermal. One temperature stands for the silicon die, the package and the board copper attached to it, which is a poor description in the first seconds of a step while the die is still running well ahead of everything bonded to it.

The measured temperature equals the modeled temperature. None of the three forms contains a sensor state, although the data generator does and a real thermistor does, so the whole of that lag is pushed into the fitted parameters. This single assumption accounts for most of the residual reported in Section 11.4.

The heat transfer coefficient is constant. Natural convection from a small package has a coefficient that grows with the temperature difference, so a single U fitted over a 26 °C rise is an average over that rise [16].

The surroundings radiate as a black body at the air temperature. The room air temperature and the temperature of the surfaces the transistor exchanges radiation with are treated as the same number, and the view factor is folded into ε.

Heater power is proportional to duty cycle. The electrical power a pulse-width-modulated transistor actually dissipates depends on its own junction temperature, so the true relation bends slightly as the device heats up.

The measurement noise is white, Gaussian and additive. A ten-bit converter reading a thermistor delivers quantized noise instead, although at $\sigma_v = 0.2\,°\mathrm{C}$ the difference is small.

11.3.10 Hardware Interface

The hardware path uses Web Serial directly from the page. The transport requests a port, opens it at 115 200 baud and falls back to 9600 baud, waits 1500 ms for the board to finish its reset, writes `VER` and accepts the connection only if a reply line matches the TCLab firmware pattern. Three commands follow, `Q1 0`, `Q2 0` and `X`, which guarantee that the run begins with both heaters off. Each second of a run it reads `T1` and `T2`, then writes the heater command that holds until the next reading, with a 1800 ms timeout on every reply and all operations serialized through a promise chain. Cable removal, a closed stream, a command timeout, the page being hidden and the page being unloaded each trigger a best-effort `Q1 0`, `Q2 0` and `X`, and the watchdog compiled into the firmware backstops all of them. Hardware always runs at 1×, because a board cannot be accelerated. The measurements reported in this chapter were taken from the simulator, since the hardware path could not be exercised in the environment used for the experiments.

11.4 Results

Five recorded runs support the measurements below and are listed in Table 11.3. Every run was taken from the built-in simulator at 10× speed with the default generator parameters, and the sensor noise was set to zero except where stated, so that each objective value is reproducible to the digit displayed. No thermistor allows zero noise, but a noise-free run shows the structural residual on its own and gives the reproducible objective values that a comparison between models needs. Objective values and fitted parameters were read from the application's own readouts, and the trajectories plotted in Figure 11.3 were regenerated from Equations (11.6) and (11.7) using the exported parameter header, which reproduces the displayed objective to 0.01 °C.

The application organizes its experiments as thirteen scenarios, numbered and named as on its Instructions page and its progress badge: 1, Record and fit the first-order form; 2, Prove the lumped groups; 3, Add radiation; 4, Check the $\varepsilon = 0$ collapse; 5, Run Best fit, then distrust it; 6, Find the identifiability valley; 7, Fix it with a better experiment; 8, Read the residual's shape; 9, Two heaters; 10, Use the 100% step; 11, Fit real hardware; 12, Export and re-fit elsewhere; and 13, Watch the fit

Table 11.3 Recorded Runs

Label	Heater program	Length	Samples
E1	Q_1 0 to 50% at 2 s	302 s	303
E2	Q_1 0 to 100% at 2 s	302 s	303
E3	manual, Q_1 80% at 2 s, 30% at 300 s	602 s	603
E4	doublet, Q_1 80, 30, 60% with Q_2 40%	600 s	601
E5	Q_1 100% at 10 s, Q_2 100% at 300 s	600 s	601

converge online. Scenarios 11 and 13 are optional, so eleven are required. A scenario is credited for a program recorded to its end, the stated model form or parameter change, and a confirmation where the reading cannot be inferred from the record. Selecting a scenario alone earns nothing. Two credits are weaker than their names. Scenario 3 records that the predicted loss split was read, which cannot establish radiation in a simulator whose balance contains none. Scenario 4 records a confirmation without checking that the other parameters match the convective form, and the collapse to Equation (11.6) is exact only when they do. The records show that an activity was done, not what was learned.

11.4.1 Structural Identifiability of the Convective Form

The convective form carries six symbols and the data can move only two combinations of them. Doubling A from 0.0012 m^2 to 0.0024 m^2 while halving U from 8.46 to 4.23 $\mathrm{W/(m^2\,K)}$ leaves the objective on E1 at 141.9 °C, unchanged in every displayed digit, and doubling m from 0.004 to 0.008 kg while halving c_p from 500 to 250 J/(kg K) does the same. Only the products UA and mc_p are identifiable, exactly as the derived readouts $\tau = mc_p/(UA)$ and $K_p = \alpha_1/(UA)$ predict. Doubling the fitted α_1 alone instead raises the objective from 141.9 to 5039.9 °C, which shows that the power factor is not part of either group.

> **TRY IT IN THE APP** **2 · Prove the Lumped Groups**
>
> Set the sensor noise to 0 in Settings and the speed to 10×, then record the Q1 = 50% · 5 min test to its end. Choose the Convective form, double U and halve A from their published values, and watch the objective and the Implied τ and Implied gain readouts. None of them moves, because only the products UA and mc_p enter the prediction.

The same identity appears between the two smallest model forms. Fitting the first-order form to E1 returns $\tau = 197$ s and $K_p = 0.676\,^\circ\mathrm{C}/\%$ with $J = 141.9\,^\circ\mathrm{C}$, and fitting the convective form to the same run returns $U = 8.46\,\mathrm{W/(m^2\,K)}$ and $\alpha_1 = 0.006\,86\,\mathrm{W}/\%$, whose derived groups are 197 s and 0.676 °C/% with the identical objective. Two independent simplex searches over different symbols reach the same model, which is the cleanest demonstration in the application that renaming parameters is not the same as adding information.

Pause. Try Exercise 11.1 before reading on.

11.4.2 Effect of the Radiative Term on the Fitted Run

Adding the Stefan-Boltzmann term reduces the objective on E1 from 141.9 to 134.2 °C, an improvement of 5.4%, and the maximum single-sample error falls from 1.67 to 1.63 °C. The radiative form introduces a nonlinear loss mechanism, but its automatic search still moves only two coefficients because emissivity is held fixed. The smaller residual therefore does not establish that radiation generated the data, which contain only linear loss. The parameters that deliver the improvement carry the argument. The heat transfer coefficient falls from 8.46 to 2.46 $\mathrm{W/(m^2\,K)}$, the derived time constant grows from 197 to 679 s, and the loss breakdown reports that 70.1% of the energy leaving heater 1 over the run went out by radiation. The truth is 0%.

Table 11.4 Objective After Fitting, by Model Form and Recorded Run

Run	Published **C**	Fitted **F**	Fitted **C**	Fitted **CR**
E1	1876.4	141.9	141.9	134.2
E2	3752.8	283.8	283.8	250.2
E3	–	567.1	567.1	658.3
E4	–	594.0	594.0	653.9
E5	–	963.2	963.2	1211.2

All entries are J in °C. E5 includes T_2 in the objective, and the other four do not.

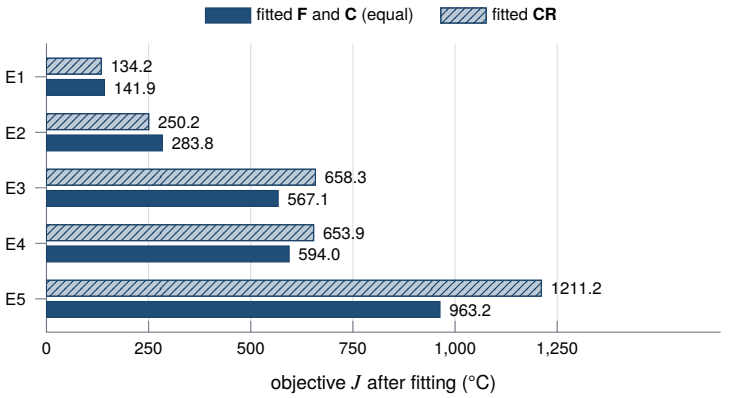


Figure 11.2 The objective after fitting, by model form and recorded run, from Table 11.4. Solid bars are the first-order and convective forms, which reach the same value on every run, and hatched bars the convective plus radiative form. Notice that the radiative term wins only on the two single steps E1 and E2 and loses on the two-level input E3, the doublet E4 and the two-heater run E5.

Results for all five runs are collected in Table 11.4. The radiative form improves both step-test fits, by 5.4% on E1 and 11.8% on E2. It loses by 16.1% on the two-level input E3, by 10.1% on the doublet E4, and by 25.7% on the two-heater run E5, where the search drives U onto its lower bound of 1 W/(m^2 K) and the application flags the parameter as unidentifiable in its status line. Figure 11.2 draws the same results as bars.

> **TRY IT IN THE APP** **5 · Run Best Fit, Then Distrust It**
>
> With the sensor noise at 0, record the Two-heater · 10 min test, tick Include T2 and choose the Convective + radiative form. Press Best fit and read the status line. The objective ends at 1211.2 °C, worse than the convective 963.2 °C, and U finishes on its lower bound of 1 W/(m^2 K), which the status line reports as a parameter the run cannot separate.

11.4.3 Cross-Validation Between the Two Step Tests

The clearest evidence sits in the pair of step tests at 50% and 100% heater power. Both models were fitted to E1 and then applied without change to E2, and results are given in Table 11.5. The convective parameters transfer perfectly, in the sense that a fresh search on E2 cannot improve on them and reports so. The objective doubles from 141.9 to 283.8 °C only because the temperature rise doubles, from 25.95 °C to 51.90 °C at the end of the run, and the measured ratio of 2.000 is itself the signature of a linear loss. The radiative parameters do not transfer. Applied to E2 they give 429.0 °C against 283.8 °C for the convective set, which is 51% worse, and the residual in Figure 11.3 shows why. The fitted radiative model runs too slowly and stays 1.49 °C below the measurement for the whole second

Table 11.5 Cross-Validation Between the Two Step Tests

Parameter set	J on E1	J on E2	$\hat{T}_1$ at $Q_1 = 100\%$
C fitted on E1	141.9	283.8	88.56
CR fitted on E1	134.2	429.0	80.48
CR fitted on E2	209.7	250.2	84.47
data generator	–	–	80.94

J in °C, steady state in °C. The convective set fitted on E2 is the set fitted on E1, because the search on E2 could not improve on it.

half of the run, while the convective residual stays inside 0.92 °C over the same window. Fitting on one step and evaluating on an untouched second step is hold-out cross-validation, and it is what makes the degradation from 283.8 to 429.0 °C a measurement rather than an anecdote.

🕮 **Pause.** Try Exercise 11.2 before reading on.

The steady-state column of Table 11.5 carries the opposite verdict. Extrapolated to a full-power steady state, the convective fit predicts 88.56 °C against the generator's true 80.94 °C, an error of 7.62 °C, while the radiative fit predicts 80.48 °C, an error of 0.46 °C. The mechanism is visible in the chord gain of the fitted radiative model, which falls from 0.655 to 0.595 °C/% between $Q_1 = 50\%$ and $Q_1 = 100\%$ while the generator holds 0.599 °C/% at both. A falling curve crosses a constant one somewhere, and E2 happens to be near the crossing. The radiative term therefore improves an extrapolation that the data never constrained and worsens a transient that they did, which is the practical shape of overfitting on a physical model.

✦ **DID YOU KNOW? Better on Its Own Record, Worse on the Next**

From the simulator: the radiative form fitted to the 50% step lowers the objective from 141.9 to 134.2 °C. Carried unchanged to the 100% step, it scores 429.0 °C against 283.8 °C for the convective set, 51% worse (Table 11.5).

11.4.4 Emissivity Scan

Emissivity is excluded from the search on purpose, and a scan explains the decision. Setting ε by hand and refitting U and α_1 at each value gives Table 11.6. Across the whole admissible range of ε from 0 to 1 the objective moves from 141.9 to 133.4 °C, a total of 6.0%, while the fitted heat transfer coefficient sweeps a factor of 4.7 from 8.46 down to 1.79 W/(m^2 K) and the derived time constant sweeps the same factor from 197 to 931 s. A student who reads only the objective sees a monotone and almost featureless slope. The reported loss breakdown at the same time swings from 100% convective to 78.1% radiative, so the application reports a completely different account of the physics for a 6% change in fit quality. The entry at $\varepsilon = 0$ reproducing the convective objective of 141.9 °C exactly is the numerical check that the two forms are nested.

The physical reason the valley is flat is the temperature range. Over a rise from 21 to 47 °C the radiative driving force $T_a^4 - T_1^4$ is close enough to linear in $T_a - T_1$ that the two loss terms are nearly proportional, so a reduction in U can be traded against an increase in ε almost exactly. Doubling the step to E2 barely helps, because at the published values the instantaneous radiative share at the end of the run moves only from 55.1% to 58.6% when the final temperature goes from 53.7 to 83.1 °C. Separating the two mechanisms needs a temperature range that a TCLab is not permitted to reach, which is a statement about the experiment rather than about the physics.

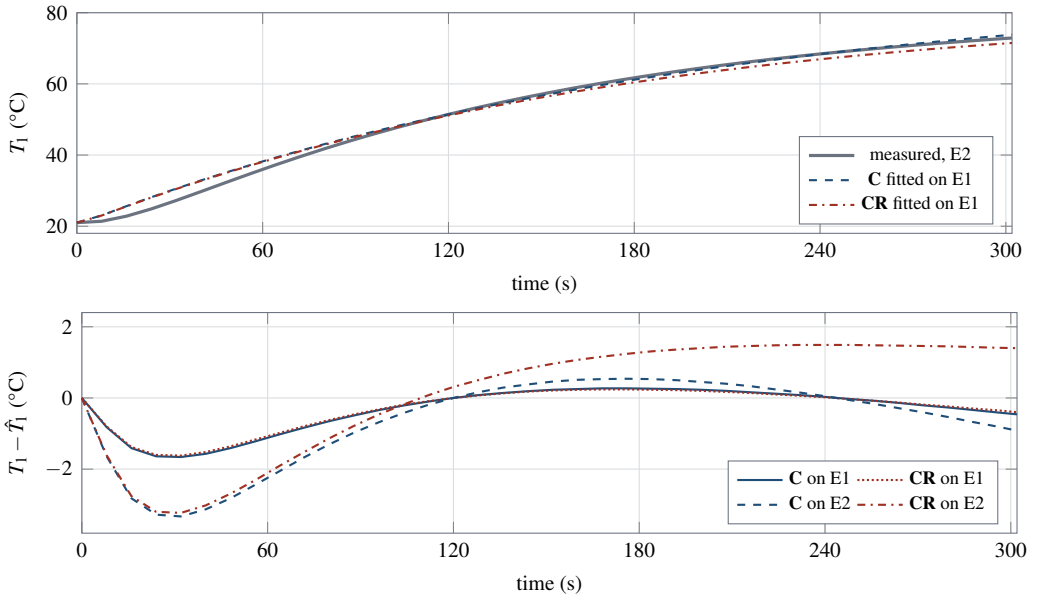


Figure 11.3 A fit made on the 50% step test carried across to the 100% step test. The upper axis plots the measured E2 response (thick gray) against the two predictions, which were fitted to E1 and then left alone: convective dashed, radiative dash-dotted. The lower axis plots the signed residual of both models on both runs: on E1 convective solid and radiative dotted, on E2 convective dashed and radiative dash-dotted. The persistent 1.49 °C bias of the radiative form on E2 is the excess that Table 11.5 counts.

Table 11.6 Emissivity Scan on Run E1, Refitting U and α_1

ε	U (W/(m^2 K))	$mc_p/(UA)$ (s)	radiative share (%)	J (°C)
0.0	8.46	197	0.0	141.9
0.2	7.12	234	15.2	140.2
0.4	5.78	288	30.6	138.5
0.6	4.45	374	46.2	136.8
0.8	3.12	534	62.1	135.1
0.9	2.46	679	70.1	134.2
1.0	1.79	931	78.1	133.4

> ✦ **DID YOU KNOW?** **A Flat Valley With Two Different Stories**
>
> From the simulator: sweeping the emissivity from 0 to 1 on E1 moves the objective by only 6.0%, while the fitted heat transfer coefficient falls by a factor of 4.7 and the reported radiative share climbs from 0.0 to 78.1% (Table 11.6).

11.4.5 Ambient Temperature and Input Design

Ambient temperature is the other parameter held out of the search, and the reason is the same trade in a different pair. Scanning T_a by hand and refitting gives Table 11.7. On the single step E1 a 6 °C error in T_a costs 2.3% of objective while moving the fitted power factor by 18%, so a search allowed to move T_a travels a long way for very little reduction in the objective. On the two-level input E3 the same 6 °C window costs 63% of objective, which is the sensitivity that makes the parameter

Table 11.7 Ambient Temperature Held at a Wrong Value, Convective Form Refitted

T_a	run E1, one heater level		run E3, two heater levels	
(°C)	α_1 (W/%)	J (°C)	α_1 (W/%)	J (°C)
19	0.0073	140.9	0.0077	476.7
21	0.0069	141.9	0.0075	567.1
25	0.0060	144.1	0.0072	778.2

identifiable. An input that visits two heater levels supplies two independent relations between T_a and α_1, while an input that holds one level supplies only the combination $UA\,T_a + \alpha_1 Q_1$.

Sensitivity is not the same as accuracy. On E3 the objective keeps falling as T_a is reduced, reaching 354.6 °C at 15 °C, which is below any plausible room temperature. The minimum of the objective over T_a sits outside the physical range because the fitted forms have no sensor state, and the search buys a better match to the lagged measurement by pretending the room is cold. Seeding T_a from the first reading of the run avoids the whole question, and it costs nothing because the board always starts at room temperature.

The two-level input also improves the fitted gain. Fitting the convective form to E1 returns a gain of 0.676 °C/% against the generator's true 0.599 °C/%, an error of 12.8%, because a five-minute step reaches only 86.6% of the asymptotic rise and the search has to guess the rest. Fitting the same form to E3 returns 0.617 °C/%, an error of 3.0%, and the doublet E4 returns 0.619 °C/%. The identified time constant improves in the same way, from 197 s on E1 to 164 s on E3 against the generator's 157.6 s mean lag.

11.4.6 The Noise Floor

Every number above was taken with the sensor noise disabled. Four further recordings use the default 0.2 °C noise and independent, reproducible pseudorandom streams. Their convective objectives are 134.6, 159.8, 161.1 and 137.4 °C, with a mean of 148.2 °C and a sample standard deviation of 14.2 °C. The automatic searches use the ambient reading taken before recording and initialize the prediction from the first stored sample. Those are different noisy readings in the simulator.

The paired radiative fits reduce the objectives by 6.7 to 7.0 °C on these same four records. This repeatable reduction is smaller than the variation between recordings, but that comparison alone does not make the paired improvement a noise artifact. The noise-free result and the hold-out test establish the relevant limitation: a smaller training objective does not verify the added mechanism. Across the four convective fits, $U = 8.39 \pm 0.28\,\mathrm{W/(m^2\,K)}$ and $\tau = 199 \pm 7\,\mathrm{s}$, where the spreads are sample standard deviations rather than confidence intervals.

A low initial reading shifts the entire predicted trajectory and can cancel part of the structural residual. The first record starts 0.33 °C below the noise-free initial temperature and fits to 134.6 °C, below the noise-free 141.9 °C. Ambient seeding provides a second source of variation. The deployed application uses unseeded browser randomness, so the exact four noisy records are a reproducibility experiment rather than a fixed sequence that every learner will obtain.

11.4.7 Two Heaters and One Shared Coefficient

Including T_2 in the objective adds two parameters and two constraints at once. On run E5 the convective form reaches 514.6 °C on T_1 and 448.6 °C on T_2 for a total of 963.2 °C, with $\alpha_2 = 0.0031\,\mathrm{W/\%}$ against $\alpha_1 = 0.0076\,\mathrm{W/\%}$ and $A_s = 0.000\,43\,\mathrm{m^2}$. The ratio α_2/α_1 of 0.41 underestimates the generator's true capacity ratio P_2/P_1 of 0.50, and the derived time constant rises to 213 s, both of which are the price of forcing one U, one A and one mc_p to serve two heaters that the generator gives different capacities. The radiative form on the same run reaches 1211.2 °C, 25.7% worse, and finishes with U against its lower bound, which the application reports in words rather than leaving the student to notice.

Table 11.8 Final Error of the Online Estimate Against Its Reference

Run	Form	KF	$L = 10$	$L = 30$	$L = 60$	noisy
S1	**C**, T_2	0.05	0.05	0.05	0.05	0.93
S2	**CR**	0.48	0.48	0.48	0.48	1.12
E1	**F**	14.43	14.43	14.43	14.43	9.35
E1	**CR**	9.91	1.42	0.95	1.20	3.39
E4	**C**	10.13	9.90	9.93	9.90	10.59
E5	**CR**, T_2	71.53	6.39	6.29	6.20	6.52
E5	**C**, T_2	65.70	35.55	27.41	27.45	25.89
$E1_{20}$	**F**	0.57	0.57	0.57	0.57	6.42

Largest final error over the estimated parameters, in percent of each range, against the truth (S1, S2) or the search's answer (the rest). KF is the unbounded window. The last column is $L = 30$, the mean of three seeds at 0.2 °C, and the others are noise-free. $E1_{20}$ is E1 with $\tau_S = 20$ s.

Table 11.9 Final Estimate When One Candidate Cause Is Changed

	whole record		walking θ		constant θ	
Run, value	$\lvert e\rvert$	e^2	window	exact	window	exact
E1 **F**, τ	197.1	221.1	111.9	10.0	146.1	174.4
E1 **CR**, U	2.456	1.551	2.826	1.513	3.588	3.522
E4 **C**, U	10.06	9.739	6.188	6.082	10.43	10.57
E5 **C**T_2, U	7.83	7.386	1.758	1.777	5.026	8.144

Noise-free. Whole record: the search on absolute or on squared errors. Window: Equation (11.12) in one pass with $L = 30$. Exact: the same problem over the whole record, relinearized to convergence. Constant: $W = 0$. τ in s, U in W/(m^2 K).

11.4.8 Online Estimate Beside the Regression

The moving horizon estimate of Section 11.3.7 was run on six records, with the answer of the search, or the truth where it is known, as the reference in every comparison. Three are the simulator runs E1, E4 and E5, which carry the 140 s thermistor lag that no fitted form contains, and a fourth repeats E1 with that lag at 20 s. Two are synthetic records generated by the fitted balances themselves, the doublet input with the convective form and T_2 at $U = 9$, $\alpha_1 = 0.008$, $\alpha_2 = 0.004$ and $A_s = 0.0004$ (S1), and the full-power step with the radiative form at $U = 6$ and $\alpha_1 = 0.009$ (S2). Each record was taken noise-free and with 0.2 °C of noise on three seeds, the truth was read by the measurement script alone, and the estimator ran unchanged on all of them. Table 11.8 lists the final errors.

On the synthetic records, where the selected form is the generating one, every estimator ends within 0.05% (S1) and 0.48% (S2) of each range of the truth without noise and within 0.93% and 1.12% on average with it, while the search ends within 0.13% and 0.11% with noise. The second power factor shows when a parameter is not yet excited. On S1, α_2 stays at its published 0.0075 W/% until heater 2 switches on at 450 s and is within 2% of its range of the truth by 452 s.

On the simulator's records the estimate settles away from the search. The first-order time constant ends at 111.9 s against 197.1 s on E1, the convective U at 6.19 against 10.06 W/(m^2 K) on E4, and the convective U on E5 with T_2 at 1.76 against 7.83. The last estimate is not a whole-record model either. Held constant over E1 it gives $J = 866.1$ °C against the search's 141.9 °C, and over E4 it gives 1918.8 °C against 594.0 °C.

Changing one candidate cause at a time locates the difference, as listed in Table 11.9. The missing structure is the precondition. With the generator's thermistor lag set to 5, 20, 50, 80 and 140 s, the first-order τ of the window ends 0.2, 0.6, 3.2, 5.6 and 14.4% of its range from the search's, and with the lag at its 20 s minimum the two agree within 0.57% noise-free, although not under noise (6.42%). The random walk carries the difference. Solving the estimator's own problem exactly over the whole of E4 gives $U = 6.08$ against the window's 6.19, so the one-step linearization costs 0.3%

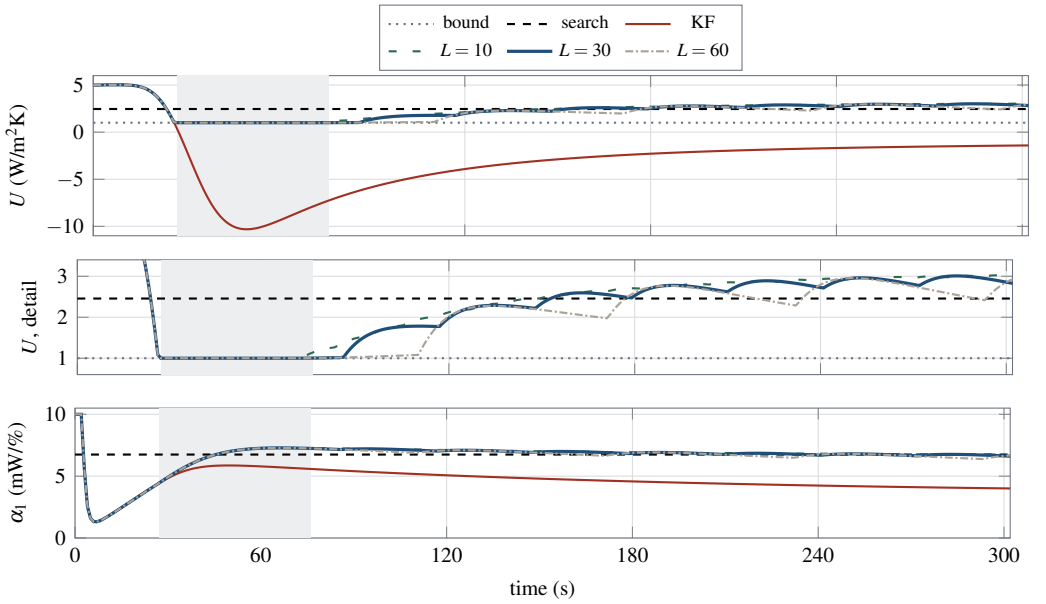


Figure 11.4 Online estimates of U and α_1 on the noise-free 50% step E1 with the radiative form. The unbounded Kalman filter (KF) takes U to -10.3 while the three windows hold it on its lower bound, shaded from 27 to 76 s for $L = 30$. The middle axis enlarges the windows. KF is thick solid, $L = 10$ loosely dashed, $L = 30$ very thick solid and $L = 60$ dash-dotted. The straight dashed lines are the search's answer and the dotted line is the bound.

of the range, while holding the parameters constant brings U to 10.43 in one pass and 10.57 solved exactly, within 1.3% of the search. With the walk off, the doublet's U stays within 0.4 to 1.2% of the search's at lags from 50 to 140 s, against 9.9 to 24.9% with it. Inside the window estimate α_1 falls from 0.0052 to 0.0016 W/% during the 30% stretch of the doublet between 300 and 450 s and then recovers, so the parameters drift to absorb the lag the form leaves out, and the last value describes the end of the record rather than the whole of it. The objective is a small part, because refitting the whole record by squares moves the answer by 0.8 (E4) to 4.1% (E1, first order) of the range. On the first-order form the walking problem is degenerate. Solved exactly it holds τ on its 10 s bound for 68% of the record and lets K_p grow from 0.013 to 0.52 °C/%, tracing the lagged rise as a growing gain, so relinearization moves the estimate further from the search rather than closer. Halving the estimator's substep or the generator's Euler step moved no value here by more than 0.02% of a range.

🕮 **Pause.** Try Exercise 11.3 before reading on.

The bound matters on E1 with the radiative form, as shown in Figure 11.4. The unbounded filter drives U below its lower bound of 1 W/(m^2 K) in 91% of its updates, reaching -10.3 at 50 s, 29% of the range below the bound, and ends at -1.41, a negative heat transfer coefficient whose pair gives $J = 1687$ °C over the record. An unbounded filter leaving the physical region in this way is a documented weakness of the extended Kalman filter that bounded moving horizon estimation avoids [18]. The window with $L = 30$ holds U on the bound from 27 to 76 s, 16.5% of its updates, then leaves it, and the three windows end at 3.01, 2.83 and 2.93 for $L = 10$, 30 and 60 against the search's 2.456, with $J = 246.0$ °C for $L = 30$ against 134.2 °C. Under noise the filter ends between -0.93 and -0.54 and the window between 3.14 and 4.16 on the three seeds. On E5 with the radiative form and T_2 the filter ends at $U = -6.65$ and $\alpha_2 = -0.000\,37$ W/%, both outside their ranges, while the window stays inside every range and ends within 6.3% of the search. Where no bound is active the window length does not matter, and the filter and all three windows agree to 10^{-6} of the range on

E1 with the first-order form, on S1 and S2, and with the 20 s lag. Where one is active, $L = 10$ and $L = 60$ differ by up to 8.1% (E5, convective), and $L = 30$ and $L = 60$ stay within 1.2% of each other in every case of Table 11.8.

The estimator's covariance exposes the trade that the search reports only through a parameter on its bound. Its Riccati covariance puts the correlation of U and α_1 at 0.89 to 0.98 on every record of the physical forms, against 0.54 to 0.59 for τ and K_p on E1, because each record fixes the steady gain $\alpha_1/(UA)$ far better than U and α_1 separately. The application does not display this correlation. On E5 with the radiative form, where the search ends with U on its lower bound, the window rides that bound for 10.5% of its updates early in the run and then leaves it, so it does not repeat the warning of the search.

An update with $L = 30$ cost a median of 1.81 ms and a 99th percentile of 3.55 ms in Node.js on the four-parameter radiative form, and 6.21 ms at the 99th percentile on the convective form with T_2, where $L = 60$ reached 43.6 ms, on a computer carrying other work (one-minute load average 1.8). In Chrome on the built application at 10× with the same four parameters, the median was 2.7 ms and the 99th percentile 6.5 ms against the 10 ms budget, with 600 of 600 samples taken, no long task and a load average of 2.7, a margin of only a factor of 1.5. At 1× the 99th percentile was the cold first update, 42.2 ms of a 100 ms budget. The estimate runs only while its view is shown, so the default view costs nothing.

11.4.9 Fitted Result in the Application

Figure 11.5 shows the current application after noise-free run E2 with the radiative form fitted. The heat transfer coefficient is 1.56 W/(m^2 K) at a held emissivity of 0.9, and the objective is 250.2 °C over 303 points. The integrated loss split is 19.7% convective and 80.3% radiative. A mean error of 0.83 °C therefore accompanies a convective time constant of 1072 s, far above the generator's 157.6 s mean lag, and a radiative loss absent from the data generator. The run and its prediction are computed in the browser and export as a CSV file for independent analysis, under comment lines naming the source, the step test, the model form and each parameter value.

11.5 Technology in Practice

This chapter has no loop to close. Its hardware is a measurement path that ends in a record and a fit, and Figure 11.6 draws that path with the devices of the board.

What Measures It

The data generator reads each thermistor through a first-order lag and white noise, Equations (11.3) and (11.4), while the three fitted forms carry no sensor at all, the assumption that Section 11.4 finds behind most of the residual. A record used to choose between models must first be trusted, and calibration compares a sensor's response with a precisely known input [19]. A zero error would shift every reading and with them the ambient value seeded from the first one. No range or accuracy for the board's sensors is given here, because none was found in an open source that names no maker.

> **TECHNOLOGY** **Sensor: A Temperature Reading Checked Against a Standard**
>
> To calibrate an instrument is to expose it to a precisely known input and adjust its response so the output matches across the whole range. A zero shift moves every reading by the same amount, while a span shift changes the slope, so its error is unequal across the range. Technicians record each calibration as found and as left, because the difference over time measures drift, and excessive drift often signals an impending failure. Water supplies two convenient reference points, its freezing and boiling points, and other substances supply the fixed points of the international temperature scale [19]. For a fitted model, a span error of the sensor would scale the measured rise and with it the identified gain.

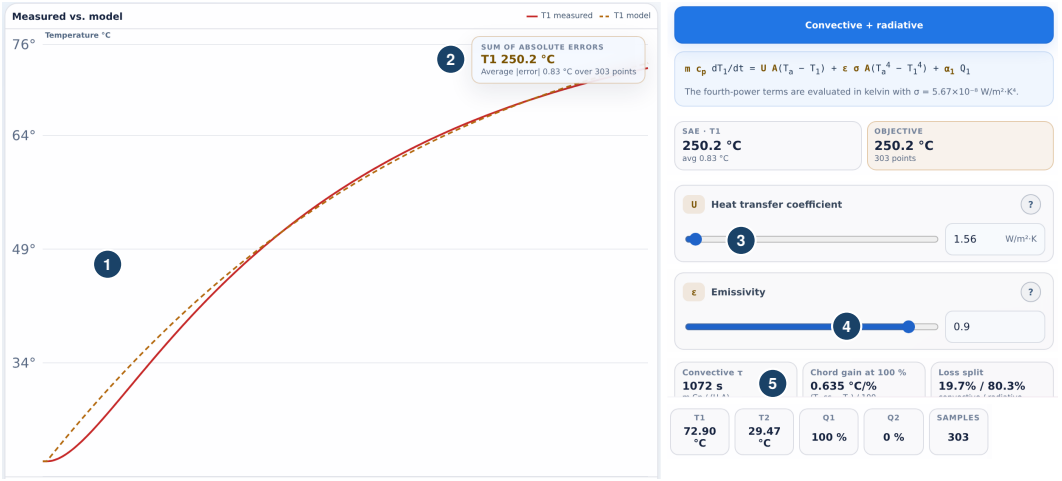


Figure 11.5 The current application after the noise-free full-power step E2, shown as a composite of six parts of the screenshot: the measured and modeled temperature on the left, and on the right the selected form, the objective tiles, two sliders, the readouts and the measurement tiles of the run. The heater trend and the time axis are cut. Heater 1 steps to 100% at 2 s (Table 11.3), and the final temperature is 72.90 °C at 302 s. The radiative best fit has an objective of 250.2 °C, $U = 1.56\,\mathrm{W/(m^2\,K)}$, and an integrated radiative loss fraction of 80.3%. Callouts: (1) the measured temperature (solid) and the model (dashed), which rises first; (2) the sum of absolute errors and its count of points; (3) U near the low end of its range; (4) the emissivity, held; (5) the readouts of the convective time constant, the chord gain and the loss split.

What Moves It

Each heater is a transistor whose duty cycle sets its share of full power, the final element that Chapter 4 describes. Here no measurement sets it: a stored program or the operator does, so the design of the input matters more than the device. Industrial test signals must excite both the steady state at low frequency and the dynamics at medium and high frequency, and step-like binary signals are the ones most used [8]. Table 11.7 shows the same need on the board: one heater level ties T_a to α_1, and two levels separate them.

> **IN PRACTICE** **Record Where the Model Was Validated**
>
> A government standard for models and simulations, NASA-STD-7009B, was first issued in 2006 and drawn up partly in answer to the findings on the Columbia accident. It requires that a model be validated and that a record of its domain of validation be kept: the region of inputs over which its responses compare favorably with a referent, the real system or an accepted stand-in. Calibration, also called tuning, adjusts parameters to improve agreement with a referent and is a separate act. A model may extrapolate outside the validated domain only within its stated limits, and the aspects not validated are to be recorded [20].

Where the Controller Runs

The platform is the browser page. Each second it reads both temperatures over the serial link and writes the held heater command (Section 11.3.10), the cycle that a standards glossary defines as reading sensor signals into memory, executing the algorithm and sending the actuator signals [21]. Plant loops run with cycle times from milliseconds to minutes [21], and the 1 s sample sits inside that range, short against the 197 s time constant fitted to E1. The simplex search and the moving horizon

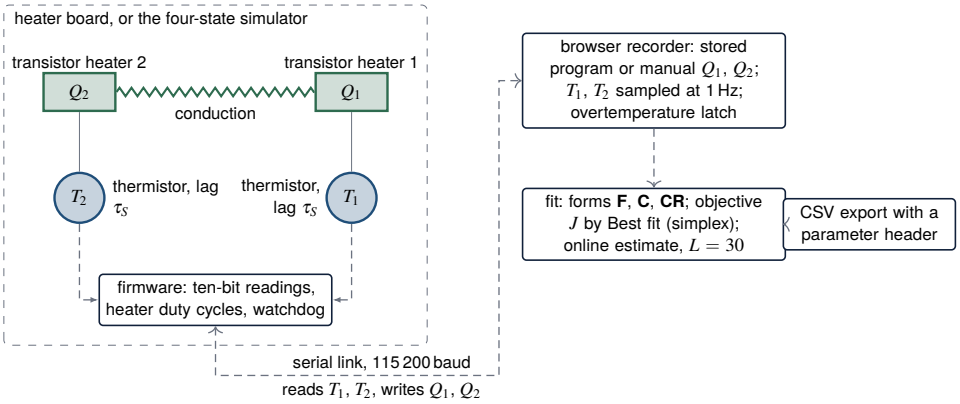


Figure 11.6 The measurement path of the chapter on the board. Each thermistor sits beside its transistor heater, the firmware converts the readings and applies the duty cycles, and the browser records both temperatures once a second and fits the three forms to the record. Notice that no arrow runs from a temperature back to a heater: the heater commands come from a program, and the only automatic action is the overtemperature latch.

estimate run in the same page. In a plant the record would come from an archive of the control system rather than from the loop itself.

> **TECHNOLOGY** **Platform: A Data Historian**
>
> A data historian is a computer that polls process data from the control system at regular intervals, stores them on a high-capacity drive, and serves them to other computers on request, typically over years. Historians existed in distributed control systems long before network security zones, because the controllers have too little memory to archive process data and plant staff wanted long-term records for running the process and diagnosing problems. In a secured network the historian sits in a buffer zone between the plant and office networks and supplies its records without any direct reading from the controllers [19]. The recorder of this chapter plays the same part at one sample per second, and its exported file names the source, the test, the model form and every parameter.

What Is Modernizing It

Three developments from the literature on digital twins reach this fit. The first is hybrid modeling: a physics-based model explains part of the data, the interpretable residual it leaves is modeled by an explainable data-driven method, and only what remains goes to a black-box model before the last residual is discarded as noise [9]. The structured residual of Figure 11.3 is such a residual, the trace of the missing sensor state. The second is the automatic recovery of governing equations by sparse regression over many candidate terms, which is fast but performs poorly on data corrupted with noise [9]. The third is data assimilation, routine in weather centers, which combines a model with observations by variational methods or Kalman filters [9]. The moving horizon estimate of Section 11.3.7 applies that idea to two parameters of one heater.

11.6 Limitations

The data generator is linear in its loss term, so the negative verdict on the radiative form in Table 11.4 is a verdict about this data set and not about a physical TCLab. Measurements on real boards show a steady-state gain that falls at elevated temperature, and radiation is the usual explanation [1], so the same experiment run on hardware may well rank the three forms differently. The method of this chapter does not depend on which way the ranking falls.

The application supplies no confidence intervals, no sensitivity matrix and no information criterion, so the comparison between models is made on a raw objective value and on cross-validation alone. A collinearity index computed from the sensitivity matrix would quantify the trade that Table 11.6 only exhibits [10], and an information criterion such as Equation (1.23) would price the third parameter directly instead of leaving the noise floor of Section 11.4 to do it [14]. Chapter 12 adds both kinds of evidence for linear models, with a moving-block bootstrap for the uncertainty of each fit and a validation interval that the fit never sees.

The derived time-constant readout is $mc_p/(UA)$ for both physical forms, and the current application labels the radiative value as the convective time constant. It ignores the fourth-power term, so the 1072 s in Figure 11.5 describes the convective path alone. The gain readout is the chord gain from the full balance, 0.635 °C/% for this fit. Using $\alpha_1/(UA)$ instead would give 3.545 °C/%, overstating that gain by a factor of 5.58.

The online estimate is a tracking estimate, not a second regression. Its parameters are random walks, so on a record that no form describes its last value summarizes the end of the record and scores $J = 866.1\,°\mathrm{C}$ over the whole of E1 against the search's 141.9 °C. Its model omits the thermistor lag exactly as the three forms do, linearizes each step once at the estimate of that moment, and carries fixed weights rather than weights fitted to the run. It reports no confidence interval, and the parameter correlation it computes is not displayed.

The three fitted forms share one heat transfer coefficient, one area and one thermal capacity between the two heaters, so a board whose two channels differ cannot be matched exactly at any parameter setting.

The simplex search moves two parameters, or four with the second heater included, and holds the rest. It is a reference answer rather than an estimator, and it carries no global-search strategy beyond two restarts from progressively smaller simplexes, so a badly placed starting point can leave it in a local minimum.

The hardware path was documented from the source code and was not exercised, because no physical board was attached during this work. Reading a board needs Web Serial, which desktop Chromium-family browsers expose and other browsers do not, and which a page reached over plain HTTP cannot use at all.

11.7 Discussion

A parameter set is a property of one data record, while a model structure is a claim about mechanism that every later use inherits. The distinction becomes expensive downstream. A controller tuned from an identified model inherits the model's gain error directly, and a predictive controller inherits its extrapolation as well, which is why model development rather than control design dominates the cost of an industrial application [8]. A digital twin makes the same inheritance permanent, because the twin is consulted precisely in the conditions where running the plant would be too expensive, and those are the conditions in which no data were ever taken [22]. Chapter 4 runs such a twin beside the board it models.

The experiment of this chapter is small enough that both outcomes can be seen in one afternoon. A model with an extra physical term fits the data it was given by 5.4% better and predicts a different experiment 51% worse, and the same model extrapolates a steady state to within 0.46 °C where the simpler one is off by 7.62 °C. Neither result is a general law about radiation on transistors. Both are what happens when a term is added to a balance and the data are not asked whether they can support it, which is the habit this chapter sets out to break.

The same choices reappear when the fit runs online. Declaring which parameters an estimator may move is the subset-selection decision that Table 11.6 forces by hand, and an estimator left free to move ε and U together would wander along the flat valley for the same reason the batch search would. Adding the sensor state that all three forms omit turns the same formulation into a soft sensor

for the transistor temperature, which no thermistor on the board reports. Volume 5, Chapter 11 builds a soft sensor of that kind for a temperature no instrument reaches.

> **OPEN CHALLENGE** **Twins That Extrapolate and Say How Far to Trust Them**
>
> A review of digital twins counts predictive capacity, transparency and generalization among the main technical challenges. Physics-based models generalize well to new problems with similar physics and allow their errors to be bounded, yet they capture only part of the known physics because of the assumptions made along the way. Data-driven models take in long histories but generalize poorly to unseen problems, and their errors cannot be bounded. The review proposes hybrid models that join the two [9]. Progress would be a twin whose predictions beyond its training record carry an error bound that holds when the device is tested there.

11.8 Conclusion

This chapter studied an application that records a Temperature Control Lab step test and fits three nested energy balances to it under one sum of absolute errors, with no controller anywhere. The first-order and convective forms are shown to be the same two-parameter model in different symbols, returning identical objectives of 141.9 °C on the 50% step and 963.2 °C on the two-heater run from independent searches, and the products UA and mc_p are shown to be the only identifiable groups by direct substitution. Adding the Stefan-Boltzmann term improves the 50% step fit by 5.4%, to 134.2 °C, while degrading the untouched 100% step from 283.8 to 429.0 °C and reporting a radiative share of 70.1% in data generated with no radiation at all.

An emissivity scan moves the objective by 6.0% across the entire range from 0 to 1 while the fitted heat transfer coefficient moves by a factor of 4.7, and four repeats of the same step test with 0.2 °C of sensor noise give objectives of 148.2 ± 14.2 °C. That spread is larger than the improvement the extra term buys, although the paired reduction of 6.7 to 7.0 °C repeats on every record, so it is the hold-out test and not the noise that rejects the term.

A moving horizon estimate of the same parameters agrees with the search within 0.48% of each range on records its own form generated, settles 14.4% of the range away on the simulator's 50% step because its drifting parameters absorb the thermistor lag, and holds U inside its range where an unbounded Kalman filter ends at -1.41 W/(m^2 K).

A fuller study would add a confidence interval on each fitted parameter, a collinearity index over the selected subset, an information criterion beside the raw objective, a sensor state as a fourth candidate structure, and the same experiments on physical boards. The question the application poses, which model a measurement can support, is answered by a second record and not by the objective of the first.

> **KEY POINTS**
>
> - Which model a measurement can support is decided by a second record, not by the objective of the first.
> - The first-order and convective forms are the same two-parameter model in different symbols, and only the products UA and mc_p are identifiable.
> - Adding radiation improves the 50% step fit by 5.4% while degrading the untouched 100% step from 283.8 to 429.0 °C, and reports a radiative share of 70.1% in data generated with none.
> - A moving horizon estimate settles 14.4% of the range away on the 50% step because its drifting parameters absorb the thermistor lag.
>
> **Where else this applies.** Chapter 12 adds bootstrap uncertainty and a validation interval for linear models, and Volume 5, Chapter 11 builds a soft sensor for a temperature that no instrument reaches.

References

[1] 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.

[2] J. A. Rossiter, B. Pasik-Duncan, S. Dormido, L. Vlacic, B. Jones, and R. Murray, "A survey of good practice in control education," *European Journal of Engineering Education*, vol. 43, no. 6, pp. 801–823, 2018.

[3] G. E. P. Box and W. J. Hill, "Discrimination among mechanistic models," *Technometrics*, vol. 9, no. 1, pp. 57–71, 1967.

[4] C. Kravaris, J. Hahn, and Y. Chu, "Advances and selected recent developments in state and parameter estimation," *Computers & Chemical Engineering*, vol. 51, pp. 111–123, 2013.

[5] 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.

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

[7] L. Ljung, "Perspectives on system identification," *Annual Reviews in Control*, vol. 34, no. 1, pp. 1–12, 2010.

[8] 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.

[9] A. Rasheed, O. San, and T. Kvamsdal, "Digital twin: Values, challenges and enablers," arXiv preprint 1910.01719 (version 1); published in revised form as IEEE Access 8, 21980–22012, 2020, 2019.

[10] R. Brun, P. Reichert, and H. R. Künsch, "Practical identifiability analysis of large environmental simulation models," *Water Resources Research*, vol. 37, no. 4, pp. 1015–1030, 2001.

[11] G. Franceschini and S. Macchietto, "Model-based design of experiments for parameter precision: State of the art," *Chemical Engineering Science*, vol. 63, no. 19, pp. 4846–4872, 2008.

[12] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[13] 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.

[14] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[15] D. W. Hahn and M. N. Özişik, *Heat Conduction*. Hoboken, NJ: Wiley, 2012.

[16] A. Bejan, *Convection Heat Transfer*. Hoboken, NJ: Wiley, 2013.

[17] 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.

[18] 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.

[19] 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.

[20] National Aeronautics and Space Administration, "Standard for models and simulations," Office of the NASA Chief Engineer, Washington, DC, NASA Technical Standard NASA-STD-7009B, 2024, approved 2024-03-05, standards.nasa.gov.

[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, Gaithersburg, MD, Special Publication 800-82r3, 2023.

[22] A. Rasheed, O. San, and T. Kvamsdal, "Digital twin: Values, challenges and enablers from a modeling perspective," *IEEE Access*, vol. 8, pp. 21 980–22 012, 2020.

11.9 Exercises

11.1 ⌕ Concept check · 2 min

On run E1 the fitted convective form has $U = 8.46\,\mathrm{W/(m^2\,K)}$ and $A = 0.0012\,\mathrm{m^2}$. Doubling A while halving U leaves J at 141.9 °C in every digit, while doubling α_1 alone raises it to 5039.9 °C. What does this show?

(a) The objective is insensitive to geometry, because the heater is small beside the board

(b) The record fixes the products UA and mc_p, not their factors, and α_1 lies outside both

(c) The search stopped in a local minimum, so the area should be added to the fitted set

(d) The area enters only through radiation, which the convective form leaves out

11.2 Concept check · 2 min

Fitting the convective plus radiative form to E1 lowers J from 141.9 to 134.2 °C and reports that 70.1 % of the loss went out by radiation. What does the lower objective establish on its own?

(a) That radiation carries most of the heat that leaves the transistor in these data

(b) That the convective form is mis-specified and should be dropped for this board

(c) Nothing about the mechanism, since a larger nested model cannot fit its own record worse

(d) That the improvement is noise, since it is smaller than the spread of the noisy repeats

11.3 Concept check · 2 min

With the noise off, the online estimate of the first-order form on E1 ends at $\tau = 111.9\,\mathrm{s}$ while the search returns 197.1 s. Which explanation does the chapter's evidence support?

(a) A window of 30 samples is too short to see a time constant near 200 s

(b) The estimator sums squared errors, and squared errors favor a short time constant

(c) The lower bound of τ is active through the run and holds the estimate down

(d) The random walk lets τ and K_p drift to absorb the thermistor lag the form omits

11.4 Theory and calculation · 7 min

Take the data generator at its defaults: $P_1 = 200$, $C_H = 5720\,\%\,\mathrm{s/°C}$, $\tau_{HA} = 20\,\mathrm{s}$, $\tau_{HH} = 100\,\mathrm{s}$ and $\tau_S = 140\,\mathrm{s}$, with $Q_2 = 0$. Write $a = 1/\tau_{HA}$ and $b = 1/\tau_{HH}$.

(a) Set the derivatives of Equations (11.1) to (11.3) to zero and find the steady-state gain from Q_1 to T_1. *Checkpoint:* 0.599 °C/%.

(b) The transfer function from Q_1 to T_{H1} is $G(s) = (P_1/C_H)(s+a+b)/\big((s+a)(s+a+2b)\big)$. Its mean lag $-G'(0)/G(0)$ is the sum of the time constants of its poles less that of its zero. Find it, and add the thermistor lag. *Checkpoint:* 17.6 s and 157.6 s.

(c) Run E1 holds $Q_1 = 50\,\%$ from 2 s to 302 s and ends 25.95 °C above its start. What fraction of the final rise is that? *Checkpoint:* 86.6 %.

Say what (c) means for a model fitted to E1 alone.

11.5 Theory and calculation · 6 min

The convective fit to E1 is $U = 8.46\,\mathrm{W/(m^2\,K)}$ and $\alpha_1 = 0.006\,86\,\mathrm{W/\%}$, with the held values $A = 0.0012\,\mathrm{m^2}$, $m = 0.004\,\mathrm{kg}$ and $c_p = 500\,\mathrm{J/(kg\,K)}$.

(a) Find the time constant and gain of the equivalent first-order form, and the gain error against the generator's 0.599 °C/%. *Checkpoint:* 197 s, 0.676 °C/%, about 13 % high.

(b) At steady state Equation (11.6) fixes only $UA\,T_a + \alpha_1 Q_1$ on a record with one heater level. With UA held, how much must α_1 change at $Q_1 = 50\,\%$ to absorb an error of 6 °C in T_a, and what fraction of the fitted value is that? *Checkpoint:* About 0.0012 W/%, about 18 %.

Compare (b) with the E1 columns of Table 11.7, and say why the doublet breaks the trade.

11.6 Theory and calculation · 7 min

The radiative term of Equation (11.7) uses kelvin.

(a) Show that $\varepsilon\sigma A(T_1^4 - T_a^4) = h_r A(T_1 - T_a)$ with $h_r = \varepsilon\sigma(T_1 + T_a)(T_1^2 + T_a^2)$, and evaluate h_r for $\varepsilon = 0.9$ and $T_a = 21$ °C at $T_1 = 21$ and at 47 °C. *Checkpoint:* 5.20 and 5.93 W/(m² K).

(b) The radiative fit to E1 has $U = 2.46\,\mathrm{W/(m^2\,K)}$. Add h_r at both ends of the rise and compare with the convective fit's 8.46. *Checkpoint:* 7.66 and 8.39 W/(m² K).

(c) At the published values $U = 5\,\mathrm{W/(m^2\,K)}$ and $\varepsilon = 0.9$, find the instantaneous radiative share $h_r/(U + h_r)$ at 53.7 and at 83.1 °C. *Checkpoint:* 55.1 % and 58.6 %.

Explain from (a) to (c) why the emissivity valley of Table 11.6 is flat.

11.7 App challenge · 8 min

Reload the app so every form holds its published values. In Settings set the Sensor noise to 0 and press Apply. With the Simulator source, choose the step test Q1 = 50% · 5 min, set Sim speed to 10× and press Run. The run ends at 5:02, about 30 s of real time, and sets T_a to 21.0 °C in every form. This is scenario 1, Record and fit the first-order form, and scenario 2, Prove the lumped groups.

(a) Select First-order and read the Objective. Then press Best fit and read τ, K_p and the Objective. *Checkpoint:* 2767.9 °C before the fit. Then $\tau = 197$ s, $K_p = 0.676\,°\mathrm{C}/\%$ and 141.9 °C, each within one unit of the last digit.

(b) Select Convective and read the Objective, Implied τ and Implied gain at the published values. Then type 20 for U and 0.0006 for A and read the three again. *Checkpoint:* 1876.4 °C, 167 s and 0.833 °C/%, both times.

(c) Leave A at 0.0006, press Best fit and read U, α_1, Implied τ, Implied gain and the Objective. *Checkpoint:* $U = 16.91\,\mathrm{W}/(\mathrm{m}^2\,\mathrm{K})$ and $\alpha_1 = 0.0069\,\mathrm{W}/\%$, then 197 s, 0.676 °C/% and 141.9 °C.

Say what (b) and (c) show together.

11.8 App challenge · 8 min

Continue from the previous exercise, or repeat its setup and Best fit on the Convective form. This is scenario 10, Use the 100% step.

(a) Select Convective + radiative and press Best fit on the E1 record. Read U, the Objective and the Loss split. *Checkpoint:* $U = 2.46\,\mathrm{W}/(\mathrm{m}^2\,\mathrm{K})$, 134.2 °C, 29.9% convective and 70.1% radiative.

(b) Choose Q1 = 100% · 5 min and press Run without touching the parameters. When the run ends, read the Objective of the radiative form, then select Convective and read its Objective. Press Best fit there and see whether it improves. *Checkpoint:* 429.0 °C and 283.8 °C, and Best fit stays at 283.8 °C.

(c) Select Convective + radiative and press Best fit on E2. Read U, the Objective and the Loss split. *Checkpoint:* $U = 1.56\,\mathrm{W}/(\mathrm{m}^2\,\mathrm{K})$, 250.2 °C, 19.7% and 80.3%.

Which model would you carry to a third experiment, and why?

11.9 App challenge · 9 min

This is scenario 13, Watch the fit converge online. Reload the app, set the Sensor noise to 0 in Settings and press Apply, choose Q1 = 50% · 5 min and the First-order form, switch the chart view to Online estimate, set Sim speed to 10× and press Run.

(a) When the run ends, read the window estimates of τ and K_p, and check whether either is marked on its bound. Then press Best fit and read its τ in the same view. Express the gap in percent of the τ range of 10 to 600 s. *Checkpoint:* $\tau = 111.9$ s and K_p about 0.582, neither on its bound. Best fit 197.1 s. Gap 14.4 %.

(b) In Settings set the Sensor response time to 20 s, press Apply and Run the same test again. Read the window τ and, after Best fit, the search's τ, and express the gap the same way. *Checkpoint:* About 31.9 s against 35.3 s, a gap of 0.57 % of the range.

What does the change from (a) to (b) say about where the disagreement comes from?

11.10 Application area · 3 min

An engineer plans to repeat the full-power step E2 on a physical board with the overtemperature limit at its default.

(a) The noise-free simulated E2 passes 70 °C about 259 s after the run starts and ends at 72.90 °C. If the board followed the simulator, what would happen? *Checkpoint:* The run stops near 259 s with both heaters off and the alarm latched.

(b) The limit can be raised to 90 °C before the board is connected. Would a longer, hotter step separate radiation from convection? *Checkpoint:* Hardly: the radiative share moves only from 55.1 to 58.6 %.

11.11 Application area · 3 min

A model for tuning is fitted to a five-minute step that reached about 87 % of its final rise, and the fitted gain is 0.676 °C/% where the true gain is 0.599 °C/%.

(a) A PI controller is tuned for a chosen loop gain $K_c K_p$ using the fitted gain. By what factor is the actual loop gain off? *Checkpoint:* $0.599/0.676 = 0.886$, about 11 % low.

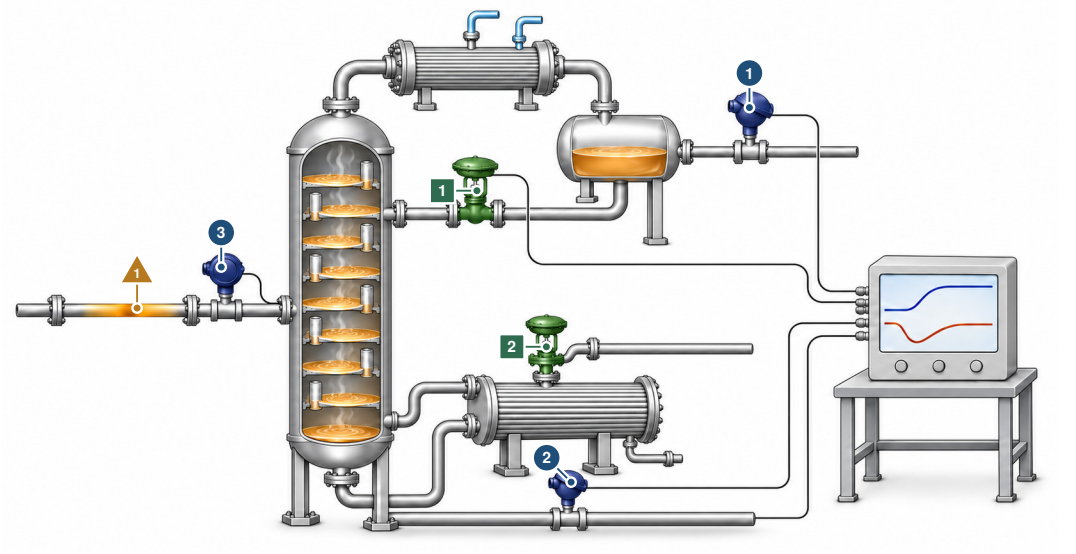
(b) If the search had also been allowed to move T_a on the two-level record E3, where would it have gone, and what practice prevents it? *Checkpoint:* Down to about 15 °C, where J is 354.6 °C. Measure T_a instead.

11.12 Review the agent · 5 min

An AI assistant reviews runs E1 and E2 and writes: “The convective plus radiative form is the better model of this board. It lowers the objective on both step tests, by 5.4 % on E1 and 11.8 % on E2, its loss breakdown shows that 70 % of the heat leaves by radiation, and its full-power steady state of 80.48 °C is within 0.46 °C of the true 80.94 °C, where the convective model is 7.62 °C off. Use it as the model for a predictive controller.” Find the flaw and say how you would check it.

CHAPTER 12

Multivariable Plant Identification Workbench with Exact Structural Zeros



The controller moves 1 reflux flow, stepped in the plant test 2 steam to the reboiler, stepped in the plant test

The controller reads 1 distillate composition analyzer 2 bottoms composition analyzer 3 feed flow meter

What disturbs it 1 feed rate changes, measured but not adjusted

Notice that reflux and steam both move each product composition, so a fit must divide the credit between them from one record; the column pressure transmitter of the chapter is not drawn. Illustration, not to scale.

IN PLAIN TERMS Which Input Moved Which Output?

The main idea. A plant keeps a record of the handles its operators moved and the readings that followed. This chapter fits models to such a record to learn how strongly, and how late, each handle moves each reading.

The challenge. When several handles move at once, a careless fit gives one handle the credit for another's work. Engineers also know that some handles cannot touch some readings. The model must leave those links out completely, and it must be tested on a part of the record it never saw.

Why it matters. A controller that steers a plant from such a model can only be as right as the model.

THREE WAYS TO READ THIS CHAPTER

⊘	**Explore**	12 min	the first page, the scale of predictive control in industry (Section 12.1), the history of Identification and Command (Section 12.2), the try-it on an absent path, the fact on true and false zeros and the try-it on the predictive controller (Section 12.6) and the key points
🕮	**Study**	65 min	the whole chapter with its three pause points and the twelve exercises at its end
⛰	**Practice**	20 min	the first page, the results (Section 12.6), Technology in Practice (Section 12.7), Limitations (Section 12.8) and the open challenge

The control question A plant record holds several inputs and outputs that move together, sometimes under feedback. How is the credit for each output divided among correlated inputs, how is an absent path removed exactly, and how is a model judged on data the fit never saw?

Learning objectives

- Explain why the inputs into a CV are fitted jointly, and why a fitted gain depends on the model structure.
- Compute the sampled form of a path with fractional dead time and the parameter count of a block state-space model.
- Tell a correct structural zero from a false one by free-run validation error, and choose a family for an integrating or ringing record.
- Pair and tune PI loops from a fit, and explain why an output-bias predictive controller loses an unmeasured load.

Prerequisites Sections 1.4.3 to 1.4.6, 1.4.8 and 1.4.9, and Chapter 11. Sections 1.6.2 and 1.6.3 only for Section 12.5.

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

See also Volume 6, Chapter 2 (loop pairing) and Volume 6, Chapter 8 (predictive control above a decentralized structure). Volume 9, Chapter 4 (offset-free predictive control with a Kalman estimator).

12.1 Introduction

The models inside industrial multivariable controllers are seldom derived from conservation laws. They are identified from plant tests in which operators step each manipulated variable in turn or together, and the result is a matrix of responses that relates every manipulated and measured disturbance variable to every controlled variable. That arrangement goes back to the first predictive controllers, which predicted from impulse-response models obtained on the plant [1], and it remained the dominant model form across the industrial vendors surveyed a quarter century later, where the identification step and its plant test were reported as the most time-consuming part of a project [2]. A textbook treatment of identification usually stops at one input and one output, and the single-input reaction curve is the model a first course teaches a student to read. Section 1.4.8 of Chapter 1 fits that case, and Chapter 11 chooses among single-input models of a heater.

DID YOU KNOW? More Than 4600 Predictive Control Applications

A survey of industrial predictive control published in 2003 counted more than 4600 applications carried out by the vendors themselves, over twice the number in a survey five years earlier. Refining held the largest single block, and the size of an application ranged from a single variable to several hundred variables [2].

Three things change when there are several inputs. The first is credit. Two inputs that moved during the same hour both explain the same output movement, and a fit of each input alone credits both with it, so the inputs entering a CV must be fitted together and their correlation becomes a property of the data that limits what any method can recover [3]. The second is structure. An engineer usually knows that some paths do not exist, for example that a reflux valve has no mechanism by which to move column pressure, and a model that estimates a small spurious gain on that path is both less interpretable and, when the controller acts on it, less safe than one that holds it at zero. The third is honesty in scoring. A model that predicts one sample ahead from the latest measurement looks accurate on almost any record, and the free-run simulation that a controller actually relies on is a much harder test [4], as Table 1.1 of Chapter 1 shows for a single heater. Distillation trains, fired heaters, reactors with jacket and feed handles, building air handlers and compressor stations all share this shape, which is why multivariable plant testing for predictive control has its own literature of test signals, order selection and error bounds [2, 3].

The misconception this application is built to correct is that a multivariable model is a collection of single-input models. Fitting each path separately and adding the answers double-counts correlated inputs, a small estimated gain is mistaken for an absent path, and a one-step-ahead fit is mistaken for a model. The workbench removes each shortcut. All active inputs into a CV are fitted jointly, a path set to zero is removed from every model family so exactly that its Markov parameters are 0.0 and not 10^{-12}, and candidates are ranked on free-run error over a validation interval that the fit did not see, with a final test interval held back until a model is chosen.

This chapter works with that workbench, its five model families and the measured behavior of each family on five records whose true structure is known. Every estimator and controller in it is an established method. Section 12.2 traces plant testing and identification for multivariable control. Section 12.3 states the identification problem: the signals and their roles, the preparation of a record, the five families, the method settings and seven assumptions. Section 12.4 gives the algorithms as coded, in enough detail to rebuild the application, among them an exact zero-order-hold discretization with fractional dead time shared by fitting and export, a structural block realization that enforces a zero path in a subspace model, and an export that proves its own equivalence to the application in GEKKO. Section 12.5 builds a predictive controller and decentralized PI loops from a chosen fit. Section 12.6 reports validation scores, recovered gains and parity errors read off the running pipeline, including a weak path that is real, whose removal raises validation error by a factor of 3.7, and measures the predictive controller against the PI loops. Section 12.7 describes the devices and platforms behind such a record, and Sections 12.8 to 12.10 give the limitations, a discussion and a conclusion, and the exercises of Section 12.11 follow the thirteen scenarios of the app.

12.2 Plant Testing and Identification for Multivariable Control

Statistical identification of dynamic systems from process data took its modern shape in the 1960s with maximum-likelihood estimation of difference equations from normal operating records, work that was aimed explicitly at industrial processes and that introduced the noise model as part of the estimation problem [5]. By 1971 a survey could already organize the field into correlation, least-squares, instrumental-variable and maximum-likelihood methods, with model structure, input design and validation as recognized topics in their own right [6]. The same decade produced the Box-Jenkins family of input and noise polynomials, and the prediction-error framework that later unified them treats every structure, ARX and output-error included, as a predictor whose one-step error is minimized [4, 7]. Realization theory developed in parallel. A finite-dimensional state-space model can be constructed from its sequence of impulse-response (Markov) parameters [8], which is the result this application relies on when it certifies a zero path by checking those parameters.

Industrial practice took a more direct route. The predictive controllers of the late 1970s stored the plant as an impulse or step response identified from a plant test [1], and finite step-response and impulse models remained the most common representation in commercial packages surveyed in 2003, alongside a growing use of state-space forms [2]. The response matrix was attractive because each cell can be read by a process engineer. A gain has a sign that can be checked against a balance, a dead time can be checked against a pipe length, and an empty cell states a structural belief about the unit. Plant testing for these controllers matured into multivariable test design with simultaneous perturbation signals, model order chosen by an asymptotic criterion and frequency-domain error bounds on each path [3].

> ⧖ HISTORY **Identification and Command, 1976 to 1978**
>
> The first applications of model predictive control were presented at a conference in 1976 and summarized in a journal paper in 1978. Their software was called IDCOM, an acronym for Identification and Command. Its plant model was a finite impulse response identified from plant test data, by an algorithm that minimized the distance between the plant and model impulse responses in parameter space. The same algorithm then solved the control problem, which its authors treated as the mathematical dual of identification. The papers

> also fixed a vocabulary this chapter still uses: manipulated variables that the controller adjusts, disturbance variables that are not available for control, and controlled variables at the outputs. The first papers on the rival step-response method, in 1979 and 1980, did not discuss their identification technology at all [2].

Subspace methods supplied the state-space alternative. MOESP estimates the extended observability matrix of a deterministic system from an LQ factorization of block Hankel matrices of input and output data [9], its past-output variant handles a system in innovations form [10], and N4SID reached the same structure through oblique projections [11]. These methods need no iterative search and deliver a coupled model of every output at once, which is their strength and also the reason a zero path cannot be imposed on them by editing one entry of a matrix. Plant records are also frequently collected while controllers run, and closed-loop identification showed that the direct prediction-error approach remains consistent under feedback only when the noise model is right, while instrumental and two-stage methods trade efficiency for robustness to that error [12].

Two benchmarks recur in teaching this material. The Wood-Berry transfer matrix of a methanol-water pilot column, with its large opposing gains and unequal dead times, has been a standard multivariable test problem since 1973 [13]. The Temperature Control Lab supplies a desktop two-input, two-output thermal plant on which ARX and other empirical models are identified and then used for control [14], and the GEKKO Optimization Suite provides the `arx` and `state_space` objects into which such models are loaded for simulation, estimation and predictive control [15].

12.3 Identification Problem and Data Model

The object identified is a linear time-invariant map from m inputs to p outputs on a uniform sample grid. The inputs are the MVs followed by the measured DVs, and the outputs are the CVs. An unmeasured disturbance and sensor noise enter the record but not the model, and every quantity the algorithms touch is labeled in Figure 12.1, using the distillation column record as the instance.

12.3.1 Signals and Roles

Every column of an imported table receives one of five roles, Time, MV, DV, CV or Ignore, together with a display name, units and a description. An MV is an input an operator or controller can change, a DV is a measured input nobody adjusts, and a CV is an output to predict. MVs and DVs both enter the input vector, MVs first, and keep their identity in every view and export. The application proposes roles from header text, such as `valve`, `Q1` or `level`, and marks each proposal as a guess until the user confirms or changes it. A column cannot be both an input and an output. During identification an MV is a recorded trajectory and never a decision variable. Table 12.1 classifies the signals of the two records used most in what follows.

12.3.2 Preparation of the Record

Raw records are kept as text and never edited, and every transformation is a field of a stored recipe, so a reopened project reproduces the same prepared data. Every model uses one uniform sample period T_s, by default the median recorded interval. MVs are held between samples as an actuator is. CVs and DVs are resampled by default with the mean over the preceding interval $(t - T_s, t]$, a causal anti-alias filter, or by linear interpolation or hold on request. Linear interpolation never reaches forward into a different split, and no method bridges an interval longer than the gap limit of three sample periods. A gap, an excluded stretch, a return to an earlier split or a record boundary starts a new segment with its own initial state, so two campaigns are never joined into one trajectory and no model history runs from held-out data back into training. Adjacent training, validation and test stretches in chronological order stay in one segment and keep their causal input history. Duplicate stamps are averaged, rows out of order are sorted, and both are reported.

Splits are chronological and never shuffled. The default marks the first 60 % of each record as training, the next 20 % as validation and the last 20 % as the final test, and a record shorter than 300

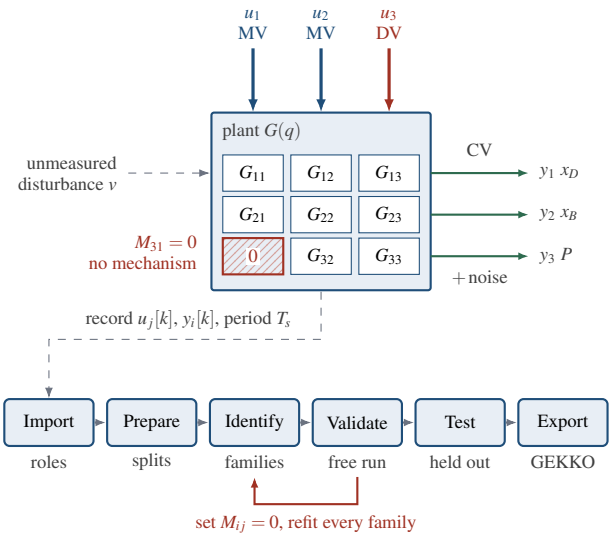


Figure 12.1 Multivariable plant as the identification sees it, and the six identification steps of the workflow, which a seventh step that builds a predictive controller from the chosen fit follows (Section 12.5). Two MVs and one measured DV enter the transfer matrix $G(q)$, whose rows are the CVs, and an unmeasured disturbance and sensor noise reach the record without entering the model. For the column record u_1 is reflux, u_2 steam and u_3 feed rate. The hatched cell is a path fixed at zero from process knowledge, which every family enforces exactly and which the validation step can contradict.

Table 12.1 Classification of Signals in the Two-Heater and Column Records. Ranges Are Those of the Recorded Test Signals.

Signal	Role	Units	Recorded range
TCLab two-heater test, $T_s = 5$ s			
Q_1 heater 1	MV	%	0 to 80, held 180 to 480 s
Q_2 heater 2	MV	%	0 to 80, held 180 to 480 s
T_1, T_2	CV	°C	measured, 0.32 °C steps
distillation column, $T_s = 1$ min			
R reflux	MV	lb/min	1.75 to 2.15, held 25 to 70 min
S steam	MV	lb/min	1.55 to 1.85, held 25 to 70 min
F feed rate	DV	lb/min	2.2 to 2.7, held 60 to 160 min
x_D distillate	CV	mol%	near 96
x_B bottoms	CV	mol%	near 0.5
P pressure	CV	bar	near 1.2
model quantities, both records			
$x[k]$	latent state	—	not a physical inventory
$e_i[k]$	residual	CV units	unmeasured
M_{ij}	structural mask	—	1 estimated, 0 fixed at zero

samples receives a 70 % and 30 % training and validation split with no test. Splits are marked by dragging across a chart or typing exact times. Each mark is a translucent band named along its length in its own color, overlaps blend and stay visible, the later mark decides, and a ribbon along the top of each chart shows the resolved split (Figure 12.2). A zoom strip under the charts starts on the whole record, so a short stretch of a long record can be marked. Final-test scores stay hidden until a separate Test step reveals them for the selected current model, and every inspection is recorded. Baselines $\bar{u}_j$

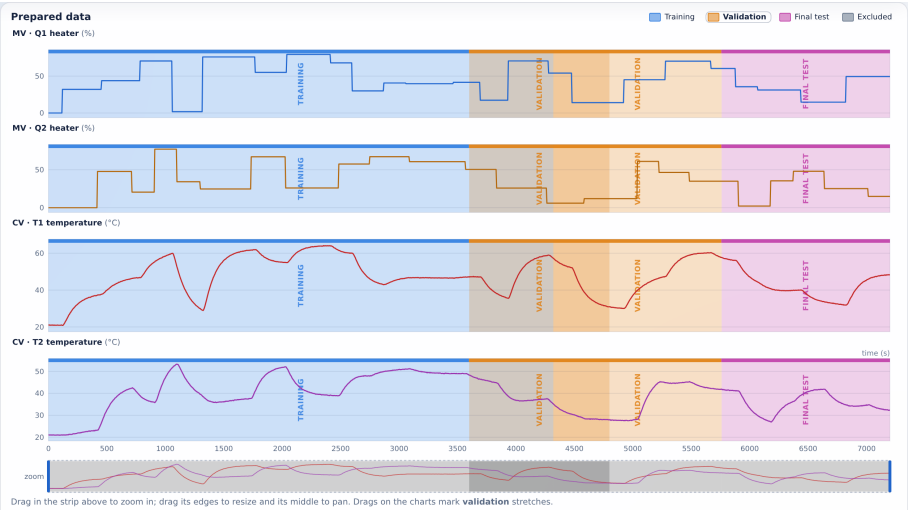


Figure 12.2 The Prepare step on the TCLab record after a validation mark from 3600 to 4800 s was added over the default splits (training to 4320 s, validation to 5760 s, final test to 7200 s). Each mark is a translucent band named along its length. Where the new mark overlaps training the two colors blend, and where it overlaps the default validation band the color deepens. The ribbon along the top of each chart shows the resolved split, validation from 3600 s because the later mark decides, and the zoom strip under the charts shows the whole record.

and $\bar{y}_i$, scales $\sigma_{u,j}$ and $\sigma_{y,i}$, optional drift slopes and optional outlier limits are computed on training samples only. A readiness summary reports, before any fit, inputs that do not move, inputs that move together (level correlation above 0.9 or move correlation above 0.8 for limited confidence, 0.995 and 0.98 for cannot estimate), channels pinned at a limit, outputs that move with no input moving, too many coefficients per training sample, dynamics slower than any training stretch, closed-loop records, and an input that tracks a CV so closely (correlation above 0.995) that it may be derived from it.

12.3.3 Model Families

All five families work in deviation variables about the training baselines, $u'_j[k] = u_j[k] - \bar{u}_j$ for each input and $\hat{y}_i[k] = \bar{y}_i + \delta_i[k]$ for each CV. The two transfer-function families give each active path a gain K_{ij}, one or two lags and a dead time θ_{ij}, as shown in Equation (12.1), the forms of Equations (1.15) and (1.11) in Chapter 1.

$$G_{ij}(s) = \frac{K_{ij}\,\mathrm{e}^{-\theta_{ij}s}}{\tau_{ij}s+1} \quad \text{or} \quad \frac{K_{ij}\,\mathrm{e}^{-\theta_{ij}s}}{\tau_{ij}^2 s^2 + 2\zeta_{ij}\tau_{ij}s+1} \tag{12.1}$$

where τ_{ij} is a time constant in record time units and ζ_{ij} is a damping factor. For $\zeta_{ij} \geq 1$ the second-order path has two real lags $\tau_{ij}(\zeta_{ij} \pm \sqrt{\zeta_{ij}^2 - 1})$, and for $\zeta_{ij} < 1$ it rings. The CV is the sum of its active paths plus an offset c_i. Any path inside the second-order candidate can be held first order, and any path can be given a gain sign, bounds on its lag and dead time, or fixed parameter values.

The ARX family writes each CV as a difference equation in its own history and every active input, as shown in Equation (12.2), of which Equation (1.20) of Chapter 1 is the case of one input with

$n_a = n_b = 1$.

$$\begin{aligned}\delta_i[k] = c_i + \sum_{l=1}^{n_a} a_{i,l}\,\delta_i[k-l] \\ + \sum_j \sum_{m=0}^{n_b-1} b_{i,j,m}\,u'_j[k-d_{ij}-m] + e_i[k]\end{aligned} \tag{12.2}$$

where n_a and n_b are the output and input orders, $d_{ij} \geq 1$ is an explicit dead time in samples (zero only when same-sample feedthrough is enabled) and e_i is the equation error. The output polynomial is diagonal, with no cross-output terms, so the transfer matrix entry is $B_{ij}(q)/A_i(q)$ and removing a path is exactly removing its numerator coefficients.

The FIR family has no assumed shape, as shown in Equation (12.3).

$$\delta_i[k] = c_i + \sum_j \sum_{l=0}^{H-1} h_{i,j,l}\,u'_j[k-d_{ij}-l] \tag{12.3}$$

where H is the horizon in samples and $h_{i,j,l}$ are the impulse-response coefficients. A step response is the running sum of $h_{i,j,l}$, and a horizon long enough to settle carries one coefficient per sample per path.

The state-space family shares a small latent state among all inputs and outputs, as shown in Equation (12.4).

$$\begin{aligned}x[k+1] = Ax[k] + B\tilde{u}[k], \qquad \tilde{u}_j[k] = u'_j[k-d_j] \\ \delta[k] = Cx[k] + D\tilde{u}[k] + c\end{aligned} \tag{12.4}$$

where d_j is a whole-sample delay per input and $D = 0$ unless feedthrough is allowed. A state block too slow for the record also adds a constant drift to its state equation (Section 12.4.5). Every family is scored with the normalized root-mean-square error of Equation (12.5).

$$\mathrm{NRMSE}_i = \frac{1}{\sigma_{y,i}} \sqrt{\frac{1}{N} \sum_k \left(y_i[k] - \hat{y}_i[k]\right)^2} \tag{12.5}$$

where the sum runs over the N samples of one split and $\sigma_{y,i}$ is the training standard deviation, so every split and every candidate is measured on one scale. The aggregate is the mean of NRMSE_i over the CVs. RMSE, MAE and bias in engineering units and a fit $100(1 - \|y_i - \hat{y}_i\|/\|y_i - \bar{y}_i\|)$, which is negative for a model worse than the split mean and undefined for a constant output, are reported beside it.

12.3.4 Method Settings

The application has no hidden plant parameters, because the plant is whatever record is loaded. Its parameters are the method settings in Table 12.2, where Status marks what the Methods panel and the Complexity selector expose, what lies behind Advanced, and what is fixed in the code.

The Complexity selector changes four of these settings together. The simple level uses 2 starting points, ARX orders up to 2 and 2 and state order up to 4, and the detailed level uses 6 starting points, ARX orders up to 5 and 6 and state order up to 12. The standard level is the default of Table 12.2.

12.3.5 Assumptions

The models rest on seven assumptions, and each one fails somewhere.

Linearity and time invariance. Every family is linear about the training baselines. A plant whose gain changes with the operating point receives an average over the training range, and extrapolation outside that range is never validated. The TCLab record is generated from an energy balance with a radiative term, so its identified gains are averages of a gain that changes with temperature.

Table 12.2 Method Settings and Defaults. S Is the Longest Training Stretch in Time Units, at Least a Quarter of the Longest Segment and $10T_s$, N Is the Number of Training Samples and N_s the Longest Segment in Samples.

Setting	Default	Status
preparation		
sample period T_s	median recorded interval	visible
CV and DV resampling	interval mean	visible
gap limit	$3T_s$	visible
splits per record	60/20/20, under 300: 70/30	visible
drift removal, outlier limit	off, 0	visible
transfer functions		
starting points per CV	4 (+ previous fit)	Advanced
iterations, scouting then total	8, then 60	Adv., fixed
bounds on $\ln\tau$	$\ln(T_s/20)$ to $\ln 3S$	per path
bounds on θ, ζ	0 to $S/2$, 0.1 to 28.3	per path, fixed
LM start damping, cost and step tolerances	10^{-3}, 10^{-9}, 10^{-8}	fixed
delays, ARX and FIR		
impulse onset, delay scan	8 % of peak, onset -3 to $+1$	fixed
orders searched n_a, n_b	1 to 3, 1 to 4	Advanced
parsimony in order search	1 %	fixed
ridge, simulation refinement	0, on	Advanced
refinement iterations, tolerance	30, 10^{-7}	fixed
FIR horizon H	$N/(3m)$, 10 to $\min(240, 720/m)$	Advanced
FIR penalty λ	7 values, 10^{-6} to 1	fixed grid
state space		
largest order per block	8	Advanced
Hankel horizon i	$\max(n_{\max}+2, 8)$, $2i(m+l) \le 480$	Advanced
refinement iterations, tolerance	25, 10^{-8}	fixed
feedthrough D	off	Advanced
starting rule, scoring and checks		
measured start when $\lvert z\rvert \ge$	$1 - 5/N_s$	fixed
unstable only when $\lvert z\rvert >$	$1 + 1/N_s$	fixed
warm-up samples not scored	3 per segment	fixed
CV weights	$1/\sigma^2_{y,i}$	Advanced
ranking parsimony	2 %	fixed
zero-path horizon	$\max(1000, 4\times$ response$)$	fixed
bootstrap replicates, block	30, $\max(10, 2N^{1/3})$	20 to 100, fixed
refit pause after an edit	650 ms	fixed

Zero-order hold. Inputs are taken as constant between samples. This is exact for an MV written by a controller and an approximation for a measured DV, which varies within the interval. The interval-mean resampling of a DV partly compensates by averaging over that interval.

Steady state at each segment start, unless the dynamics are too slow. A free run begins with every input held at its first value for all earlier time and the model at the matching steady state, so a record that starts mid-transient puts that transient into the early residual. An ARX output or state block slower than a fifth of the longest segment, an integrator included, starts from its first measurements instead (Section 12.4.7).

No noise model. The transfer-function, FIR and state-space families model no noise, so their predictor is their simulation. ARX starts from the equation-error least-squares estimate, the one-step fit of Section 1.4.8, which is biased when the disturbance is colored, and by default refines it on simulation error, which makes its final estimate an output-error fit with ARX polynomials. Output-error, ARMAX and Box-Jenkins noise models and instrumental variables are not implemented.

Open-loop excitation unless declared otherwise. Under feedback the MVs are correlated with the

disturbance through the controller, and every estimator here can be biased toward the inverse of the controller unless external excitation such as a setpoint step dominates the input [12]. A closed-loop declaration adds a readiness finding and changes how the residual-input correlation is described, but it does not change the estimator.

Integrating paths have no finite gain. The transfer-function families contain no integrator, and a long lag looks like one over a short record (Section 1.4.4). A path whose fitted time constant exceeds the longest training stretch S is reported as integrating with its largest step-response slope rather than with a gain the data never saw settle. ARX and state space can hold the integrator itself, and a pole within T_s/S of the unit circle is reported the same way.

A zero path is a structural claim. Fixing $M_{ij} = 0$ asserts that no mechanism connects input j to CV i over the whole operating range. It is not a statistical result, and the validation score is the only evidence the application can offer for or against it.

12.4 Identification Algorithms as Implemented

12.4.1 Exact Sampling with Fractional Dead Time

Every transfer-function path is sampled exactly for a held input, with no Padé approximation and no rounding of the dead time. The dead time is split as $\theta = dT_s + \phi$ with $0 \le \phi < T_s$, so over one interval the delayed input equals $u[k-d-1]$ for the first ϕ and $u[k-d]$ for the remaining $T_s - \phi$, as shown in Equation (12.6).

$$\begin{aligned} x[k+1] &= \Phi x[k] + \Gamma_1 u[k-d-1] + \Gamma_0 u[k-d] \\ y[k] &= x_1[k] \\ \Phi &= \mathrm{e}^{AT_s}, \qquad \Gamma_0 = \int_0^{T_s-\phi} \mathrm{e}^{As}\,\mathrm{d}s\,B \\ \Gamma_1 &= \mathrm{e}^{A(T_s-\phi)} \int_0^{\phi} \mathrm{e}^{As}\,\mathrm{d}s\,B \end{aligned} \tag{12.6}$$

where the unit-gain continuous realization is $A = -1/\tau$, $B = 1/\tau$ for first order and, on the states y and $\tau\dot{y}$ for second order, is given by Equation (12.7).

$$A = \begin{bmatrix} 0 & 1/\tau \\ -1/\tau & -2\zeta/\tau \end{bmatrix}, \qquad B = \begin{bmatrix} 0 \\ 1/\tau \end{bmatrix} \tag{12.7}$$

Each pair e^{Ah} and $\int_0^h \mathrm{e}^{As}\mathrm{d}s\,B$ comes from one exponential of the augmented matrix $\left[\begin{smallmatrix} Ah & Bh \\ 0 & 0 \end{smallmatrix}\right]$, computed by a degree-6 Padé approximant with scaling and squaring. The path output is K times $y[k]$. A step applied at $k = 0$ therefore first appears in the sampled output at $k = d + 1$, the first sample after the dead time, and θ keeps its continuous meaning without rounding. Equation (12.6) is also the realization written into the export, so the fitted model and the exported model are the same object.

12.4.2 Transfer Functions by Variable Projection

All active paths into one CV are fitted together. For fixed lags, damping factors and dead times, the free-run output is linear in the gains and the offset, so those are solved exactly by least squares inside the residual function and the outer search moves only the nonlinear parameters, which is the variable projection idea for separable problems [16]. With $s_{ij}[k]$ the unit-gain response of path j to the recorded u'_j for nonlinear parameters p, the reduced residual is given by Equation (12.8).

$$r_i(p) = \sqrt{w_i}\, W^{1/2} \left(\delta_i - c_i^{\star}(p) - \sum_j K_{ij}^{\star}(p)\, s_{ij}(p) \right) \tag{12.8}$$

where δ_i is the training target, W holds the per-sample weights (zero for a down-weighted outlier), $w_i = 1/\sigma_{y,i}^2$ is the CV weight, and $K_{ij}^{\star}$ and $c_i^{\star}$ are the least-squares gains and offset for that p. The

nonlinear vector holds $\ln \tau_{ij}$, a damping coordinate q_{ij} (second order) and θ_{ij} for every parameter not fixed by the user, with the damping given by Equation (12.9).

$$\zeta_{ij} = \begin{cases} \mathrm{e}^{q_{ij}}, & q_{ij} \le 0 \\ 1 + \sinh q_{ij}, & q_{ij} > 0 \end{cases} \tag{12.9}$$

The coordinate is monotone with unit slope on both sides of $\zeta = 1$, linear where a path rings and nearly logarithmic in the ratio of the two real lags, which grows as e^{2q}. Its bounds $q = \ln 0.1$ and $q = 4$ give ζ from 0.1 to 28.3, a lag ratio of about 3200, so a path whose second lag the data cannot see runs down a straight valley to the upper bound and stops, whereas in (τ, ζ) coordinates the same path crept along a curved valley to its iteration limit. A gain sign constraint is enforced by an active set that pins a gain of the forbidden sign at zero and re-solves the others. Each path's response column is cached by its exact parameters, so a difference step on one path re-simulates only that path.

The gains come from cached weighted inner products of the path columns by an equilibrated Cholesky factorization, and the Jacobian is the variable projection Jacobian of Golub and Pereyra [16] rather than a least-squares re-solve per parameter. For a parameter of path j with column derivative D, one forward difference of that column alone, the change of the fit is given by Equation (12.10).

$$\begin{aligned} \partial f &= K_j D + A\, \partial c \\ \partial c &= M^{-1} \left(e_j \langle D, We \rangle - K_j A^{\mathsf{T}} W D \right) \end{aligned} \tag{12.10}$$

where A holds the free columns and the constant, $M = A^{\mathsf{T}} W A$, c the solved gains, e the misfit and e_j the unit vector of path j. A fixed gain keeps only the second term inside the parentheses, and a pinned gain contributes nothing. Tests check (12.10) against differences of the residual to 10^{-4} relative and check that it reaches the same minimum as a plain route that re-solves the gains by QR for every perturbation.

The outer search is a bounded Levenberg-Marquardt method [17] with (12.10) as its Jacobian and an active set at the bounds. A parameter on a bound, within 10^{-9} of its span, whose gradient would carry it outside is held for the iteration and the step is computed for the others. Clipping a joint step instead, as the first release did, spoils the step of every other parameter and lets a search creep to its limit. Differences use a step of $10^{-6} \max(|p|, 1)$, at most 1 % of the bound span and inward at an upper bound. The damped normal equations add $\lambda \max(d_{aa}, 10^{-12} d_{\max})$ to each diagonal entry of $J^{\mathsf{T}} J$ and are solved by Cholesky, with λ starting at 10^{-3}, divided by 5 after an accepted step and multiplied by 8 after a rejected one (10 if the factorization fails), up to 12 attempts per iteration, and trial points projected onto the bounds. The search stops on a relative cost decrease below 10^{-9}, a relative step $|\Delta p| / (|p| + 1)$ below 10^{-8}, a vanishing gradient, every free parameter held, or no damping giving a decrease. A parameter that ends within 10^{-6} of its bound span from a bound is reported in words, except a dead time of zero, which is simply a path without dead time.

Starting points come from a regularized FIR fit with the automatic horizon. Each path's FIR step response gives a gain from its last 15 % and a lag from its 63.2 % crossing, and the dead time is $\theta_{ij} = (o_{ij} - 1) T_s$, where the onset o_{ij} is the FIR delay plus the first lag whose impulse coefficient reaches 8 % of the largest. The onset is right for an integrating path, whose 5 % step crossing lies far up its ramp, where the first release put its dead time. For second order, an overshoot above 5 % gives ζ from the logarithmic decrement and τ from the time to peak, and otherwise $\zeta = 1$ and τ is the first-order lag divided by 2.15. Four multistarts scale the lag and dead time and shift the damping coordinate by $(1, 1, 0)$, $(0.4, 1, 1.5)$, $(2.5, 1, 0)$ and $(1, 0.5, -0.7)$, and a refit adds the previous valid model as a fifth. Every start receives 8 scouting iterations and the best continues for the remaining 52 of its 60, which is what lets a record of overlapping random steps, rather than isolated clean steps, be fitted at all.

12.4.3 Delays, then ARX by QR Least Squares

Each CV's recurrence (12.2) is solved by streaming the rows into a Givens QR factorization with column equilibration and a rank decision on the diagonal of R at 10^{-11} relative, so a regressor that carries no information is dropped instead of receiving a large coefficient. A row is used only when its target and every lagged output lie in the training split.

The dead times d_{ij}, shared by ARX and state space, start from the onsets o_{ij} and are refined for each CV by a one-step loss. With the other inputs at their current delays, every candidate from $o_{ij}-3$ to $o_{ij}+1$ is scored by the weighted residual sum of squares of a one-step ARX with $n_a = n_b = 2$, in two passes over the inputs. The candidates share one parameter count, so no validation data are needed, a one-step loss is not inflated by an integrator, and one factorization per input covers all of its candidates as column subsets of the triangular factor.

All twelve order pairs with $n_a \le 3$ and $n_b \le 4$ are solved from one factorization of the largest pair, each smaller pair by a small QR of the matching columns of its R plus the few rows that a shorter output history makes usable after an excluded stretch. Each pair is simulated once from the starting rule of Section 12.4.7 and scored on training and validation, a pole more than $1/N_s$ outside the unit circle scores as infinite, and a pair replaces the best so far only if it lowers the validation NRMSE by 1 %, or at all with no more coefficients.

The chosen pair is then refined on simulation error by default, by an unbounded Levenberg-Marquardt search over every a, b and c of the CV with the exact Jacobian of the sensitivity recursion (12.11).

$$s_\vartheta[k] = \sum_{l=1}^{n_a} a_{i,l}\, s_\vartheta[k-l] + r_\vartheta[k] \tag{12.11}$$

where $s_\vartheta = \partial\hat{\delta}_i/\partial\vartheta$, r_ϑ is the regressor of coefficient ϑ in the free run ($\hat{\delta}_i[k-l]$ for $a_{i,l}$, the lagged input for a b coefficient, 1 for c_i), and the history before the record is the derivative of the starting value, zero for a measured start. Tests check it against differences for both starting rules. The search runs for at most 30 iterations at a cost tolerance of 10^{-7}, and its result is kept only if the training free-run error does not rise and no pole moves more than $1/N_s$ outside the circle. On the bundled records it converged in 3 to 24 iterations per CV. A pole within T_s/S of the circle makes a path integrating, with the ramp slope $B_{ij}(1)/(T_s\tilde{A}_i(1))$ where $A_i(q) = (1-q^{-1})\tilde{A}_i(q)$ and $\tilde{A}_i(1) = \sum_l l\, a_{i,l}$. ARX alone also reports one-step and h-step predictions ($h = 10$ by default), each labeled with the measured-output advantage it carries.

12.4.4 FIR with Penalties Chosen on Validation

The FIR regression (12.3) is formed in normalized units, with inputs divided by $\sigma_{u,j}$ and the CV by $\sigma_{y,i}$, accumulated as normal equations, and solved by Cholesky with the penalty of Equation (12.12).

$$\min_h \|\delta_i - Xh\|^2 + \lambda\bar{s}\left(0.1\|h\|^2 + 10\|D_2 h\|^2\right) \tag{12.12}$$

where D_2 takes second differences $h_l - 2h_{l+1} + h_{l+2}$ within each path, $\bar{s}$ is the mean diagonal of $X^\mathsf{T}X$ so that λ is dimensionless, and λ is chosen from $10^{-6}, 10^{-5}, \ldots, 1$ by the lowest validation mean-square error. Every regressor is a shifted copy of one input, so each entry of $X^\mathsf{T}X$ is a lagged cross product. A sum then follows from its predecessor on the same diagonal with two end corrections per run of equally weighted rows, and only the first sum of each diagonal takes a full pass, which turns about $Np^2/2$ multiplications into about Nmp for m inputs. When per-sample weights break the training rows into runs shorter than the horizon on average, the rows are accumulated one by one instead. The horizon is $N/(3m)$ samples, clipped to between 10 and $\min(240, 720/m)$, which holds the coefficient count below about 720 per CV so the normal equations stay interactive. A FIR fit notes a path whose step response still moves by more than 10 % of its largest value over the last fifth of the horizon, because its gain is then a truncated sum. A zero path's coefficients are simply absent from the regressors.

12.4.5 Subspace State Space with Prediction-Error Refinement

The state-space fit proceeds in four steps. Each input first receives one delay d_j, the shortest refined delay d_{ij} over the CVs it drives less the one sample a strictly proper model already has. A past-output MOESP step then stacks future inputs, past inputs, past outputs and future outputs of every training run into block Hankel rows with horizon i, each channel scaled by its training standard deviation, and an LQ factorization (a streamed QR) isolates the block that relates past data to future outputs after the future inputs are projected out [10]. A one-sided Jacobi SVD of that block suggests the order, and $\Gamma = U_1 S_1^{1/2}$ gives C from its first block row and A from shift invariance by least squares.

An integrating mode lands on either side of the unit circle by noise, so when the spectral radius ρ of A exceeds 1, A is scaled by $1/\rho$, which puts that mode on the circle and moves the others by the same small factor. With A and C fixed, B, D and the offsets c are linear in the free-run output and are solved jointly by least squares on every training sample of the block, using the state-sensitivity recursion $S_j \leftarrow AS_j + Iu_j[k]$ from its steady state. A block whose slowest mode has $|z| \geq 1 - 5/N_s$ starts instead from Equation (12.13).

$$x_0 = G\,(y_N - Z_N \vartheta)\,, \qquad G = \left(\Gamma_N^{\mathsf{T}} \Gamma_N\right)^{-1} \Gamma_N^{\mathsf{T}} \tag{12.13}$$

where y_N stacks the block's first $N = 2n$ measured outputs, Γ_N stacks $C, CA, \ldots, CA^{N-1}$, and $Z_N \vartheta$ is the response of the unknowns $\vartheta = (B, D, c)$ over those samples from zero state, so the output stays linear in ϑ. Such a block also receives a constant state drift, solved as the B column of a virtual input fixed at 1, for the integrated mean of its inputs that an output offset cannot carry. Orders 1 through $\min(8, il - 1)$ are scored on validation free run, an order with a pole more than $1/N_s$ outside the circle is rejected, and the 1 % parsimony rule of ARX picks the order, while the largest log gap among the singular values is only displayed. A block of order n with m inputs and l outputs carries $n^2 + nm + ln + l$ parameters, plus n for a drift.

The chosen model is then refined on the weighted free-run error of the training samples over every entry of A, B, C, the offsets, a drift and, when enabled, D, with 25 Levenberg-Marquardt iterations at a tolerance of 10^{-8}. Trial models with spectral radius of at least 1.2 are rejected inside the search, and the result is kept only if the training error does not rise and no pole lies more than $1/N_s$ outside the circle. The refinement is skipped and reported when the parameter count n_p and residual count N_r make $n_p N_r (n^2 + nm + 1)$ exceed 1.5×10^9 or $N_r n_p^2$ exceed 10^9, which is the interactive cost budget.

12.4.6 Exact Zero Paths in Every Family

A zero entry $M_{ij} = 0$ removes the entire deterministic path, delayed effects and feedthrough included, from every candidate family. In the transfer-function families the path and its parameters leave the problem. In ARX and FIR its coefficients are not regressors, and the remaining coefficients of that CV are re-estimated. A state-space model cannot be treated by deleting one entry of B, C or D, because the input can reach the output through shared states, and the constraint that matters is on the whole transfer entry, as shown in Equation (12.14).

$$D_{ij} = 0, \qquad C_i A^k B_j = 0 \quad \text{for } k = 0, 1, \ldots, n-1 \tag{12.14}$$

where C_i is row i of C and B_j is column j of B. By the Cayley-Hamilton theorem the first n Markov parameters vanishing makes every later one vanish in exact arithmetic, which is the realization argument of Ho and Kalman applied to one entry [8]. The application satisfies (12.14) by construction. Outputs are grouped by their set of allowed inputs, each group receives its own block of states driven only by those inputs, and the full model is block diagonal, so a prohibited input has an identically zero column in the block that its CV reads. One group, with no zeros, is the fully coupled model. The price is that dynamics shared across blocks are not modeled and the total order can rise, which the candidate's notes state.

🕮 **Pause.** Try Exercise 12.2 before reading on.

Every zero path is then verified end to end rather than by inspecting a coefficient. The whole fitted model is simulated with every input at its baseline, again with a unit impulse at $k = 1$ on input j, and again with a unit step from $k = 1$, over $\max(1000, 4n_r)$ samples where n_r is the response horizon, and the CV's trajectories must agree exactly, to 0.0. For state space the Markov parameters of (12.14) and D_{ij} are checked as well. A zero that fails is labeled as a violated constraint and the candidate becomes ineligible for ranking and export. Each matrix cell carries one of nine statuses in words and an icon, among them fixed at zero, small estimated gain (an effect at one input standard deviation below 2 % of the CV standard deviation), cannot estimate (an input with fewer than two moves in training), integrating, unstable, failed and out of date, so a small gain is never displayed as a zero. By default each cell draws its unit-step response in its own frame, with the vertical extent and the time span written in the cell and the span running until that response settles, so a path with a small gain or a fast response is not a flat line in a corner. A shared scale per CV, a scale in CV standard deviations and a shared horizon remain as view settings.

12.4.7 Initialization, Ranking and Uncertainty

Every free run starts from the steady state for the first inputs of its segment, held before the record begins, unless the model is too slow to show one. An ARX output or a state block whose slowest pole has $|z| \geq 1 - 5/N_s$, a time constant beyond a fifth of the longest segment, starts from its measurements, an ARX history at its first measurement and a state block at (12.13). The formal steady state of such a model is an extrapolation far beyond the record and was one reason the first release's ARX and state-space models drifted on the surge tank. The flag is set at fit time and stored in the model, so the export applies it too. An ARX with an exact unit root also holds its first measurement, and an unmarked block with $I - A$ singular starts at zero. This one convention makes the five families comparable and is what the GEKKO export reproduces. The first 3 samples of each segment are not scored.

Candidates are ranked only if they are current, meaning computed from the present data, splits and mask, without a pole more than $1/N_s$ outside the unit circle (an integrator counts as stable, and FIR has no poles) and with every zero verified exactly. Among those, the lowest aggregate validation NRMSE of the free run leads, and a candidate within 2 % of the best with fewer parameters is preferred. Two baselines frame every ranking, persistence (the first measurement held, or the previous one for one-step prediction) and a static regression $\hat{y}_i = \bar{y}_i + c_i + \sum_j K_{ij} u'_j$ under the same mask. Residual autocorrelation and residual-input cross-correlation over 25 lags are computed within runs of one split and never across a gap, against the $\pm 1.96/\sqrt{N}$ band.

Uncertainty comes from a moving-block bootstrap, because residuals of a dynamic model are autocorrelated and resampling single samples would treat them as independent draws [18]. The free-run residuals are cut into blocks of $\max(10, \mathrm{round}(2N^{1/3}))$ consecutive samples, which is 19 for the 864 training samples of the TCLab record. Blocks are drawn only from the training stretches of each segment, a source stretch chosen in proportion to its length and a start chosen so the block stays inside it, and they replace the residual on training samples only, so no validation or test residual enters a replicate. Each replicate refits the fitted model's own structure, with each CV's ARX orders and delays, each CV's FIR horizon and chosen penalty, each state block's order, the input delays and the mask held at their fitted values. Transfer-function lags and dead times are re-estimated, starting from the fitted values. The default is 30 replicates with a fixed seed, and the 5th, 50th and 95th percentiles of every gain and the pointwise 5th and 95th percentiles of every step response are reported. The application's test of this rule adds a sinusoid of amplitude 500 to the validation and test outputs of a single-path record with a true gain of 2 and requires the 5th to 95th percentile gain interval of 12 replicates to stay between 1.8 and 2.2, which the ARX, FIR, state-space and first-order families all satisfy.

12.4.8 Worker Protocol and Stale Results

Fitting runs in a browser module worker. The prepared data are posted once with a data key, and each job carries only settings, the mask and a revision key. A result is published only if both keys still match the interface state, so a fit that finishes after a newer edit can never be shown as current, and the last valid result of an older revision is displayed as out of date and is excluded from ranking and export. Starting a job cancels the one running, and cancellation terminates the worker and starts a fresh one with the data re-posted, so nothing from a cancelled job can arrive later. A structural edit saves an undoable revision, marks every candidate out of date and refits the family on screen after a 650 ms pause, which batches rapid edits.

12.4.9 Export Realization and GEKKO Adapters

Every family is written as one discrete realization $x[k+1] = Ax[k] + Bu'[k]$, $y[k] = y_0 + Cx[k] + Du'[k]$ with typed states. Dynamic states carry lags, oscillations and ARX output history, register states form shift registers that realize whole-sample dead times exactly, and a constant state fixed at 1 carries the ARX constant term and the drift of a measured-start state block. Registers start at the first input, the constant at 1, measured-start states as in Section 12.4.7, and the remaining dynamic states at the solution of $(I - A_{dd})x_d = A_{d,o}x_o + B_d u'[0]$, with the fallbacks described there. A measured-start block is solved last, by Equation (12.13) with the response of every other state as the known part, and the realization carries the per-state flags and the per-block sample counts that this needs. The generated `model.py` needs only numpy and GEKKO. It simulates the realization by a numpy recursion and compares it with the application's free-run predictions at a tolerance of $10^{-8}\max(1, \sigma_{y,i})$, builds the same model in GEKKO with `remote=False` in simulation mode (`IMODE` 4, `NODES` 2, `RTOL` and `OTOL` 10^{-10}) with the MVs and DVs supplied as recorded values and never as decisions, compares GEKKO with numpy at $10^{-5}\max(1, \sigma_{y,i})$, recreates every step-response cell, checks that every zero path is exactly zero, and exits with an error status if any check fails [15].

Two GEKKO conventions were measured rather than assumed. The discrete `state_space` object takes the value assigned to each state at the first time point as $x[0]$ and does not evaluate the output equation at that point, so the script computes $y[0] = y_0 + Cx[0] + Du'[0]$ itself. The `arx` object implements Equation (12.15) with zero history before the first point.

$$y_i[k] = \sum_{l=1}^{n_a} a_{l,i}\, y_i[k-l] + \sum_g \sum_j b^{\mathrm{G}}_{i,g,j}\, u_j[k-g-1] + c^{\mathrm{G}}_i \tag{12.15}$$

The adapter maps the canonical coefficients by $b^{\mathrm{G}}_{i,g,j} = b_{i,j,g+1-d_{ij}}$, so row g of the GEKKO input matrix is lag $g+1$, and it works in deviation-from-start form, with inputs $u'_j[k] - u'_j[0]$, outputs measured from the output history y^s_i that the application starts from, steady or measured, and the constant $c^{\mathrm{G}}_i = c_i - (1 - \sum_l a_{i,l})(y^s_i - \bar{y}_i) + \sum_g \sum_j b^{\mathrm{G}}_{i,g,j} u'_j[0]$. Zero history in that form is the start of the application. An ARX with same-sample feedthrough uses the state-space route instead.

12.5 Predictive Control from the Fitted Model

A seventh step uses a chosen fit twice, as the internal model of a model predictive controller (MPC) and as the plant of a closed-loop simulation, and tunes a decentralized PI from the same fit for comparison. The dynamic-matrix form (DMC) predicts with the fitted step responses and a constant output bias, the form of most industrial packages [2], posed with input bounds as a quadratic program [19]. The state-space form predicts with the exported realization and a Kalman filter. Both minimize one objective under one set of bounds, so a difference between them is one of prediction and estimation. The objective has the form of Equation (1.29) in Chapter 1 without slack terms, and the state-space form's estimator is a steady-state Kalman filter with added disturbance states, as in Section 1.6.3.

12.5.1 Controller Model, Objective and Constraints

The controller works in the deviation variables u'_j and $y'_i = y_i - \bar{y}_i$ with the fit's record constants (ARX constant, state drift, offset) dropped, so rest is an equilibrium of every family. The first CVs in role order, as many as the MVs, receive setpoints y_i^{sp}, and the realization is pruned to the states that reach them, which keeps the 270 min dead time of the spurious pressure path (Table 12.5) out of the column's controller. The controller reads the fit, $\sigma_{y,i}$, the range each MV covered in the records and, at each sample, the measured CVs, setpoints and DVs, never a record's generator. In the nominal cases the simulated plant is the fit itself, so only the cases with the MV gains scaled by 0.7 and 1.5 or with another family's fit as the plant measure model error. At every sample k both forms choose input levels $v_{j,\nu}$ on move blocks $\nu = 1, \ldots, N_b$ over a horizon of N_p samples, as shown in Equation (12.16).

$$\begin{aligned} \min_{v} \; & \sum_{\ell=1}^{N_p} \sum_{i \in \mathscr{C}} w_i \left(\hat{y}'_i[k+\ell] - y_i^{\mathrm{sp}} \right)^2 \\ & + \sum_{\nu=1}^{N_b} \sum_{j} R_j \left(v_{j,\nu} - v_{j,\nu-1} \right)^2 \\ \text{s.t. } & u_j^{\min} \le v_{j,\nu} \le u_j^{\max}, \qquad R_j = \lambda_u \textstyle\sum_{i \in \mathscr{C}} w_i K_{ij}^2 \end{aligned} \tag{12.16}$$

where $\mathscr{C}$ holds the controlled CVs, $w_i = 1/\sigma_{y,i}^2$ is the weight of Equation (12.8), $v_{j,0}$ is the input applied now, K_{ij} is the fitted gain and $\lambda_u = 1$ makes a move cost the weighted steady error it could remove. The twelve blocks last $1, 1, 1, 1, 2, 2, 3, 4, 6, 9, 15$ samples and then the remainder, and only the first is applied before a dense dual active-set solver repeats the problem at $k+1$. The default horizon $N_p = T_{\mathrm{set}}/T_s$, with T_{set} the slowest 2 % settling time of an MV-to-CV step response, was fixed before any loop was closed. The application declares no MV limits, rate limits or CV bands, so each MV moves inside the range it covered in the records, the evidence the fit rests on, and the PI is clipped to the same box. The TCLab FOPDT fit ($T_s = 5\,\mathrm{s}$) gives $T_{\mathrm{set}} = 790\,\mathrm{s}$, $N_p = 158$ and boxes of 0 to 79.4 and 0 to 77.2 % for Q_1 and Q_2, and the column fit ($T_s = 1\,\mathrm{min}$) gives $N_p = 85$ and 1.77 to 2.13 and 1.55 to 1.84 lb/min for reflux and steam. The present input is held whenever the solver status is not optimal or a planned level is not finite or leaves its box by more than 10^{-6} of the width.

12.5.2 Dynamic-Matrix and State-Space Predictions

The DMC predicts from the step-response coefficients $g_{ij}[\ell]$ of the fit, as shown in Equation (12.17).

$$\begin{aligned} \hat{y}'_i[k+\ell] &= \beta_i[k] + \sum_j \sum_{t<k} g_{ij}[k+\ell-t]\,\Delta u'_j[t] \\ &\quad + \sum_j \sum_{t=k}^{k+\ell-1} g_{ij}[k+\ell-t]\,\Delta u'_j[t] \\ \beta_i[k] &= y'_i[k] - \sum_j \sum_{t<k} g_{ij}[k-t]\,\Delta u'_j[t] \end{aligned} \tag{12.17}$$

where $\Delta u'_j[t]$ is the move of input j at sample t, planned only at block starts, g_{ij} is held at its last value beyond the model length N_m, 316 on the TCLab fit and 170 on the column, and $\beta_i[k]$ is the output bias. Measured DVs enter the first sum with their present move, and their future moves are zero. The prediction matrix is fixed, so the Hessian is formed once. The bias is the whole estimator of this form and assumes that the discrepancy seen now persists over the horizon. The DMC is refused for an integrating path, as on the surge tank.

The state-space form runs the pruned realization, with dead times as exact shift registers, augmented by one constant disturbance d_j per MV that enters where a held input of that MV reaches the dynamic states, as shown in Equation (12.18).

$$\begin{aligned} x[k+1] &= Ax[k] + B_u u'[k] + B_w w'[k] + B_d d[k] \\ d[k+1] &= d[k], \qquad y'_{\mathscr{C}}[k] = Cx[k] \end{aligned} \tag{12.18}$$

where w' holds the measured DVs, held over the horizon. A steady-state Kalman filter estimates the dynamic states and d with process variances of 10^{-6} of the mean CV variance per state and $10/\sum_i w_i K_{ij}^2$ per disturbance and measurement variances $\sigma_{y,i}^2$, converging in 253 Riccati iterations on the TCLab fit and 182 on the column. The disturbance estimate removes the steady offset of a model error or an unmeasured load [20], the mechanism Volume 9, Chapter 4 uses for a fuel bias. The realization has 27 states on the TCLab fit, 23 of them registers, and 26 on the column, and the form is refused for FIR fits and above 90 states. With the output bias selected in place of the filter both forms solve the same program, and they agree to 1.0×10^{-5} of an MV's box width on the TCLab fit and 4.3×10^{-5} on the column, the difference being the truncation of the step model.

12.5.3 Decentralized PI and Controller Export

The comparison is a decentralized PI without a decoupler, paired by the relative gain array [21] on the diagonal of both records ($\Lambda_{11} = 1.039$ and 2.019), the one-to-one pairing with positive relative gains closest to 1. Volume 6, Chapter 2 treats this pairing choice on a plant whose relative gain hides a one-way interaction. Each loop is tuned by SIMC [22] on the paired path's FOPDT equivalent, with $\theta_e = \theta + T_s/2$, $\tau_c = \theta_e$, $K_c = \tau/(K(\tau_c + \theta_e))$ and $\tau_I = \min(\tau, 4(\tau_c + \theta_e))$, which gives $K_c = 7.61\,\%/^\circ\mathrm{C}$ and $\tau_I = 104\,\mathrm{s}$ on T_1, $K_c = 12.0$ and $\tau_I = 94.9\,\mathrm{s}$ on T_2, and $K_c = 0.441$ and $-0.104\,(\mathrm{lb/min})/\mathrm{mol}\%$ with $\tau_I = 12.0$ and 14.3 min on x_D and x_B. Each loop runs in velocity form with the integral alone absorbing the clipping (Section 1.6.2), and one column row adds the static feedforward $-(K_{iF}/K_{ij})w'_F$ of the measured feed that the MPC predicts with.

A second export, `mpc.py`, carries the controller to Python. Its plain-Python checks confirm that the realization reproduces the DMC's coefficients, that the first move equals the optimum of the same quadratic program rebuilt in the script, with an active set on the bounds when one binds, and the SIMC gains. On the bundled fits these agreed to 2.0×10^{-14} relative (21 fits with a DMC) and 6.5×10^{-16} of the box width (all 25). With numpy and GEKKO the script also builds the controller on the `state_space` object in control mode (`IMODE` 6, output bias by `FSTATUS` 1) and compares its closed loop with the application's within 15 % of IAE [15], since GEKKO moves at every sample where the application holds blocks.

12.6 Measured Results

12.6.1 Five Records with Known Structure

Every bundled record is generated from a stated model with seeded noise, so every visitor sees the same data and every answer can be checked against the generator. The five records and the behavior each one tests are listed in Table 12.3. All results below use default settings, the default chronological splits and the full mask unless a zero is stated, and were produced by the application's own fitting pipeline in a measurement test that calls the same job functions as the worker.

Validation scores for all five families on all five records are collected in Table 12.4. No family wins everywhere. The first-order matrix ranks first on the two-heater and column records, the state-space model on the tank and the stage, and the first-order matrix again on the exchanger, where three candidates fall within 2 % of the best (0.0939 and 0.0940 twice) and parsimony selects 7 parameters over 9 and 19. Every candidate beats both baselines on every record. The one-step persistence baseline reaches 0.0260 on the TCLab record, close to the best free-run model, which is why a one-step score is never used for ranking. Without its simulation-error refinement ARX scores 0.0643, 0.1208, 1.546, 0.0355 and 0.0971 in table order, so the refinement helps on four records and costs 0.008 on the exchanger, which was recorded under feedback.

🕮 **Pause.** Try Exercise 12.1 before reading on.

Fit times on the measurement host ranged from 4 ms (first order, exchanger) to 771 ms (state space, column), and the first-order matrix of the column took 59 ms. The FIR candidate reuses the preliminary FIR fit that supplies every family's delays and starting points. Every first-order CV fit and eight of the nine second-order ones stop before the iteration limit. The exception is the column's

Table 12.3 Bundled Records. Samples Are Training, Validation and Final Test on the Uniform Grid.

Record	Inputs, CVs	T_s, samples	Generator and purpose
TCLab	2 MV, 2 CV	5 s, 864/288/289	radiative two-heater balance, 12 s sensor lag, 0.32 °C quantization, weak real cross paths
Column	2 MV, 1 DV, 3 CV	1 min, 900/300/300	Wood-Berry matrix plus pressure with no reflux path and a weak feed path, two campaigns
Surge tank	1 MV, 1 DV, 1 CV	2 s, 1080/360/360	integrating level of a 12 m^3 tank, no steady state
Stage	1 MV, 1 DV, 2 CV	0.02 s, 1800/600/600	table at 1.6 Hz with $\zeta = 0.15$, 40 ms transport
Feedback	1 MV, 1 DV, 1 CV	0.5 min, 864/288/288	exchanger under PI control with three setpoint steps and a drifting unmeasured disturbance

Table 12.4 Free-Run Validation NRMSE by Family and Record, With Estimated Parameter Counts in Parentheses. Bold Marks the Candidate Ranked First.

	TCLab	Column	Tank	Stage	Feedback
FOPDT	**0.0232** (14)	**0.1093** (30)	0.396 (7)	0.1203 (14)	**0.0940** (7)
SOPDT	0.0275 (18)	0.1096 (39)	0.125 (9)	0.0254 (18)	0.0940 (9)
ARX	0.0523 (20)	0.1146 (36)	0.190 (4)	0.0057 (16)	0.1053 (6)
FIR	0.0283 (574)	0.1166 (894)	4.386 (359)	0.0060 (962)	0.1216 (287)
SS	0.0295 (14)	0.1112 (75)	**0.062** (46)	**0.0056** (23)	0.0939 (19)
static	0.622	0.660	5.960	0.405	0.316
persistence	2.251	1.057	6.258	1.039	0.784

pressure CV, whose absent reflux path runs to the lower damping bound and whose search stops at 60 iterations, 84 with scouting. The final-test NRMSE of each first-ranked candidate is 0.0267 (TCLab), 0.1166 (column), 0.196 (tank), 0.0056 (stage) and 0.0651 (exchanger), the lowest final-test score on four records. On the exchanger, where the first three candidates lie within 0.0001 on validation, ARX, ranked fourth, has the lowest at 0.0619. On the tank the final-test NRMSE of every family that follows the level is 2.8 to 3.5 times its validation score, because an integrating level drifts further the longer the free run continues.

12.6.2 Recovering the Wood-Berry Matrix

The column record is the most direct check of credit assignment, because three inputs move independently but overlap in time and two of the paths into each composition have opposite signs. Table 12.5 compares the first-order matrix with the generator. All six Wood-Berry gains are recovered within 0.9 %, the dead times of six of the seven strong paths within 0.10 min, and every strong-path time constant within 8 %. The weak feed path into pressure is where the record runs out of information. Its gain of 0.0349 against 0.035 is close, but its dead time of 3.51 min against 2 min and lag of 5.69 min against 9 min trade against each other, because the full 0.5 lb/min range of feed moves pressure by at most 0.0175 bar against 0.004 bar of noise.

The feed dead time into distillate composition, 8.87 min against 8.1 min, is the only strong-path dead time off by more than 0.1 min, and its lag is 1.14 min short, so the two errors partly cancel in the delayed rise. The absent reflux path into pressure is estimated at 0.0141 bar/(lb/min) with a lag of 114.6 min and a dead time on its upper bound of 270 min, half the longest training stretch. That is a physically meaningless shape, and its steady effect over the full 0.4 lb/min reflux range would be 0.006 bar against 0.004 bar of noise. That entry is what a structural zero replaces.

Table 12.5 First-Order Matrix of the Column Record Against the Generator. Gains in mol%/(lb/min) for x_D and x_B and bar/(lb/min) for P, Times in Minutes.

Path	K true	K fit	τ true	τ fit	θ true	θ fit
$x_D \leftarrow R$	12.8	12.69	16.7	16.80	1	1.00
$x_D \leftarrow S$	−18.9	−18.74	21.0	20.81	3	3.10
$x_D \leftarrow F$	3.8	3.78	14.9	13.76	8.1	8.87
$x_B \leftarrow R$	6.6	6.64	10.9	10.93	7	7.00
$x_B \leftarrow S$	−19.4	−19.43	14.4	14.31	3	3.06
$x_B \leftarrow F$	4.9	4.91	13.2	12.96	3.4	3.36
$P \leftarrow R$	0	0.0141	—	114.6	—	270.0
$P \leftarrow S$	0.35	0.348	6.0	5.89	1	1.00
$P \leftarrow F$	0.035	0.0349	9.0	5.69	2	3.51

Table 12.6 Validation NRMSE Before and After a Structural Zero, With Parameter Counts. Every Zero Verified Exactly in Every Family.

	Column, $P \leftarrow R$ absent path	Column, $x_D \leftarrow S$ strong path	TCLab, $T_1 \leftarrow Q_2$ weak real path
FOPDT	0.1093→0.1085 (27)	0.1093→0.3892 (27)	0.0232→0.0853 (11)
SOPDT	0.1096→0.1091 (35)	0.1096→0.3913 (35)	0.0275→0.0867 (14)
ARX	0.1146→0.1145 (35)	0.1146→0.3989 (27)	0.0523→0.1081 (16)
FIR	0.1166→0.1154 (795)	0.1166→0.3914 (795)	0.0283→0.1075 (431)
SS	0.1112→0.1128 (57)	0.1112→0.3994 (31)	0.0295→0.0838 (33)

TRY IT IN THE APP 3 · A Path That Is Not There

On 1 · Import load the Distillation column example, and on 3 · Identify fit First order + dead time with the defaults. Open the Pressure ← Reflux cell and watch its dead time sit on the upper bound of 270 min. Press Set response to zero: the cell reads Fixed at zero, verified exactly, and the validation NRMSE falls slightly, from 0.1093 to 0.1085.

12.6.3 Structural Zeros, True and False

Three structural edits were applied to every family, and the results are collected in Table 12.6. Setting the absent pressure path to zero removes 3 parameters from the first-order matrix and moves its validation NRMSE from 0.1093 to 0.1085, and three other families improve slightly, by 0.00002 (ARX) to 0.0013 (FIR). The rows of distillate and bottoms composition are identical before and after in the transfer-function, ARX and FIR families, because those families fit each CV separately, and only the pressure row is re-estimated, with the steam gain moving from 0.3477 to 0.3482. Figure 12.5 captures the application immediately after this first-order refit. The state-space model is the exception that the method predicts. The zero splits its single 6-state block into a 5-state block for the two compositions and a 1-state block for pressure, so the parameter count falls from 75 to 57, but the composition block no longer shares its states with pressure, its distillate row rises from 0.1007 to 0.1090, and the correct zero costs this family 0.0016 of validation NRMSE. Figure 12.3 draws the scores of Table 12.6 as bars on one scale.

Removing the strong steam path into distillate composition is a false structural claim, and every family reports it. The first-order validation NRMSE rises from 0.1093 to 0.3892, and the distillate row alone rises from 0.0996 to 0.9394, close to the CV's own training standard deviation. The refit also shows what credit reassignment looks like. With steam gone, the reflux gain into distillate climbs from 12.69 to 19.47 and its lag from 16.8 min to 42.7 min, and the feed gain from 3.78 to 5.23, because the remaining paths are refitted to absorb as much of the missing steam response as their own movement in this record permits. A fitted gain is therefore a property of the model structure as

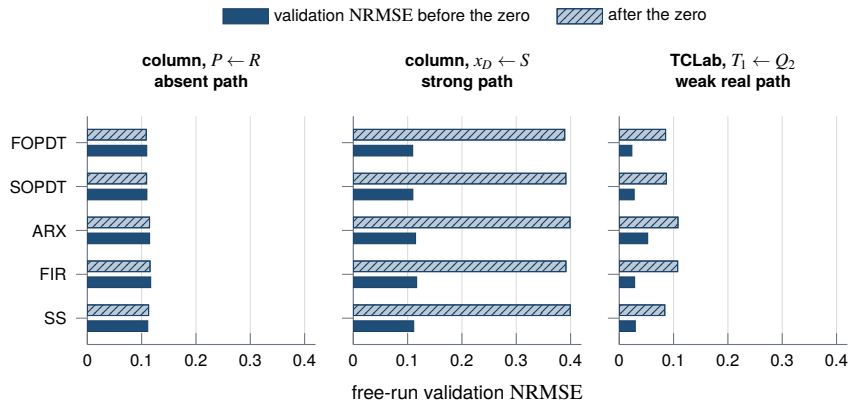


Figure 12.3 Free-run validation NRMSE of every family before (solid) and after (hatched) a structural zero, from Table 12.6, on one scale. Notice that the absent path changes almost nothing, while removing the strong steam path raises the error of every family to between 0.3892 and 0.3994, and removing the weak but real heater path raises it in every family, by a factor of 3.7 for the first-order matrix.

well as of the plant.

🕮 **Pause.** Try Exercise 12.3 before reading on.

The weak heat path from heater 2 to sensor 1 on the TCLab board is real, with a fitted gain of 0.079 °C/% against 0.512 °C/% for the direct path, a dead time of 52.6 s against 10.9 s and a lag of 181 s against 104 s. Setting it to zero raises the first-order validation NRMSE from 0.0232 to 0.0853, a factor of 3.7, and the T_1 row from 0.0299 to 0.1541, while the T_2 row is unchanged at 0.0165. The direct gain rises from 0.512 to 0.542 °C/% and its dead time falls from 10.9 to 4.8 s as the model tries to explain heat that arrives late through the board with heat that arrives early. The generator contains no transport delay at all, so both identified dead times are the apparent delay of a 12 s sensor lag and the transistor's own thermal capacity seen through a first-order model. In the state-space family this zero raises the parameter count from 14 to 33, by splitting the 2-state model into blocks of 1 and 4 states.

✦ **DID YOU KNOW? Removing a Path Costs Little or a Lot**

From the simulator: fixing the absent reflux path into column pressure at zero moves the first-order validation NRMSE from 0.1093 to 0.1085. Fixing the weak but real path from heater 2 to sensor 1 at zero raises it from 0.0232 to 0.0853, a factor of 3.7 (Table 12.6).

12.6.4 Oscillation, Integration and Feedback

The stage record separates the families by model shape. Its table position rings at 1.6 Hz with $\zeta = 0.15$, and the first-order matrix cannot represent that, scoring 0.1203 on validation against 0.0254 for second order. The second-order position path returns $\tau = 0.100$ s against the 0.0995 s of the generator's natural frequency and $\zeta = 0.196$ against 0.15, with a gain of 0.589 mm/V against 0.6, while both motor-speed paths run to the upper damping bound and are reported as effectively first order. The state-space model, at order 3 with 23 parameters, ARX with 16 coefficients and FIR with 962 all reach 0.0056 to 0.0060, likely because the drive enters the table through a motor lag and a transport delay that neither transfer-function shape describes exactly. Without its simulation-error refinement ARX scores 0.0355 on this record.

The surge tank has no steady state. The first-order fit drives both lags to 6049 and 6480 s, the second at its upper bound of three training spans, and the application reports both paths as integrating

Table 12.7 Maximum Absolute Errors Printed by Each Exported `model.py`, Outputs Outlet and Jacket. Step Columns Are the Worst Matrix Cell.

Family	numpy vs app	GEKKO vs numpy	step, numpy	step, GEKKO
FOPDT	7.1e-15, 7.1e-15	4.7e-10, 5.0e-11	5.6e-16	5.0e-12
SOPDT	2.6e-13, 4.3e-13	4.9e-10, 5.0e-11	7.8e-16	5.0e-12
ARX	2.1e-14, 7.1e-15	4.9e-10, 5.0e-11	6.7e-16	5.0e-12
FIR	1.4e-14, 7.1e-15	4.9e-10, 5.0e-11	3.3e-16	5.0e-12
SS	1.4e-14, 1.4e-14	4.8e-10, 5.0e-11	7.8e-16	9.3e-12

with initial slopes of 0.0013861 and −0.0022985 %/s per unit input against the generator's 0.0013889 and −0.0023148, errors of 0.20 % and 0.70 %. The second-order fit reaches 0.125 with slopes of 0.0013904 and −0.0023083. ARX and state space now hold the integrator itself. The chosen ARX, with $n_a = n_b = 1$ and 4 coefficients, has its pole at $|z| = 1.0003$, within $1/N_s$ of the circle, starts from the first measured level and scores 0.190, with slopes 8 to 10 % low. The subspace step placed the slowest mode at $|z| = 1.0018$, the scaling moved it onto the circle, and the 5-state block, with its measured start and drift and 46 parameters, follows the level at 0.062, the best score of any family, with slopes of 0.0013884 and −0.0022376. The FIR model alone cannot represent an integrator within its horizon and scores 4.386, and even the best model reaches 0.196 on the final test as its slope error accumulates.

The exchanger record was taken under PI control with $K_c = 0.9\,\%/^\circ\mathrm{C}$ and $\tau_I = 8\,\mathrm{min}$, three setpoint changes and a drifting unmeasured disturbance, and the readiness summary flags it as closed loop. The setpoint steps supply external excitation, and the valve gains agree across families at 0.911 (first order), 0.896 (ARX) and 0.911 (state space) against the generator's 0.9 °C/%, with a first-order lag of 5.97 min and dead time of 1.46 min against 6 and 1.5 min. The inlet-temperature path is recovered at 0.446 against 0.45. ARX is the one family that the refinement hurts here, at 0.1053 against 0.0971 for its equation-error estimate with a valve gain of 0.907, which is why the app's guide advises switching the refinement off for a record taken under feedback. The FIR model is the outlier, at 0.856 and 0.246, and it also has the worst validation score of 0.1216, so the agreement among the parametric families is the check that the record supports.

12.6.5 Export Parity

The export was verified on a two-input, two-output test system with a zero path, a negative second-order path and fractional dead times, with every family exported and its `model.py` run under numpy 2.5.4 and GEKKO 1.3.2. Its state-space model has a 6-state block for the jacket with a measured start and a drift, so the starting rule and the constant state are tested as well. Before export, the realization of every family reproduces the application's free run and step responses to below 10^{-9}. The errors reported by the scripts themselves are listed in Table 12.7. The numpy recursion matches the application to at most 4.3×10^{-13}, which is double-precision round-off on outputs of order 20 to 40, GEKKO matches numpy to at most 4.9×10^{-10}, consistent with its 10^{-10} relative tolerance, and the zero path is exactly 0.0 in both numpy and GEKKO for every family.

12.6.6 Scale

The specification called for an interactive benchmark of 50 000 samples with 8 inputs (5 MVs and 3 DVs) and 4 CVs. A synthetic record of that size, with first-order paths, dead times of 0 to 4 samples, six true zero paths left estimated, and noise of 0.05, was fitted family by family with default settings in Node.js 26.12 on the development host. The results are given in Table 12.8. Every family finishes in between 2.5 s and 15.0 s, and a second run on the same host agreed to within 0.4 s per family. Every family but FIR includes the preliminary FIR fit, which alone takes the 2.5 s of the FIR row, and the delay scan. The resident memory of the whole test process stayed between 323 and 370 MB. Most of the ARX time is its simulation-error refinement, which the sensitivity recursion (12.11) keeps to

Table 12.8 Fit Time and Validation NRMSE on 50 000 Samples, 8 Inputs and 4 CVs

Family	Time (s)	Parameters	Validation NRMSE
FOPDT	8.9	100	0.0379
SOPDT	15.0	132	0.0379
ARX	7.7	139	0.0615
FIR	2.5	2884	0.0390
SS	7.0	164	0.0539

Table 12.9 Closed-Loop Comparison on the FOPDT Fits at Default Settings. IAE in °C s and mol% min, Interaction in °C and mol%, Overshoot of the Stepped CV in Percent, Input Travel $\sum|\Delta u|$ in Percent and lb/min. The Mismatch Rows Repeat the First Setpoint Step, and DMC, SS Marks Rows Where the Two Forms Agree to Three Digits. Interaction Is the Largest Absolute Deviation of the CV Whose Setpoint Holds.

		TCLab two-heater fit, $T_s = 5$ s						Column fit, $T_s = 1$ min					
Case	Controller	IAE T_1	IAE T_2	Inter.	Ovs.	Q_1	Q_2	IAE x_D	IAE x_B	Inter.	Ovs.	R	S
T_1 or x_D step	PI	347	29.6	0.344	1.5	49.2	8.36	4.10	5.19	0.475	8.0	0.557	0.139
	DMC, SS	392	7.51	0.137	2.1	47.8	5.73	5.49	0.978	0.0739	0.17	0.156	0.0633
T_2 or x_B step	PI	38.9	335	0.387	3.7	5.94	58.4	2.03	7.00	0.142	0	0.166	0.134
	DMC, SS	6.10	366	0.121	1.9	4.54	56.6	0.573	4.92	0.0413	0.20	0.0828	0.0645
Load or feed	PI	217	9.33	—	—	17.4	4.03	0.667	4.59	—	—	0.0682	0.0496
	PI + FF	—	—	—	—	—	—	1.46	1.37	—	—	0.218	0.0788
	DMC	372	72.3	—	—	16.7	1.18	0.414	0.890	—	—	0.0169	0.0428
	SS	132	14.4	—	—	28.0	3.01	0.414	0.890	—	—	0.0169	0.0428
Residual	PI	138	114	—	—	182	256	22.5	21.3	—	—	11.5	2.17
	DMC	156	129	—	—	37.6	43.6	20.8	16.1	—	—	1.33	0.594
	SS	161	134	—	—	64.6	70.1	22.1	17.3	—	—	1.54	0.850
Gain ×0.7	PI	494	42.3	0.293	0	40.5	7.00	4.98	7.20	0.399	3.7	0.425	0.0819
	DMC	548	4.61	0.0961	0	40.3	5.99	7.78	1.18	0.0546	0.03	0.171	0.0673
	SS	548	6.59	0.0734	4.0	41.8	6.64	7.00	1.34	0.0706	2.1	0.203	0.0835
Gain ×1.5	PI	272	24.3	0.434	9.8	65.6	14.3	6.05	10.9	0.593	20	1.12	0.436
	DMC	337	15.0	0.198	14	60.0	5.58	4.29	1.24	0.110	7.2	0.164	0.0562
	SS	337	23.7	0.261	2.4	75.7	9.85	4.66	0.926	0.0734	0.47	0.183	0.0566
SS fit as plant	PI	318	28.2	0.181	0.37	47.0	3.71	3.81	5.82	0.409	9.4	0.516	0.120
	DMC	383	17.7	0.135	0.94	47.0	6.96	5.98	1.21	0.0677	2.0	0.161	0.0679
	SS	390	16.5	0.158	3.1	47.8	8.36	6.01	1.13	0.0743	3.0	0.176	0.0730

one filtered pass per coefficient. ARX still trails the transfer functions here, at 0.0615, likely because each CV shares one output polynomial of order at most 3 among eight inputs whose true lags range from 8 to 26 samples. The state-space refinement was skipped for all four CVs by the cost budget and reported, which explains its validation score of 0.0539 against 0.0379 for the first-order matrix. The FIR horizon was capped at 90 samples by the 720-coefficient rule, giving 2884 coefficients.

12.6.7 Predictive Control Against Decentralized PI

The controllers were compared on the FOPDT fit of the two records with two MVs, ranked first on both, at the defaults ($N_p = T_{\text{set}}/T_s$, $\lambda_u = 1$, Kalman filter). Each case starts at rest and runs for two settling times, 316 samples on the TCLab fit and 170 on the column, or for a replayed stretch, with one plant, box and case for all three controllers. A setpoint step is the smaller of one training standard deviation and the change whose steady inputs use half of the MVs' room in the box, 7.66 °C on T_1, 7.04 °C on T_2, 0.673 on x_D and 0.488 mol% on x_B. This rule replaced one standard deviation after the first column run held both MVs at a bound for 89 to 99 % of the samples, and the interaction finding below also holds under the original rule (0.587 under PI against 0.242 mol%). The disturbances are an unmeasured step of 15.9 %, a fifth of Q_1's recorded range, on the TCLab fit and the measured feed record of the validation stretch on the column. The validation residual is replayed on every CV, and the first step is repeated on plants with the MV gains at 0.7 and 1.5 times the fit's and on the state-space fit, as collected in Table 12.9.

Both MPC forms hold the CV whose setpoint stays far closer than the PI, which has no decoupler,

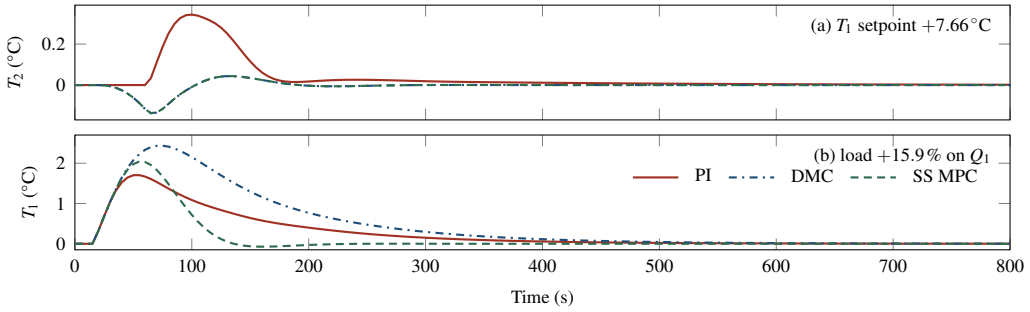


Figure 12.4 Deviation of the CVs of the TCLab FOPDT fit under the decentralized PI (solid), the DMC (dash-dotted) and the state-space MPC with its Kalman filter (densely dashed). (a) T_2 while the T_1 setpoint steps, with a largest deviation of 0.344 °C under PI and –0.137 °C under both MPC forms, whose traces coincide. (b) T_1 after an unmeasured heater load, with an IAE of 217, 372 and 132 °C s over the whole 1575 s case, of which the first 800 s are drawn.

at 0.137 °C against 0.344 °C on T_2 when T_1 steps and 0.0739 against 0.475 mol% on x_B when x_D steps, as shown in Figure 12.4(a), while moving the other MV less (5.73 % of Q_2 against 8.36 %). On the stepped CV the PI is faster in three of four cases, 347 against 392 °C s on T_1, 335 against 366 on T_2 and 4.10 against 5.49 mol% min on x_D, while the MPC is faster on x_B, 4.92 against 7.00. The move suppression prices moves the SIMC loops make freely, and at $\lambda_u = 0.1$, a counterfactual outside the table, both forms beat the PI on T_1 (342) and x_D (3.24). At $\lambda_u = 10$ the T_1 IAE rises to 598 °C s, and at each of the three settings 0.1, 1 and 10 the interaction stays below the PI's.

The DMC loses the unmeasured load, with an IAE on T_1 of 372 °C s against 217 for the PI and 132 for the Kalman form and a peak of 2.43 °C against 1.70 and 2.04, as shown in Figure 12.4(b). The load reaches T_1 through a 104 s lag, at 0.682, 2.55 and 4.67 °C after 20, 50 and 100 s of a final 8.13 °C in open loop, while the bias of Equation (12.17) predicts that each discrepancy stays as it is. Two counterfactual runs place the loss in the estimator. The state-space form with the output bias gives the same 372, and when the same disturbance arrives as an output step, for which a constant bias is the correct model, the DMC reaches 396 against 373 and returns within 10 % of its peak first, at 80 against 115 s. On the column both forms predict the measured feed and reach 0.414 and 0.890 mol% min against 0.667 and 4.59 for the PI. Static feedforward helps the PI on x_B (1.37) but harms x_D (1.46), likely because feed reaches x_D 7.9 min after reflux does (Table 12.5).

> **TRY IT IN THE APP** **13 · MPC From the Fit**
>
> With the TCLab two-heater test fitted by First order + dead time, open 7 · MPC at the defaults and wait for the table. In Setpoint step on T1 temperature, read the Interaction column: 0.344 °C under PI and 0.137 °C under both predictive forms. Then compare the T_1 IAE in Unmeasured load on Q1 heater: 217 for the PI, 372 for the DMC and 132 °C s for the state-space form.

On the replayed residual the PI has the lowest IAE on the TCLab fit, 138 and 114 against 156 and 129 for the DMC, with five to six times the DMC's input travel, and the DMC the lowest on the column. At 1.5 times the gain the Kalman form overshoots least, 2.4 % against 9.8 for the PI and 14 for the DMC on T_1 and 0.47 % against 20 and 7.2 on x_D, and the MPC interaction stays below the PI's in all six mismatch rows. All 444 runs over the 25 bundled fits stayed finite and inside the box with no plan held. In Node.js on a shared host a DMC decision took 18 µs on average and a state-space decision 2.0 ms on the TCLab fit and 1.0 ms on the column. At half the sample period these findings hold (interaction 0.121 against 0.309 °C, load IAE 297 for the DMC against 184 and 79), and only the order of the two MPC forms where they agree to three digits reverses.

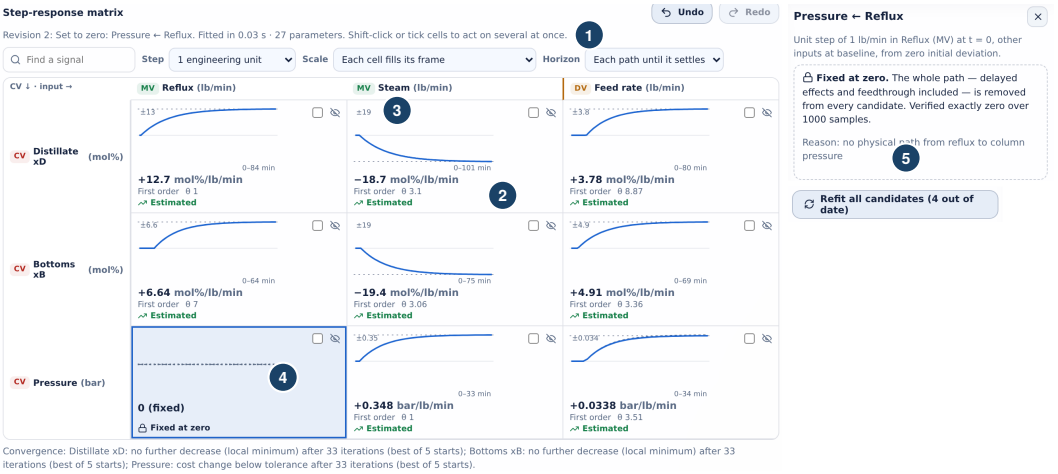


Figure 12.5 The Identify step on the distillation column record with the first order plus dead time candidate after Pressure ← Reflux was set to zero with the reason "no physical path from reflux to column pressure", shown as a composite of three parts of the screenshot: the step-response matrix at left, and the path inspector's report and the refit button at right. (1) The matrix shows revision 2, refitted in 0.03 s with 27 parameters. (2) Each cell prints its gain, 12.7, −18.7 and 3.78 into distillate, 6.64, −19.4 and 4.91 into bottoms, and 0, 0.348 and 0.0338 into pressure. (3) Each cell has its own vertical extent and time span, written in its corners. (4) The zeroed cell reads Fixed at zero. (5) The inspector reports the zero verified exactly over 1000 samples, and the button below it shows the other four candidates out of date until refitted.

12.7 Technology in Practice

The workbench has no instruments of its own. It reads a record, and the devices that made the record decide what any family can recover. Figure 12.6 draws the column record as it would be taken on a plant.

What Measures It

The bundled records carry their sensors in the generator: seeded noise on every record, and a 12 s lag and 0.32 °C steps on the TCLab temperatures (Table 12.3). On a real column the two compositions would come from an online analyzer that separates a sample over time [23], so each reading arrives late and only once per analysis cycle, a delay that the fit cannot tell from the process, as the TCLab sensor lag shows in Section 12.6.

> **TECHNOLOGY** **Sensor: An Online Composition Analyzer**
>
> Product purity at the top and bottom of a column is often measured by a process gas chromatograph. A small sample is injected at the head of a long, narrow column, and a carrier gas washes its compounds along so that they separate over time and reach the detector one by one. One analyzer reports several species, which suits the products of a distillation column. The separation takes time: cycles in the range of minutes are not uncommon, so a certain measurement dead time is inevitable. The sample line needs a continuous flow to waste, and without it the readings become grossly delayed [23]. An identified model cannot tell that delay from the process, so it would appear in every fitted dead time into x_D and x_B.

What Moves It

During identification an MV is a recorded trajectory, held between samples as an actuator is. In a plant it is often the setpoint of a flow, pressure, temperature or level loop of the regulatory layer, so

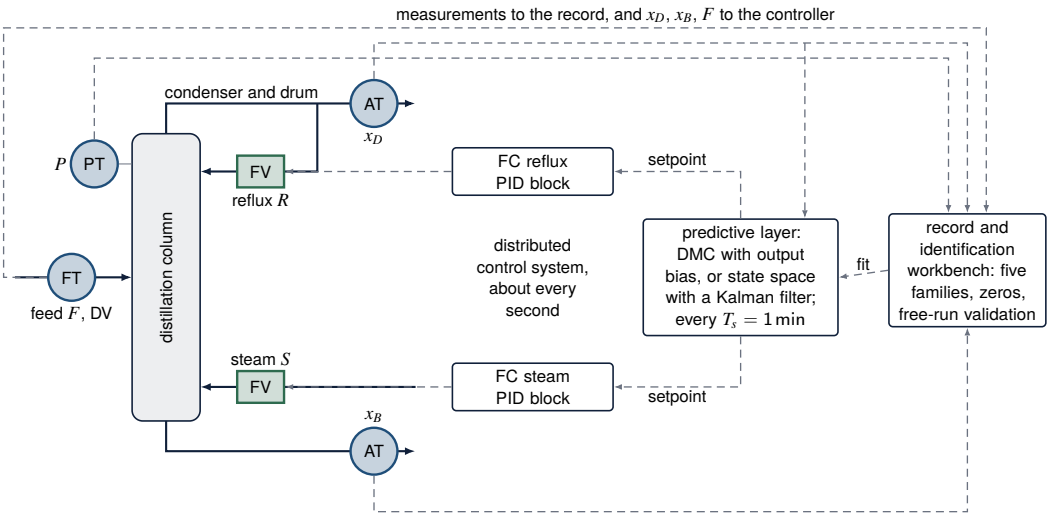


Figure 12.6 The column record of the chapter as it would be taken on a plant. An analyzer (AT) reads each product composition and a pressure transmitter (PT) the column top, a flow meter (FT) measures the feed, and the reflux and steam valves (FV) are positioned by flow loops of the distributed control system. The predictive layer above them writes the two flow setpoints once a minute from a model that the workbench identified on the record. Notice that the identified paths start at the setpoints, so they include the flow loops, and that every composition reading carries the analyzer's own delay.

the identified path runs through that loop and its valve [2]. Industrial packages excite the plant with step-like binary signals or steps of random size, because the test must reveal both the steady state and the dynamics [2].

> **IN PRACTICE** **Cut, Merge and Constrain a Plant Test**
>
> Most industrial identification packages step one, or at most a few, MVs at a time and hold the others at steady state, which is valid as long as the process is linear and superposition holds. A few move several MVs together with uncorrelated signals. Records from several test periods may be merged, which lets the engineer cut out a stretch corrupted by disturbances. The fit mostly minimizes a least-squares criterion, in an equation-error form that feeds past measured outputs back into the model or in an output-error form, and empirical models are often adjusted to known physics, for example a gain known to have a certain sign or value [2].

Where the Controller Runs

The workbench fits in a browser module worker, from 4 ms to 771 ms per family on the bundled records. The controller built from a fit then decides once per sample, every 5 s on the TCLab fit and every minute on the column, in 18 µs for a DMC decision and 2.0 ms or less for a state-space one. Plant control loops run with cycle times from milliseconds to minutes [24], and the predictive layer of a plant, which holds the structure of Section 12.5, sits above the PID function blocks of the distributed control system [2]. The decentralized PI of the comparison corresponds to those function blocks with no predictive layer above them.

TECHNOLOGY Final Element: The Setpoint of a Regulatory Loop

In the control hierarchy of a processing plant, a predictive controller sits above the PID loops of the distributed control system, and those loops position the valves. A survey of industrial practice draws the predictive layer acting about every minute and the PID layer about every second. The moves of the predictive controller are therefore setpoints of lower loops, and its identified paths include them. If the tuning of a lower PID loop changes significantly, a new process model may be needed, built by combining the identified model with a model from the new setpoint to the input [2]. The reflux and steam of the column record are flows in lb/min, the kind of variable such a loop holds.

What Is Modernizing It

Two developments reach this workbench. The first is regularization. Tikhonov proposed it for inverse problems in the 1940s, it became known as ridge regression in linear regression, and around 2000 kernel methods turned it into a kernel approach to identification whose core idea is to accept some bias for a smaller variance and a smaller mean square error. The penalty can also encode structural properties such as stability [25]. The FIR penalty of Equation (12.12), a ridge term plus a smoothness term with its weight chosen on validation, is a simple member of that family. The second is deep learning. Feedforward, convolutional and recurrent networks widen the range of models to choose from, a survey connects them to the kernel methods, and models with very many parameters show counterintuitive behavior such as benign overfitting and double descent [26].

12.8 Limitations

The workbench identifies linear, time-invariant models only. Hammerstein, Wiener, nonlinear ARX and grey-box structures are absent, so a plant whose gain changes with operating point is represented by an average gain and its validity range is the training range. No noise model is estimated. ARMAX and Box-Jenkins structures, instrumental variables and feedback-aware subspace methods are not in this release, so the equation-error ARX start is biased under colored noise, its simulation-error refinement removes that bias only in open loop, and every estimate can be biased by feedback without external excitation [12]. The transfer-function families contain no integrator and no process zero, so integrating paths are recognized and reported by slope rather than modeled in that family, and an inverse response is left to ARX, FIR and state space.

The structural block realization enforces a zero exactly but cannot share dynamics across blocks, so a zero can raise the state-space parameter count, from 14 to 33 on the TCLab record, and a correct zero can cost validation accuracy, 0.0016 on the column. The bootstrap is conditional on the chosen structure, with orders, delays, penalties and mask held fixed, so it understates uncertainty about the structure itself. The step-response dead time is read differently by family, as the fitted θ for transfer functions, the integer d_{ij} for ARX, the first sample above 5 % of the final value for FIR and above 1 % of the peak for state space, so dead times are comparable across families only approximately. A revealed final test is recorded with its revision but cannot be unrevealed. The measured start makes an integrating ARX or state-space model depend on its first measurements, one for ARX and $2n$ for a state block, so a noisy start shifts the whole free run. The 50 000-sample benchmark took 2.5 to 15.0 s per family, which is interactive only in the sense that the page remains responsive while the worker computes, and the state-space refinement is still beyond its budget at that size.

The predictive controller's model is the fit, which is also the nominal plant, so only the scaled-gain and other-fit cases measure model error. There are no rate limits, CV bands or user-set MV limits, and one tuning rule serves every fit. The GEKKO part of `mpc.py` was checked for syntax only, because the development host had neither numpy nor GEKKO, so its closed loop remains unverified.

The records are stylized, and each is defended on its own terms. A pilot column would not be accompanied by a pressure transmitter with exactly no reflux dependence, but the absent path serves to illustrate what a correct structural zero costs and buys. A surge tank whose level is never controlled would overflow in practice, but its record serves to illustrate why an integrator has a slope and not a

gain. The TCLab record is simulated rather than recorded, so that the weak path is known to be real, and the same procedure applies unchanged to a record taken on a physical board (Chapter 4).

12.9 Discussion

A structural zero is the place where process knowledge enters a black-box identification, and the measurements show both sides of that bargain. On the column, a correct zero improved validation NRMSE by at most 0.0013 in four families and cost the state-space model 0.0016. Those changes are small because the absent path was estimated near zero anyway, and the more important effect is that the controller built on the model can no longer act through a gain of 0.014 bar/(lb/min) at a dead time of 270 min that exists only in the noise. A false zero on a strong path raised validation NRMSE by a factor of 3.6 and silently moved the reflux gain from 12.69 to 19.47, which is a model that would mislead a predictive controller into using reflux to do steam's work. The asymmetry is the practical lesson. Removing an absent path costs little and protects the controller, while removing a real one is expensive and visible, but only if the model is judged on data it did not see.

The same discipline bears on model maintenance. A predictive controller's model is identified once and then trusted for years while the unit changes, and the identification step was already the costliest part of a commissioning when surveyed [2]. A record that keeps its roles, recipe, splits, mask with reasons, revision history and an export that checks itself can be re-identified and compared when the plant changes, which is the practical value of making every one of those choices explicit and reproducible.

> ⚠ **OPEN CHALLENGE** **When a Linear Model Is Not Enough**
>
> A road map of nonlinear identification notes that the real world is nonlinear and time-varying, and names distillation columns among the systems with nonlinear dynamic behavior. Compared with linear identification, the experiment design is more tedious, model selection much more involved and estimation more difficult, and very many nonlinear structures have been proposed, so the step raises the cost of the whole effort. Where a periodic test signal can be applied, a nonparametric analysis can detect the nonlinear distortions and estimate how much a nonlinear model could gain [27]. Progress would be a workbench that measures this gap on a plant record before a nonlinear family is chosen.

12.10 Conclusion

This chapter studied a browser workbench for multivariable identification, with first- and second-order transfer-function matrices, ARX, FIR and subspace state-space models fitted to the same prepared record and ranked on free-run validation error. The transfer functions are sampled exactly with fractional dead time and fitted by variable projection with a bounded multistart Levenberg-Marquardt search, and on the column record they recover all six Wood-Berry gains within 0.9 %. A path fixed at zero is removed from every family and verified to 0.0, with state-space zeros enforced by structural blocks whose Markov parameters vanish identically. A correct zero on the column changed the first-order validation error from 0.1093 to 0.1085, a false zero raised it to 0.3892, and a weak but real path on the TCLab record raised it by a factor of 3.7 when removed. An active-set search with a damping coordinate stops eight of the nine second-order CV fits on the bundled records before their iteration limit, and a measured start with a state drift lets ARX and state space follow an integrating level at 0.190 and 0.062, where the first release's models drifted. Oscillatory and closed-loop records show where each family fails, and every exported model reproduces the application to 4.3×10^{-13} in numpy and 4.9×10^{-10} in GEKKO. As a predictive controller the first-order fit cut the interaction to 0.137 from 0.344 °C on the TCLab record and to 0.0739 from 0.475 mol% on the column, while the dynamic-matrix output bias lost a heater load to the PI, 372 against 217 °C s. The models are linear and noise-free, the state-space zero cannot share dynamics across blocks, and the bootstrap is conditional on the chosen structure. A fuller workbench would add ARMAX and Box-Jenkins noise models, instrumental-variable and closed-loop subspace methods, integrating and inverse-response

transfer functions, constrained prediction-error refinement on a coupled state, and recorded rather than generated example data. The workbench is an accessible platform for identifying multivariable models from plant records and for testing what a structural assumption is worth.

KEY POINTS

- A multivariable model is not a collection of single-input models: the inputs into one output are fitted together.
- A fitted gain is a property of the model structure as well as of the plant: the reflux gain climbs from 12.69 to 19.47 when the steam path is falsely removed.
- Removing an absent path costs little and protects the controller, while removing a real one is expensive and visible, but only on data the fit did not see.
- The dynamic-matrix controller, whose output bias assumes that the present discrepancy persists, lost an unmeasured heater load to the PI, 372 against 217 °C s, while the Kalman form reached 132.

Where else this applies. A predictive controller built on an identified response matrix also runs above a decentralized structure in Volume 6, Chapter 8, and Volume 9, Chapter 4 removes an unmeasured fuel bias with a Kalman disturbance estimate, the remedy for the output bias measured here.

References

[1] J. Richalet, A. Rault, J. L. Testud, and J. Papon, "Model predictive heuristic control," *Automatica*, vol. 14, no. 5, pp. 413–428, 1978.

[2] 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.

[3] Y. Zhu, "Multivariable process identification for MPC: the asymptotic method and its applications," *Journal of Process Control*, vol. 8, no. 2, pp. 101–115, 1998.

[4] L. Ljung, "Perspectives on system identification," *Annual Reviews in Control*, vol. 34, no. 1, pp. 1–12, 2010.

[5] K.-J. Åström and T. Bohlin, "Numerical identification of linear dynamic systems from normal operating records," in *Theory of Self-Adaptive Control Systems*. Boston, MA: Springer US, 1966, pp. 96–111.

[6] K. J. Åström and P. Eykhoff, "System identification—a survey," *Automatica*, vol. 7, no. 2, pp. 123–162, 1971.

[7] M. R. Gevers, "A personal view of the development of system identification: A 30-year journey through an exciting field," *IEEE Control Systems*, vol. 26, no. 6, pp. 93–105, 2006.

[8] B. L. Ho and R. E. Kalman, "Effective construction of linear state-variable models from input/output functions," *Regelungstechnik (now at – Automatisierungstechnik)*, vol. 14, pp. 545–548, 1966.

[9] M. Verhaegen and P. Dewilde, "Subspace model identification Part 1. The output-error state-space model identification class of algorithms," *International Journal of Control*, vol. 56, no. 5, pp. 1187–1210, 1992.

[10] M. Verhaegen, "Identification of the deterministic part of MIMO state space models given in innovations form from input-output data," *Automatica*, vol. 30, no. 1, pp. 61–74, 1994.

[11] P. Van Overschee and B. De Moor, "N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems," *Automatica*, vol. 30, no. 1, pp. 75–93, 1994.

[12] U. Forssell and L. Ljung, "Closed-loop identification revisited," *Automatica*, vol. 35, no. 7, pp. 1215–1241, 1999.

[13] R. K. Wood and M. W. Berry, "Terminal composition control of a binary distillation column," *Chemical Engineering Science*, vol. 28, no. 9, pp. 1707–1717, 1973.

[14] 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.

[15] 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.

[16] G. Golub and V. Pereyra, "Separable nonlinear least squares: the variable projection method and its applications," *Inverse Problems*, vol. 19, no. 2, pp. R1–R26, 2003.

[17] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[18] H. R. Künsch, "The jackknife and the bootstrap for general stationary observations," *The Annals of Statistics*, vol. 17, no. 3, pp. 1217–1241, 1989.

[19] C. E. Garcia and A. M. Morshedi, "Quadratic programming solution of dynamic matrix control (QDMC)," *Chemical Engineering Communications*, vol. 46, no. 1–3, pp. 73–87, 1986.

[20] K. R. Muske and T. A. Badgwell, "Disturbance modeling for offset-free linear model predictive control," *Journal of Process Control*, vol. 12, no. 5, pp. 617–632, 2002.

[21] E. Bristol, "On a new measure of interaction for multivariable process control," *IEEE Transactions on Automatic Control*, vol. 11, no. 1, pp. 133–134, 1966.

[22] S. Skogestad, "Simple analytic rules for model reduction and PID controller tuning," *Journal of Process Control*, vol. 13, no. 4, pp. 291–309, 2003.

[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, Gaithersburg, MD, Special Publication 800-82r3, 2023.

[25] G. Pillonetto, T. Chen, A. Chiuso, G. De Nicolao, and L. Ljung, *Regularized System Identification: Learning Dynamic Models from Data*, ser. Communications and Control Engineering. Cham: Springer, 2022.

[26] G. Pillonetto, A. Aravkin, D. Gedon, L. Ljung, A. H. Ribeiro, and T. B. Schön, "Deep networks for system identification: A survey," arXiv preprint 2301.12832, 2023, published in revised form in *Automatica*, vol. 171, art. 111907, 2025, doi:10.1016/j.automatica.2024.111907.

[27] J. Schoukens and L. Ljung, "Nonlinear system identification: A user-oriented roadmap," arXiv preprint 1902.00683, 2019, published in revised form as "Nonlinear system identification: A user-oriented road map," *IEEE Control Systems*, vol. 39, no. 6, pp. 28–99, 2019, doi:10.1109/MCS.2019.2938121.

12.11 Exercises

12.1 ○ Concept check · 2 min

On the TCLab record the one-step persistence baseline, which predicts each sample as the last measured one, scores a validation NRMSE of 0.0260. The best free-run candidate scores 0.0232. Why does the workbench rank candidates on free-run error and never on one-step error?

(a) A one-step prediction carries the last measurement forward, so it scores well with no model, while a controller relies on the free run from the inputs

(b) A one-step prediction draws on final-test samples, which must stay hidden until a model has been chosen and the Test step reveals them

(c) A one-step error is always larger than a free-run error on the same split, so ranking on it would reject every candidate on every record

(d) A one-step prediction cannot be computed for a record with two inputs and two outputs, so it does not cover every record

12.2 ○ Concept check · 2 min

The reflux path into column pressure is fixed at zero and the state-space family is refitted. How does the workbench make that path exactly zero?

(a) It sets the reflux column of B to zero and keeps the single coupled block of six states

(b) It gives pressure its own block of states driven only by steam and feed, so the Markov parameters of the path vanish identically

(c) It adds a large penalty on the reflux gain into pressure, which drives that gain below 10^{-12} in the refit

(d) It sets the pressure row of C to zero, so that no state of the model reaches the pressure output

12.3 Concept check · 1 min

When the strong steam path into distillate composition is falsely fixed at zero, the refitted reflux gain into x_D rises from 12.69 to 19.47 mol%/(lb/min). What does the rise show?

(a) The reflux valve has become more effective since the record was taken

(b) The search stopped early and would return 12.69 given more iterations

(c) The remaining paths absorb what they can of the missing steam response

(d) The zero removed noise that had hidden the true reflux gain all along

12.4 Theory and calculation · 7 min

The first-order path from heater 1 to sensor 1 of the TCLab record has $K = 0.512\,^{\circ}\mathrm{C}/\%$, $\tau = 104\,\mathrm{s}$ and $\theta = 10.9\,\mathrm{s}$, sampled at $T_s = 5\,\mathrm{s}$.

(a) Split the dead time as $\theta = dT_s + \phi$. *Checkpoint:* $d = 2$, $\phi = 0.9\,\mathrm{s}$.

(b) With $A = -1/\tau$ and $B = 1/\tau$, evaluate Φ, Γ_0 and Γ_1 of Equation (12.6), and show that $\Gamma_0 + \Gamma_1 = 1 - \Phi$. *Checkpoint:* $\Phi = 0.95306$, $\Gamma_0 = 0.038656$, $\Gamma_1 = 0.0082834$, sum 0.046940.

(c) Heater 1 steps by 10 % at $k = 0$. Find the first sample at which T_1 moves, the deviation there and one sample later, and compare both with the continuous response $K\Delta u(1 - \mathrm{e}^{-(t-\theta)/\tau})$. What would rounding the dead time to 10 s give at $k = 3$? *Checkpoint:* $k = 3$: 0.198 °C; $k = 4$: 0.429 °C, equal to the continuous values at 15 and 20 s. Rounded: 0.240 °C.

12.5 Theory and calculation · 6 min

(a) Evaluate the damping factor of Equation (12.9) at the lower bound $q = \ln 0.1$, at $q = 0$ and at the upper bound $q = 4$. At the upper bound, find the ratio of the two real lags $\tau(\zeta \pm \sqrt{\zeta^2 - 1})$ and compare it with e^{2q}. *Checkpoint:* $\zeta = 0.1$, 1 and 28.3. Ratio 3199, against $\mathrm{e}^8 = 2981$.

(b) The stage generator rings at a natural frequency of 1.6 Hz with $\zeta = 0.15$, and the second-order fit returns $\tau = 0.100\,\mathrm{s}$ and $\zeta = 0.196$. Find the generator's τ, and from Equation (1.12) the overshoot and period of the generator and of the fit. *Checkpoint:* $\tau = 0.0995\,\mathrm{s}$. Overshoot 62.1 % against 53.4 %, period 0.632 against 0.641 s.

12.6 Theory and calculation · 6 min

Use the fitted gains, lags and dead times of Table 12.5 with $T_s = 1\,\mathrm{min}$.

(a) Compute the relative gain Λ_{11} of the reflux and steam pair on x_D and x_B from the fitted gains, and again from the generator's. *Checkpoint:* 2.019 from the fit, 2.009 from the generator.

(b) Tune the $x_D \leftarrow R$ loop by the SIMC rule of Section 12.5. *Checkpoint:* $\theta_e = 1.5\,\mathrm{min}$, $K_c = 0.441$ (lb/min)/mol%, $\tau_I = 12.0\,\mathrm{min}$.

(c) Tune the $x_B \leftarrow S$ loop the same way, and say what the sign of its gain means. *Checkpoint:* $\theta_e = 3.56\,\mathrm{min}$, $K_c = -0.103$ (lb/min)/mol%, $\tau_I = 14.3\,\mathrm{min}$.

12.7 App challenge · 8 min

Use scenarios 1, Read a two-heater matrix, and 2, A weak path that is real. The page opens on the TCLab two-heater test with the default splits. On 3 · Identify select First order + dead time and press Fit selected methods. The bundled records carry fixed noise, so the fit repeats, and every value below holds within one unit in its last digit.

(a) Open the cells $T_1 \leftarrow Q_1$ and $T_1 \leftarrow Q_2$ and read K, τ and θ of each. *Checkpoint:* 0.512, 104 s, 10.9 s; and 0.0789, 181 s, 52.6 s.

(b) On 4 · Validate read the candidate's validation NRMSE, and its T_1 and T_2 rows in the per-CV table. *Checkpoint:* 0.0232; T_1 0.0299, T_2 0.0165.

(c) Back on Identify, open $T_1 \leftarrow Q_2$, press Set response to zero and let the model refit. Read the validation NRMSE and its rows again, and K and θ of $T_1 \leftarrow Q_1$. *Checkpoint:* 0.0853; T_1 0.154, T_2 0.0165; $K = 0.542$, $\theta = 4.79$ s.

Explain why the direct path's dead time shrinks when the cross path goes.

12.8 App challenge · 8 min

Use scenario 3, A path that is not there. On 1 · Import load the Distillation column (2 MVs, feed DV) example, and on 3 · Identify fit First order + dead time with the defaults. Values hold within one unit in the last digit shown.

(a) Open Pressure ← Reflux and read K, τ and θ. From the reflux range of Table 12.1, find the largest steady pressure change this path predicts. *Checkpoint:* $K = 0.0141$ bar/(lb/min), $\tau = 114.6$ min, $\theta = 270$ min on a bound; 0.0056 bar.

(b) Press Set response to zero on that path and let the model refit. Read the zero check in the details, and on 4 · Validate the parameter count and validation NRMSE. *Checkpoint:* Verified exactly zero over 1000 samples; 27 parameters, 0.1085 (from 30 and 0.1093).

(c) Press Restore response, then set Distillate xD ← Steam to zero instead. Read the validation NRMSE, its x_D row, and K and τ of Distillate xD ← Reflux. *Checkpoint:* 0.389; x_D 0.939; $K = 19.47$, $\tau = 42.7$ min.

12.9 App challenge · 9 min

Use scenario 13, MPC from the fit, on the TCLab two-heater test with its First order + dead time fit. Open 7 · MPC at the defaults (horizon 1, $\lambda = 1$, Kalman filter) and wait for the table, about 8 s. Values hold within one unit in the last digit shown.

(a) In the rows of Setpoint step on T1 temperature, read the IAE of T_1 and the Interaction for PI, DMC and State space. *Checkpoint:* PI 347 and 0.344 °C; DMC and State space 392 and 0.137 °C.

(b) In the rows of Unmeasured load on Q1 heater, read the IAE of T_1 for the three controllers. *Checkpoint:* PI 217, DMC 372, State space 132 °C s.

(c) Set the move suppression to $\lambda = 0.1$, wait for the table, and read the first row pair of (a) again. Repeat at $\lambda = 10$. *Checkpoint:* $\lambda = 0.1$: MPC 342 and 0.0610. $\lambda = 10$: MPC 598 and 0.133. PI unchanged.

Explain why the PI wins on the stepped CV at $\lambda = 1$ and why the DMC loses the load.

12.10 Application area · 3 min

The feed path into column pressure has a true gain of 0.035 bar/(lb/min), a lag of 9 min and a dead time of 2 min. The record moves feed over 2.2 to 2.7 lb/min with 0.004 bar of pressure noise, and the first-order fit returns 0.0349, 5.69 min and 3.51 min.

(a) Find the largest steady pressure change that feed causes in the record, as a multiple of the noise. *Checkpoint:* 0.0175 bar, about 4.4 times the noise.

(b) Name two things the chapter offers an engineer who needs this path's dynamics better than the record gives them.

12.11 Application area · 4 min

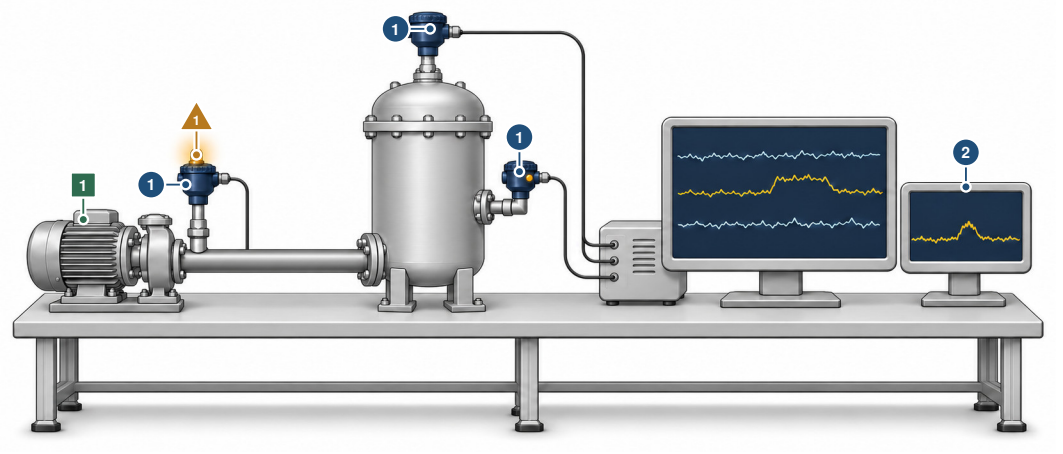
A plant engineer has a month of historian data from the exchanger's outlet temperature loop, always on automatic at one setpoint, with the steam valve moving only to reject an unmeasured drift. The engineer wants the valve gain for a predictive controller. Using the chapter's assumptions and its exchanger record, say what can go wrong and what to do instead.

12.12 Review the agent · 5 min

An AI assistant is asked to simplify the TCLab models. It writes: "The cross path $T_1 \leftarrow Q_2$ is weak. Fixing it at zero removes its three parameters from the first-order matrix, from 14 to 11. In the 2-state state-space model it removes the path's entry of B and its effect through C, so that model falls from 14 parameters to 12 or fewer." Find the flaw and say how you would check it.

CHAPTER 13

Process Monitoring Workbench That Diagnoses Records From Declared Simulated Plants



The controller moves 1 pumps whose schedule sets the flow through the network

The controller reads 1 flow, pressure and level transmitters whose readings the record keeps

2 a residual statistic on its own screen

What disturbs it 1 one meter that reads a few percent high

Notice that the meter with the amber light lifts one trend and the residual, yet neither screen says which meter is wrong; the chapter's records come from simulators, not from a bench like this one. Illustration, not to scale.

IN PLAIN TERMS **Can the Plant's Own Record Be Trusted?**

The main idea. A plant stores every reading it takes, often for years. This chapter checks such stored records for a meter that reads wrong, a fault that is starting, a valve that sticks and a burst of warnings.

The challenge. A warning does not name its cause. One bad meter can make every number that is added up with it look wrong. Readings taken one after another are also linked to each other, so false warnings come far more often than the usual rule of thumb promises.

Why it matters. An operator who trusts the wrong warning, or who gets too many, can miss the one that matters.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the fact on reviewing alarms (Section 13.1), the histories of three-sigma limits and of an alarm flood (Section 13.2), the try-it on the water network, the fact on false alarms and the try-it on the alarm flood (Section 13.5) 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 results (Section 13.5), Technology in Practice (Section 13.8), Limitations (Section 13.9) and the open challenge

The control question A plant record holds the evidence of a biased flowmeter, a developing fault, a sticking valve and an alarm flood. What does each monitoring statistic actually say about such a record, when one gross error spreads into every balance it enters and the samples of a closed-loop plant are not independent?

Learning objectives

- Reconcile a linear flow network, apply the global and measurement tests, and explain why one biased meter fails every test.
- Train a principal-component model and tell what the T^2 and Q statistics each detect.
- Update EWMA and CUSUM charts by hand and convert a run length to time.
- Estimate the false detections per day of a run-of-three rule, and explain why a closed-loop record raises far more.
- Tell a sticking valve from an aggressive loop, and rationalize an alarm flood with deadbands and delays.

Prerequisites Sections 1.4.8 (least squares, ARX) and 1.4.9 (residuals and their autocorrelation). Section 1.6.3 is needed only for the MPC health tool. The chapter is best read after Chapter 12.

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

See also Volume 5, Chapter 4, whose drifting kappa analyzer is the digester record here, and Volume 5, Chapter 3 for control through a slow analyzer. Volume 2, Chapter 4, whose sticking steam valve produces loop A, and Volume 2, Chapter 6 for a gain-driven limit cycle that a sticking valve would mimic.

13.1 Introduction

Every measurement a plant makes is stored, usually for years, and the questions asked of that store have a long history of their own. Flow and inventory data are adjusted to close their balances before they are used for accounting [1, 2], many correlated signals are charted together against a model of normal operation [3], a loop's routine record is compared with the best variance any controller could reach [4], and the alarm log is judged against the rates an operator can handle [5]. None of these tasks needs a plant model. Each needs a record, a declared structure and a statistic with a limit.

DID YOU KNOW? Fifty Alarms a Shift, or One Shift an Alarm

Fixing an alarm system takes time. Guidance on alarm handling from the Health and Safety Executive estimates that a quick first pass may cover perhaps 50 alarms per shift, while a thorough review and redesign may take more than one shift per alarm [6].

The misconception this workbench is built to correct is that a statistic above its limit names a cause, and that a 99 % limit fires on one sample in a hundred. On the shipped water-network record a single 3 % calibration shift on one flowmeter puts all seven measurement tests above their limit, from 11.68 to 17.71 against 2.68, because a gross error spreads into every stream that shares a balance with it. On normal days of the port of the Tennessee Eastman plant [7] studied in Volume 6, Chapter 8, the residual statistic sits above its 99 % limit on 4.3 % of samples, because samples of a closed-loop plant are not independent. The same two facts govern any unit with redundant meters and feedback loops, from a refinery's material balance to a utility's distribution network.

However, a student usually meets these methods one at a time, on small static examples, and almost never on a record whose cause is known and declared. Benchmark plant data exist, but the cause of an anomaly in a real historian is rarely known with certainty, and a synthetic example built for one method teaches nothing about what the others would have said. This chapter works with a workbench in which five tools act on records generated by other simulators of the collection that accompanies this series, with every injected change written in the file. It reports what each tool measures on each record. The workbench simulates no plant itself: it reads records, as the identification workbench of Chapter 12 does, and asks whether the evidence in them can be trusted.

The estimators and tests are the standard ones. The results read off the shipped records include a negative result (one Tennessee Eastman fault is not detectable at the statistics' false-alarm rate, over 61 seeds) and two discrepancies in published sources, each checked by independent routes. Section 13.2 follows the methods' history. Section 13.3 sets out the records and the scripts that generate them, which run other simulators' engines headlessly and declare every change. Section 13.4 states the five tools as coded, in enough detail to rebuild them. Section 13.5 measures the ten cases, each of which has a completion rule in the app, and Section 13.6 adds a sixth tool that reads an exported predictive-control run. Section 13.7 reports the verification, Section 13.8 sets the records beside the devices and systems of a plant, and Sections 13.9 and 13.10 state the limitations and the conclusions.

13.2 Balances, Charts and Principal Components on Plant Records

Reconciliation of process data was set out for plant networks in 1976, when flow and inventory measurements of a plant were adjusted by weighted least squares so that the material balances closed, each adjustment scaled by its meter's variance [1]. A test on each adjustment followed, so that a single gross error could be detected and attributed to a measurement [8], and matrix projection removed unmeasured flows from the balances so that the measured ones could be reconciled alone and the unmeasured ones recovered afterward where they were observable [9]. By the end of the century reconciliation and gross-error detection were collected in a textbook that also covered dynamic systems and industrial applications [2]. The link between reconciliation and principal-component analysis was later made explicit, with a six-stream recycle network whose constraints principal components can recover [10].

The univariate charts are older. The cumulative sum was proposed in 1954 as a continuous inspection scheme [11], the geometric (now exponentially weighted) moving average in 1959 [12], and the run length of a CUSUM was computed by a Markov chain on the discretized statistic in 1972 [13]. The EWMA's properties and enhancements were worked out in detail later [14], and the NIST/SEMATECH e-Handbook prints the CUSUM run lengths that practitioners consult [15]. Multivariate monitoring added a limit for the residual of a principal-component model [16], was carried into chemical processes with charts of scores and residuals from correlated plant data [3], and gained lagged variables for autocorrelated records [17]. The Tennessee Eastman problem [7], published as a plantwide control challenge with a list of disturbances, became a common test bed for such monitors.

> ⧖ HISTORY **Three-Sigma Limits From the 1920s**
>
> During the 1920s Walter A. Shewhart proposed a general model for control charts. A sample statistic is plotted in time order with a center line at its mean and an upper and a lower control limit a chosen number of standard deviations away. Three became the accepted standard in industry. For a normal variable the two three-sigma limits are crossed by chance with a probability of 0.0027, and the handbook that sets this out calls such a risk purely arbitrary. Points inside the limits that form a nonrandom pattern also mean that something is wrong [15]. Shewhart rule 1 of the workbench is this limit, and the other Western Electric rules are tests for such patterns.

Loop assessment began with the observation that routine closed-loop data already contain the minimum variance any controller could reach, given the process delay [4]. Oscillation detection followed from the autocorrelation of the control error and the regularity of its zero crossings [18], and valve stiction was characterized by its stick band and slip jump in the plot of valve position against controller output [19]. Alarm management, finally, became a lifecycle with its own standard, ANSI/ISA-18.2-2016, issued by the International Society of Automation [5]. The workbench names that standard as the class of source of its alarm metrics, and every threshold it applies is a teaching value, not a number read from the standard.

Table 13.1 Shipped Records. Every Record's Header Names Its Generating Application, Scenario, Seed and Script and Declares Every Change Listed Here.

Record	Generator, case	Rows, period	Declared change
Tennessee Eastman, six records	tennessee-eastman, case 4 settings, seed 20261014	1441, 3 min	case event removed, one IDV switched on at 48 h (none in the normal record)
Water network	water-network-pumping, case 7, seed 20261030	119, 15 min means	station meter $\times 1.03$ from 4 h, three derived streams
Digester	kraft-digester, case 5, seed 20261030	1441, 1 min	none
Loop A	heat-exchanger, case 15, seed 20260825	1201, 0.1 min	none (the app's own 3 % stiction)
Loop B	continuous-caster, case 5, seed 20260923	6001, 0.1 s	none (gain 5, clean gate)
Alarm log	turbine-trip case 4 and smr-steam-header case 10, seed 20260906	3601, 1 s	trips moved to 1800 and 1500 s, colored noise τ 2 s, σ 0.2 % of span
Cabin nitrogen	space-station-eclss, case 7, seed 20261030	577, 5 min	none (extra record, unscored)

> ⧖ HISTORY **Too Many Alarms at a Refinery, 1994**
>
> In 1994 an explosion and fires at a refinery in Wales injured twenty-six people and caused damage of around £48 million. The investigation by the Health and Safety Executive found three key factors. There were too many alarms and they were poorly prioritized, the control room displays did not help the operators understand what was happening, and training for a stressful and sustained upset had been inadequate. In the last 11 minutes before the explosion the two operators had to recognize, acknowledge and act on 275 alarms. The regulator then issued guidance on better alarm handling to prevent similar accidents, and it recommends an industry guide to alarm systems, published in 1999, to which it contributed [6]. The alarm log of case 8 asks the same question in one simulated hour.

13.3 The Records and Their Declarations

The workbench reads a comma-separated file with a header row, a time column and any number of signals. Rows need not be evenly spaced, blank, `NaN` and text cells are missing values, and date-times become hours from the first row. A line `# monitor:` carries a JSON declaration of balances, alarm limits or loop roles, and other lines starting with `#` are the header. The shipped records, listed in Table 13.1, are generated by scripts that import another simulator of the collection read-only, run its engine at its own step, keep the measured columns a historian would hold and write the header with the generating application, scenario, seed, sampling, every change the script made and the script's name. No record carries a truth column. When the records were built, regenerating all thirteen files (twelve records and the answer file the tests use) reproduced every byte. Section 13.6 notes the one record that no longer does.

13.3.1 Six Tennessee Eastman Records

The Tennessee Eastman records come from the port of the plant in Volume 6, Chapter 8, under its decentralized structure with the pressure override, run for 72 h at the engine's 1 s step and sampled every 3 min from its 1 min historian. Each row holds the 41 measurements (XMEAS 1 to 41, noise and analyzer holds included) and the 11 valve commands (XMV 1 to 11), the 52 signals of the monitoring literature. The script removes the case's own disturbance and switches one IDV on at 48 h, so the first two days of every record are identical normal operation and serve as the training window. The faults are IDV(1), an A/C feed ratio step, IDV(5), a condenser cooling water inlet temperature step, IDV(13), a slow drift of the reaction kinetics, IDV(3), a D feed temperature step,

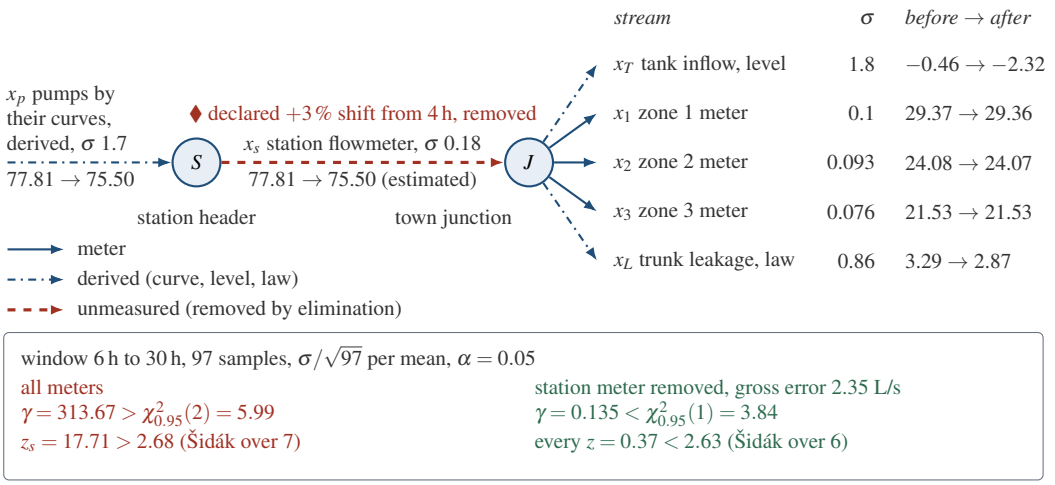


Figure 13.1 The water-network record as two balances, the station header S and the town junction J, with seven streams in L/s. Line style gives each value's source, and σ is the declared standard deviation of one 15 min value. Each stream carries its reconciled value before and after serial elimination on the window from 6 to 30 h. The station meter carries the declared 3 % shift and is the stream elimination removes. The clean window from 0 to 4 h gives $\gamma = 0.07$.

and, for the unlabeled challenge, IDV(4), a reactor cooling water inlet temperature step [7]. The challenge record's header says only that its scenario is withheld.

13.3.2 A Water Network With a Declared Meter Fault

The water-network record is case 7 of the distribution simulator of Volume 9, Chapter 11: scheduled pumping and a main break in zone 3 at 10 h, run 30 h. Every 10 s step the script reads the simulator's measurement packet and averages it into 15 min blocks stamped at each block's end. Four streams are meters (the station flowmeter and three zone meters). Three more are derived, as a utility engineer would write them down, from measured heads and the level, as shown in Equation (13.1).

$$
\begin{aligned}
x_p &= \sum_{i\,\text{running}} \sqrt{\left(a_i \omega_i^2 - (H_S - z_{cw})\right)/b_i} \\
x_T &= A_T \,\frac{\bar{h}_{k+1} - \bar{h}_{k-1}}{2\Delta t}, \qquad x_L = C_T\, p_J^{N_1}
\end{aligned}
\tag{13.1}
$$

where x_p is the pumps' flow from their declared curves (a_i, b_i and speed ω_i) at the measured station head H_S above the clearwell level z_{cw}, x_T is the tank inflow from a centered difference of the block-mean levels $\bar{h}$ with tank area $A_T = 600\,\text{m}^2$ and block $\Delta t = 15\,\text{min}$, and x_L is the trunk-main leakage from its declared law at the measured junction pressure head p_J with $N_1 = 1.18$, carried as an allowance. A one-sided difference lagged the block means by half a block and failed the clean 0 h to 4 h window with $\gamma = 9.12$, which is why the difference is centered. Each stream's standard deviation was measured against the engine's truth, then the truth was dropped, with the pump curves floored at 2 % and the leakage law at 30 % of their means as teaching values. The simulator's meters have no bias option and the simulator is not edited, so the script multiplies the station flowmeter by 1.03 from 4 h and says so in the header. The two balances, the seven streams and their declared σ are drawn in Figure 13.1.

13.3.3 The Other Records

The digester record is case 5 of the kraft digester simulator of Volume 5, Chapter 4, in which the kappa loop runs on an analyzer whose bias drifts from 1 h. The analyzer samples every 15 min and reports 20 min later, the laboratory samples every 4 h and reports 1 h later, and each column holds its last report beside the time it was sampled. The two loop records keep setpoint, measurement, controller output and valve (or gate) position. Loop A is a steam heater's outlet-temperature loop whose valve has the simulator's own 3 % stiction deadband, and loop B is a caster's mold-level loop with a proportional gain of 5 and a clean gate. The alarm log puts two simulators on one clock at 1 s, a steam-generating unit that trips at 1500 s and a turbine at 1800 s, with thirteen alarm limits set close to normal operation. Both simulators' measurements are exact at steady state, which left 26 raw annunciations and one flood that no deadband or delay could reduce, so the script adds declared colored noise with a 2 s time constant and a standard deviation of 0.2 % of each tag's span (seed 20261030). The cabin record, the nitrogen ledger from the sealed-cabin simulator of Volume 6, Chapter 10, is an extra, unscored record for the charts.

13.3.4 Truth and the Challenge's Answer

Truth stays in the generating simulators. The answers the completion rules need are in a separate file that the shipped code does not import. The labeled cases' answers are copied into a small key, and the challenge's fault class and three responsible variables are kept only as one-way digests. With seven classes and 52 variables a learner could recover them by hashing every option, so the digest guards against a casual look and is not a security measure. A test re-derives every key entry from the answer file and fails if any bundled source imports it.

13.4 Five Tools as Implemented

Each tool reads only the record's measured columns. The settings each exposes and their defaults are listed in Table 13.2. Every statistic is computed in the browser in double precision by code written for the workbench, with no numerical library.

13.4.1 Reconciliation and Gross-Error Tests

The balances of a record are linear in its stream flows x, with one row of the incidence matrix per node, $+1$ for a stream entering and -1 for one leaving. For the water record they are given by Equation (13.2).

$$\begin{aligned} S:&\quad x_p - x_s = 0 \\ J:&\quad x_s - x_T - x_1 - x_2 - x_3 - x_L = 0 \end{aligned} \tag{13.2}$$

where the symbols are those of Figure 13.1. Splitting the incidence matrix into its measured and unmeasured columns, $A = [A_k\ A_u]$, a basis P of the left null space of A_u removes the unmeasured streams, and the reduced balances B are an orthonormal basis of the rows of PA_k. With y the measured values (window means) and Σ the diagonal matrix of their variances, the reconciled values, the global test and the measurement test on each adjustment a_i are given by Equation (13.3).

$$\begin{aligned} \hat{x}_k &= y - \Sigma B^{\mathsf{T}} \left(B \Sigma B^{\mathsf{T}} \right)^{-1} B y \\ \gamma &= (By)^{\mathsf{T}} \left(B \Sigma B^{\mathsf{T}} \right)^{-1} B y \sim \chi^2(r) \\ z_i &= \frac{|a_i|}{\sqrt{\left(\Sigma B^{\mathsf{T}} (B \Sigma B^{\mathsf{T}})^{-1} B \Sigma \right)_{ii}}} \end{aligned} \tag{13.3}$$

where r is the number of reduced balances and $a = y - \hat{x}_k$. The global test fails when γ exceeds the $1-\alpha$ quantile of $\chi^2(r)$. Each z_i is compared with a two-sided normal limit Šidák-corrected over the m redundant measurements, $z_c = \Phi^{-1}\left(1 - [1-(1-\alpha)^{1/m}]/2\right)$, which is 2.68 for $m = 7$ and 2.63 for $m = 6$ (a Bonferroni correction gives 2.69 for $m = 7$). A window mean of n samples is given

Table 13.2 Tool Settings and Defaults. Visible Settings Are on the Tool's Panel, Declared Ones Come From the Record, Fixed Ones Are in the Code.

Setting	Default	Status
reconciliation		
balances, streams, kinds	from the record	visible
σ per sample, window	declared, 6 to 30 h (water)	visible
test level α (both tests)	0.05, range 0.001 to 0.2	visible
principal components		
training window	0 to 48 h	visible
components k	50 % cumulative variance	visible
lags l	0, range 0 to 4	visible
confidence of both limits	99 %, range 90 to 99.9 %	visible
detection run	3 samples	visible
charts		
EWMA λ, L	0.2, 2.962	visible
CUSUM k, h (in σ)	0.5, 5	visible
Shewhart rules	rule 1 only	visible
chain states, CUSUM and EWMA	120 and 151	fixed
loops		
delay d, AR order p	declared (10 samples), 10	visible
window	last two thirds of record	visible
moved threshold ε	0.02 % of output range	fixed
stiction screen	held $>$ 25 %, band $>$ 0.5 %	fixed
alarms		
deadband, on- and off-delay, filter	0, 0, 0, 0	visible
flood	$>$ 10 in a 10 min period	teaching
chattering	$\geq$ 3 within 60 s	teaching
rate target	1 per 10 min, 144 per day	teaching

$\sigma/\sqrt{n}$. An unmeasured stream is recovered from $A_u x_u = -A_k \hat{x}_k$ by the pseudo-inverse and reported as observable only when no direction of the null space of A_u moves it. A measured stream absent from every reduced balance is non-redundant and keeps its measurement. An allowance, such as the leakage law, is carried as a measurement with its own large σ.

Serial elimination reconciles, and while the global test fails it moves the redundant measurement with the largest z_i to the unmeasured set and reconciles again. Each step reports the removed stream, its estimated gross error (its measurement less its estimate once removed) and any stream whose test equals the largest to within 10^{-7} relative, an equivalent set that no test can split. The loop stops when the test passes or no redundancy remains. A per-sample global test evaluates γ on every row with the per-sample σ and $(B\Sigma B^{\mathsf{T}})^{-1}$ formed once. For a record with a fast and a slow instrument of one quantity, a second mode pairs each laboratory result with the analyzer reading whose sample time is nearest its own and fits a least-squares line to the paired differences, whose slope is the analyzer's drift rate.

13.4.2 Principal-Component Monitoring

Each variable that moves in the training window is standardized by its training mean and standard deviation, a variable that does not move is left out and listed, and l lagged copies may be appended for dynamic PCA [17]. The eigenvalues λ_i and eigenvectors of the training correlation matrix come from the cyclic Jacobi method. The model keeps the smallest k whose eigenvalues reach the chosen share of the total, or a fixed k. For a standardized row z with scores $t = P_k^{\mathsf{T}} z$ and residual $e = z - P_k t$,

the two statistics and their limits are given by Equation (13.4).

$$\begin{aligned} T^2 &= \sum_{i=1}^{k} \frac{t_i^2}{\lambda_i}, \qquad Q = \|e\|^2 \\ T_\alpha^2 &= \frac{k(n^2-1)}{n(n-k)} F_\alpha(k, n-k) \\ Q_\alpha &= \theta_1 \Big[\frac{c_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \Big]^{1/h_0} \end{aligned} \tag{13.4}$$

where n is the number of training rows, $\theta_j = \sum_{i>k} \lambda_i^j$, $h_0 = 1 - 2\theta_1\theta_3/(3\theta_2^2)$ and c_α is the normal quantile, the residual limit of Jackson and Mudholkar [16]. A detection is the first sample after training that begins a run of three consecutive samples above a limit, a declared persistence rule that keeps isolated crossings from counting. Contributions to a chosen sample are the squares of the elements of $D^{1/2}z$, with $D = P_k \Lambda_k^{-1} P_k^{\mathsf{T}}$, for T^2 and the squares e_j^2 for Q. Both sets are non-negative, sum to their statistic, are shown as shares, and sum a variable's lagged copies into the variable.

13.4.3 Control Charts and Their Run Lengths

A chart is trained on a window, from which the mean μ and standard deviation σ of one signal set $z_t = (x_t - \mu)/\sigma$, and its statistic starts from zero at the window's end. The Shewhart chart signals when $|z_t| > 3$ (rule 1), and the other three Western Electric rules (two of three beyond 2, four of five beyond 1, eight in a row on one side) are a switch. The EWMA and the two-sided CUSUM are given by Equation (13.5).

$$\begin{aligned} w_t &= \lambda z_t + (1-\lambda) w_{t-1}, \qquad |w_t| > L\sqrt{\lambda/(2-\lambda)} \\ C_t^{\pm} &= \max\big(0, C_{t-1}^{\pm} \pm z_t - k\big), \qquad C_t^{\pm} > h \end{aligned} \tag{13.5}$$

where the EWMA uses its asymptotic limit, the one run-length tables assume. The average run length (ARL) of each chart comes from a Markov chain on its discretized statistic [13], with transition matrix R and run lengths a from $(I - R)a = \mathbf{1}$, on 120 states for a one-sided CUSUM and 151 for the EWMA. The two sides of a CUSUM combine as $1/\mathrm{ARL} = 1/\mathrm{ARL}^{+} + 1/\mathrm{ARL}^{-}$, and a side whose chain is singular to working precision, because the shift lies far on the other side, is given an infinite run length.

13.4.4 Loop-Performance Indices

On the last two thirds of a loop record the workbench forms the control error $e = \mathrm{SP} - \mathrm{PV}$ and its autocorrelation up to a third of the window, at most 400 lags, of which the lag-one value of Equation (1.22) is the first. The decay ratio is the height of the first autocorrelation peak above the line joining the minima on either side, divided by the depth of the first minimum below the line from lag 0 to that peak, so 1 is an undamped cycle and 0 a damped one. The regularity of the zero crossings is $\bar{T}_p/(3 s_{T_p})$, with T_p twice each interval between crossings, after Thornhill, Huang and Zhang [18]. A sample counts as held when the controller output moved by more than ε while the valve moved by no more than ε. The held share of all samples on which the output moved, the median output travel during a hold (the stick band) and the number of holds ending in a jump describe the stick-slip pattern [19]. The minimum-variance benchmark fits an autoregressive model of order p to the mean-removed error by linear least squares, the same fit as the ARX model of Section 1.4.8 without an input, as shown in Equation (13.6).

$$e_t = \frac{a_t}{A(q^{-1})}, \quad \sigma_{mv}^2 = \sigma_a^2 \sum_{j=0}^{d-1} \psi_j^2, \quad \eta = \frac{\sigma_{mv}^2}{\sigma_e^2} \tag{13.6}$$

where ψ_j is the impulse response of $1/A$, d is the declared process delay in samples, σ_a^2 the innovation variance and η the Harris index [4], 1 at minimum variance and near 0 far from it.

13.4.5 Alarm Metrics

An alarm watches one signal, optionally filtered by $y_k = y_{k-1} + \Delta t\,(x_k - y_{k-1})/(\tau_f + \Delta t)$. A high alarm's condition sets when $y > L$ and clears only when $y < L - \mathrm{DB}$, with the deadband DB in percent of the tag's span, and a low alarm mirrors it. The alarm annunciates once its condition has held for the on-delay and returns once the clear condition has held for the off-delay. Metrics are counted in fixed ten-minute periods from the record's start: annunciations, the rate per ten minutes and per day, the busiest period, flood periods, chattering tags and the annunciations in their bursts, the alarms standing at the end and each tag's share. "Standing at the end" is the set active when an hour-long record stops, not the standard's stale alarm, which an hour cannot show.

13.4.6 Assumptions

Reconciliation assumes linear balances at steady state over the window and independent errors with the declared σ, so component, energy and bilinear balances are outside it, and a window that contains a transient reads as a gross error. Principal-component limits and chart limits assume independent samples from a stationary normal operation. A closed-loop plant wanders between days and its samples are correlated, unlike the independent residuals of a correct model in Section 1.4.9, so both kinds of limit are crossed more often than their nominal rate, which Section 13.5.3 measures. A contribution plot ranks the variables that carry the statistic at one sample, which is where a fault shows and not necessarily where it starts. The loop indices assume uniform sampling (a non-uniform record is resampled to its median interval) and a declared delay, and the stiction screen needs the valve position. Alarms are limit crossings on single signals, with no state-based suppression, shelving or priority.

13.5 Ten Cases, Measured

The ten cases and what each measures at the default settings are listed in Table 13.3. Each number below comes from the shipped records, measured with the workbench's own functions in a test that imports them, and the reconciliation, PCA, chart, loop and alarm numbers were also re-derived by independent code (Section 13.7).

13.5.1 The Meter That Lies

On the window from 6 to 30 h, 97 samples per stream, the two balances have rank 2 and the global test gives $\gamma = 313.67$ against the $\chi^2(2)$ limit of 5.99 at $\alpha = 0.05$. All seven measurement tests exceed 2.68, the station meter at 17.71, the pump curves at 13.20 and the five junction streams at 11.68 each, because the five form an equivalent set within one balance and one error spreads equally into them. Serial elimination removes the station meter first, with no tie, and estimates its gross error at 2.35 L/s, close to 3 % of its 77.85 L/s mean. The balances then close with $\gamma = 0.135$ against the $\chi^2(1)$ limit of 3.84, and every remaining test reads 0.37 against 2.63. Removal moves the reconciled station and pump flows from 77.81 to 75.50 L/s, the tank inflow from −0.46 to −2.32 L/s and the trunk leakage from 3.29 to 2.87 L/s, as drawn in Figure 13.1. The window from 0 to 4 h, before the shift, passes with $\gamma = 0.07$. The same state of the workbench is captured in Figure 13.2.

🕮 **Pause.** Try Exercise 13.1 before reading on.

Two further runs show what the structure allows. With the trunk leakage declared unmeasured, the junction balance projects away, one reduced balance remains ($\gamma = 177.21$), and the first elimination step reports the pump curves and the station meter as tied, because a meter in series with the only other measurement of its flow cannot be told from it. Around the main break, from 9 to 12 h with the station meter removed, $\gamma = 0.021$ against 3.84, because the broken main's flow passes through the zone 3 meter (22.33 L/s on that window) and the balance still closes.

Table 13.3 The Ten Cases, What the Default Settings Measure on the Shipped Record, and What Completes Each Case

#	Case and record	Measured with the defaults	Required to complete
1	False-alarm rate, normal days	$k = 11$, limits 25.22 and 42.59, no detection, Q above its limit on 5.0 % of the third day, one-day training detects at 29.70 h	Train on 0 to 48 h, then on one day or less, then confirm the rate
2	IDV(1) feed ratio step	Q 48.05 h, T^2 48.15 h, a feed variable first in Q's top three at 49.80 h	Find Q's detection, then pick a sample from about 50 h whose top three include a feed variable
3	IDV(5) cooling water step	Q 48.05 h, T^2 never (Q 12.5 %, T^2 0.8 % above)	Train with Q detecting within an hour and T^2 not at all
4	IDV(13) slow drift, product flow	EWMA and CUSUM 52.00 h, Shewhart rule 1 52.45 h, all four rules 49.15 h	Chart one signal with Shewhart and EWMA, the EWMA first
5	Water network	$\gamma = 313.67$, elimination removes the station meter, $\gamma = 0.135$	Fail the global test, then let elimination name the meter
6	Digester, analyzer and laboratory	six pairs, drift 0.2944 kappa/h	Name the drifting instrument and its rate within 50 %
7	Loops A and B	held 93.9 % against 0.4 %, η 0.009 and 0.341	Diagnose both loops
8	Alarm flood	494 annunciations and 6 floods fall to 14 and 0	No flood and at most a tenth of the raw count
9	Unlabeled record (IDV(4))	Q and T^2 48.05 h	Class, detection within 15 min, one of three variables
10	IDV(3) D feed temperature	Q detects at 54.75 h, above its limit on 8.3 % (normal 5.0 %)	Train on the fault record and on the normal record, then confirm

TRY IT IN THE APP 5 · The Water Network: Find the Meter That Lies

Load the case and press Reconcile. The global test reads 313.67 against its limit of 5.99, and every measurement test is above its limit. Then press Serial elimination and watch the station flowmeter turn red: it is removed with a gross error of 2.35, and every remaining stream passes its test.

The per-sample global test is a warning about the derived tank inflow. Before the shift, 3 of 15 samples exceed 5.99, each within one block of a step of the pump flow, at 1.25, 2.25 and 3.25 h. With the declared shift removed over the whole record, 11 of the 21 samples within one block of the seven steps exceed it, against 1 of the 98 samples away from them, whose mean γ is 0.84 against the $\chi^2(2)$ mean of 2. A centered difference spreads each level step over two blocks, and away from the steps the declared σ are conservative rather than small.

13.5.2 Two Statistics and a Contribution

On every Tennessee Eastman record 50 of the 52 signals vary in training (the compressor recycle and stripper steam valves are constant), the training window holds 961 rows, and 11 components carry 50.3 % of the variance. The limits are 25.22 for T^2, from $F(11, 950)$, and 42.59 for Q, from the 39 discarded eigenvalues. On the normal record nothing is detected on the third day, while Q lies above its limit on 5.0 % of samples and T^2 on 0.2 % (0.8 % for Q inside training). Two departures from a first design are measured choices. A model trained on one day detects falsely at 29.70 h (T^2 at 29.35 h), because the closed-loop plant wanders between days, and a model at 90 % variance keeps 33 components and detects falsely at 65.05 h.

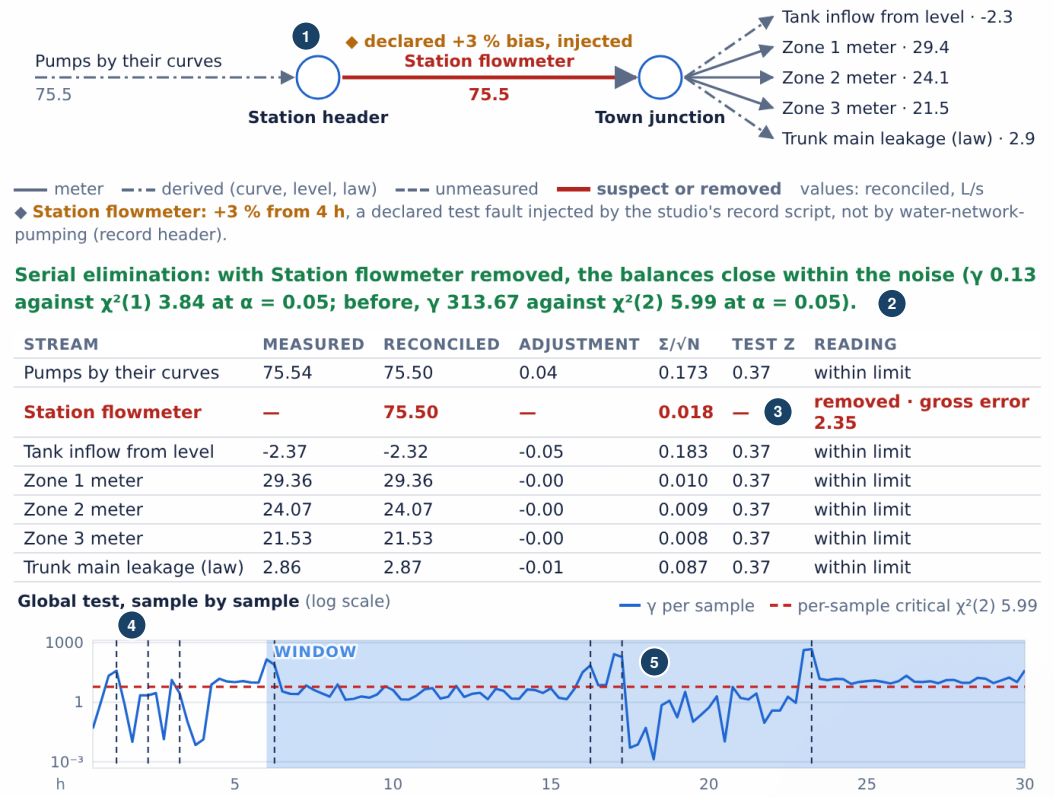


Figure 13.2 The Reconcile tool on case 5 after serial elimination, window 6 to 30 h at $\alpha = 0.05$, cropped to its result panel. (1) The diagram tags the station flowmeter with the declared 3 % shift once elimination has run and shows reconciled values in L/s. (2) The verdict line reads $\gamma = 0.13$ against $\chi^2(1)$ 3.84 after removal and 313.67 against $\chi^2(2)$ 5.99 before. (3) The table reads the station meter removed with a gross error 2.35 and every other test 0.37 within its limit. (4) The per-sample global test, on a log scale, marks the steps of the pump flow with dotted lines and (5) shades the window.

> ✦ **DID YOU KNOW?** **Five Percent Above a 99 Percent Limit**
>
> From the simulator. A 99 % limit on independent samples is crossed on about one sample in a hundred. On the normal Tennessee Eastman record, Q lies above its limit on 5.0 % of the third day (case 1 in Table 13.3), and over 61 normal days on 4.3 % of samples (Section 13.5.3).

The cooling water step of case 3 separates the two statistics, as shown in Figure 13.3. Q detects at 48.05 h, the first sample after the step, and lies above its limit on 12.5 % of the following day, while T^2 never detects and lies above on 0.8 %. The likely mechanism is that the temperature loops absorb the step in directions the 11 retained components did not span in training, which moves the residual and leaves the scores inside their limit. With two lags the model has 150 columns and 17 components, and T^2 then detects at 48.10 h. A 99.9 % limit would remove most false detections on normal days but would detect this fault with neither statistic.

🕮 **Pause.** Try Exercise 13.2 before reading on.

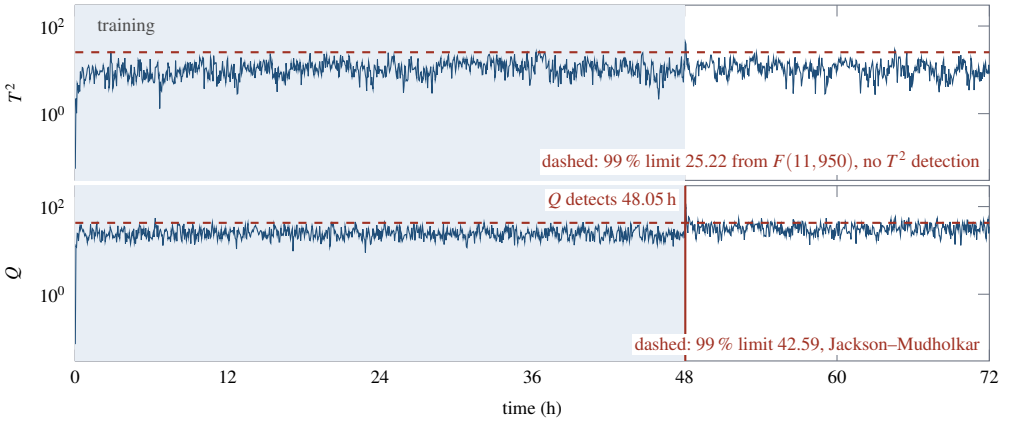


Figure 13.3 T^2 and Q on the cooling water step record (case 3), trained on 0 to 48 h (shaded) with 11 components, against their 99 % limits. Q detects at 48.05 h, the first sample after the step, and T^2 never detects.

On the feed ratio step of case 2, Q detects at 48.05 h and T^2 at 48.15 h, but the three largest Q contributions at detection are the reactor temperature, the reactor cooling water valve and its outlet temperature. None of the 35 samples between the step and 49.80 h has a feed variable among its three largest, and from 49.80 h on every sample has one, the reactor feed analysis of A beside the compressor work and the separator liquid valve. The variable that moves first is where the temperature loops answer within a sample or two, while the reactor feed analyzer samples every 0.1 h and reports 0.1 h later. On the unlabeled challenge record Q and T^2 both detect at 48.05 h, and the largest mean Q contributions over the following hour are the reactor cooling water valve, the reactor temperature and the reactor cooling water outlet temperature, which is the answer the challenge accepts within 15 min and one of three variables.

13.5.3 A Fault the Statistics Cannot See

The D feed temperature step of case 10 is absorbed by the control system. On the shipped seed Q detects at 54.75 h and lies above its limit on 8.3 % of samples after 48 h, against 5.0 % on the normal record, and T^2 never detects. One seed cannot decide whether that is a detection, so both records were regenerated for 61 seeds (1 to 60 and the shipped one) with the workbench's own generator and judged with its own monitor at the defaults. Normal days raise 76 detection episodes in 61 days, Q or T^2 runs of three merged when they start within three samples, which is 1.25 per day (95 % interval 0.7 to 1.8) on 28 of the 61 days, with Q above its limit on 4.3 ± 2.4 % of samples against the nominal 1 %. Independent samples above the limit that often would give about 0.04 runs of three per day. The fault days raise 2.02 episodes per day on 36 of 61 days, with Q above on 7.0 ± 3.5 % of samples. The paired difference in Q's share is positive on all 61 seeds, by 2.7 points on average, but a monitor never has the paired normal day. A whole-day test on Q's share, set at the normal days' 95th percentile of 8.1 %, flags 20 of the 61 fault days, and only once the day is over. This fault is therefore not detectable at the statistics' false-alarm rate, which is the result the case asks the learner to state.

🕮 **Pause.** Try Exercise 13.3 before reading on.

13.5.4 Drift in a Chart and in Time

The product flow of the slow-drift record (XMEAS 17), trained on 0 to 48 h with a mean of 22.9494 and a standard deviation of 0.1169, gives the EWMA and the CUSUM their first signal at 52.00 h and Shewhart rule 1 at 52.45 h. The two accumulating charts signal on the same sample because of the record, where one sample at $z = -2.46$ takes C^- from 3.25 to 5.21 and the EWMA from -0.80 to

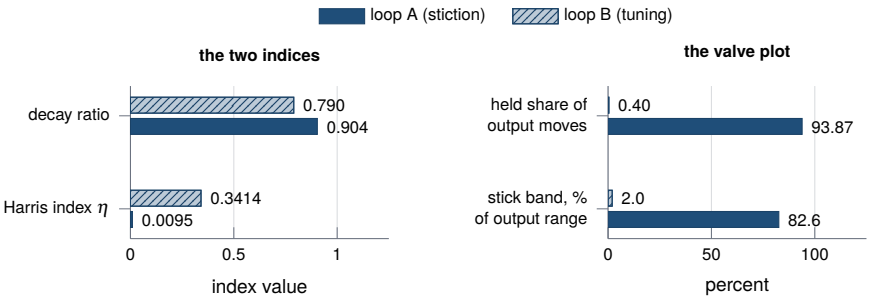


Figure 13.4 Loop indices of the two loop records, from Table 13.4. Notice that both loops show a high decay ratio and a Harris index far from 1, so the indices do not tell them apart, while the valve plot does: loop A's valve held still on 93.87 % of the samples on which the output moved, loop B's on 0.40 %.

Table 13.4 Loop Indices on the Last Two Thirds of Each Record, Delay 10 Samples, AR Order 10

Index	Loop A (stiction)	Loop B (tuning)
samples, period	801, 0.1 min	4001, 0.1 s
decay ratio	0.904	0.790
regularity (crossings)	81.99 (4)	12.88 (33)
oscillation period	13.82 min	2.42 s
held share of output moves	93.87 %	0.40 %
stick band, % of output range	82.6	2.0
Harris index η	0.0095	0.3414
stiction screen	stiction	none

-1.13 against a limit of 0.987. On the normal record neither the EWMA nor rule 1 signals, and the CUSUM signals at 64.85 h. The four Western Electric rules together fire at 49.15 h on both records, which is why rule 1 alone is the default. The EWMA's chain gives an in-control run length of 499.3 samples, about 25 h at 3 min, against 370.4 for rule 1. On the extra cabin record, a CUSUM on the nitrogen balance residual trained on the first day signals at 25.25 h, 1.25 h after the leak step, while the cabin simulator's own leak flag appears at 27.08 h in the five-minute record.

In the digester's in-time mode the six laboratory samples at 0, 4, 8, 12, 16 and 20 h pair with analyzer readings whose differences grow 0.05, 0.67, 2.60, 3.37, 4.35 and 5.93, a slope of 0.2944 kappa/h against the 0.3 kappa/h drift the simulator declares. The laboratory reads 29.81 at the first pair and 26.90 at the last, while the loop holds the drifting analyzer at its setpoint.

13.5.5 Loops and Alarms

The two loop records give the indices in Table 13.4. Both cycle with a high decay ratio and both read far from minimum variance, so neither index tells them apart. The valve plot does. On loop A the valve held still on 93.9 % of the samples on which the output moved, with a stick band of 3.56 output points (83 % of the output's range) over 11 holds ending in a jump. On loop B it held on 0.4 %, and the gate follows its command. Two caveats apply. Loop A's regularity rests on four zero crossings, three intervals, and its Harris index depends on how the autoregressive model is fitted, 0.0095 by least squares against 0.118 by Yule-Walker equations, although both read far from 1. Figure 13.4 draws the indices and the valve plot's shares of Table 13.4 side by side.

The raw alarm log annunciates 494 times in the hour, 73, 77, 30, 56, 137 and 121 per ten-minute period, so all six periods are floods, and five chattering tags carry 473 of the annunciations. A deadband of 2 %, an on-delay of 15 s, an off-delay of 60 s and a 5 s filter leave 14 (1, 0, 3, 10, 0 and

0), no flood and no chattering tag, while 1 %, 10 s and 30 s without the filter leave 16 with one flood of 11 between 30 and 40 min. The limits sit close to normal operation on purpose, as unrationalized limits often do, and every threshold is a teaching value of the kind the standard's class of documents defines [5].

> **TRY IT IN THE APP 8 · An Alarm Flood: Rationalize It**
>
> Load the case and read the raw count: 494 annunciations in the hour, and every ten-minute period is a flood. Set Deadband to 2 % of span, On-delay to 15 s, Off-delay to 60 s and Filter τ to 5 s. The count falls to 14, with no flood and no chattering tag.

13.6 MPC Health From an Exported Run

A sixth tool reads an MPC run without simulating its plant. The Health tab maps measured output, set point and applied input to the record's own columns. Where present, `ctrl_mpc`, `mpc_fallback`, `mpc_status`, `mpc_slack_*`, `mpc_J` and `mpc_iters` supply active time, fallback episodes, soft-band violations, objective and solver work. The time of a flag is the interval from its timestamp to the next, and the final row adds no duration. Consecutive fallback flags count as one episode, not as several failed solves. Positive slack is a violation, while zero slack does not identify a binding constraint. Measured time at or beyond a bound requires an explicitly mapped bound column. Missing quantities are reported unavailable.

A forecast column is scored against a measured output only at its declared future timestamp. With measured y and forecast $\hat{y}_{k+H|k}$, the residual is $y_{k+H} - \hat{y}_{k+H|k}$. RMSE and mean residual use exact matching future timestamps, not interpolation across data gaps. A held endpoint whose lead time changes between solves cannot be scored with one constant H. Input travel sums successive absolute changes and breaks at missing input samples. Categorical solver status stays aligned when rows are sorted into time order.

A CSV file downloaded from an app with the MPC option (Sections 1.6.3 and 1.6.4) can carry the accepted controller plan in a comment header. This is a copy of the plan in charge, not a retrospective simulation or an exported optimization model. Output predictions and held future moves share the measured run's time scale. A rejected or unavailable plan contributes no snapshot. The tab marks its origin and extends the set point by an explicitly labeled hold of the last recorded value, without claiming recipe preview. A saved snapshot may also be opened beside a longer matching record to assess its future predictions, and input changes after the snapshot contribute to that residual too. The tab links to a GEKKO resource for further study and does not produce a GEKKO model.

The workbench carries this tool as an extra case, "MPC health", shown in Figure 13.5, which asks the learner to inspect the difference between a prediction and a measured outcome and to record an analysis. Numerical checks exercise unequal sampling intervals, missing values, duplicate timestamps, a fixed-horizon forecast, grouped fallback flags, malformed snapshots and held moves. The case's record is regenerated from the microgrid's own engine with wall-clock costs removed. The original ten records and their completion judgments are unchanged. The alarm record no longer regenerates identically from its upstream plants. It is retained rather than silently replacing the evidence behind the alarm results of Section 13.5.

13.7 Verification

13.7.1 Code Verification

When the records were built, the workbench's own suite passed 79 tests (one large-data test is skipped without its environment), and regenerating the records reproduced all 13 files byte for byte. The reconciliation was re-derived independently by solving the Karush-Kuhn-Tucker system of the constrained least squares and taking the adjustment covariance from the solution operator. On 138 random networks it agrees with the projection route to 3.7×10^{-15} in γ and 8.2×10^{-14} in z, and on

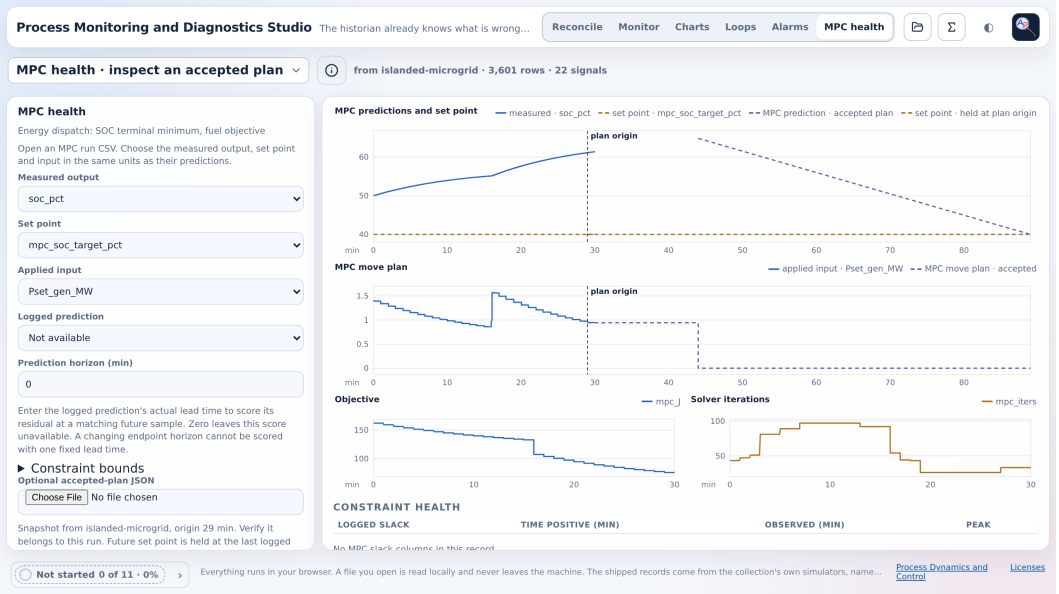


Figure 13.5 The MPC health case: a measured islanded-microgrid MPC run and its accepted future snapshot. The upper plot compares battery SOC with its terminal SOC target and the accepted prediction. The lower plot compares the genset command with its held move plan. The origin marks when this plan was accepted. Its future is a controller prediction, not a measured outcome. Objective and solver work are read from the exported record.

the water record it gives the same γ, tests, gross error and reconciled values. Under simulated noise on a four-balance network the global test's 20 000 draws have a mean of 3.992 and a variance of 7.99 against 4 and 8, with 5.13 % false alarms at $\alpha = 0.05$. On a synthetic rank-2 set of eight variables the PCA keeps two components and Q exceeds its 99 % limit on 0.920 % of 20 000 new normal rows (T^2 on 0.995 %). An independent Householder and QL solver reproduces the T^2 and Q limits on the plant records to the quoted digits, and the Jacobi solver reconstructs 27 symmetric matrices of order 2 to 96, random, clustered within 10^{-9} and rank-deficient, to 6×10^{-12} of their largest eigenvalue. The Harris index of a first-order moving-average error with coefficient 0.5 is 0.800 at $d = 1$ and 1.000 at $d = 2$, the exact values, and a synthetic alarm log reproduces its hand counts.

The run lengths were checked three ways. For the one-sided CUSUM with $k = 0.5$ the chain, a Nyström solution of the run-length integral equation on Gauss-Legendre nodes, and a Monte Carlo of 400 000 runs give 77.07, 77.08 and 77.06 ± 0.11 at $h = 4$ and a shift of 0.25, and 38.01, 38.01 and 38.00 ± 0.05 at $h = 5$ and a shift of 0.5. Doubling the chain from 120 to 240 states moves the first value from 77.073 to 77.077. The EWMA with $\lambda = 0.2$ and $L = 2.962$ has an in-control run length of 499.3 by the chain and 499.7 by the integral equation, and the two-sided CUSUM with $h = 5$ has 465.2 by the chain.

13.7.2 Agreement With Published Sources

The CUSUM chain agrees with the run-length table printed in the NIST/SEMATECH e-Handbook for $k = 0.5$ within 1.2 % in 18 of its 20 cells [15]. Two printed cells disagree, 74.2 at $h = 4$ and a shift of 0.25 and 30.0 at $h = 5$ and a shift of 0.5, where the three independent routes above agree on 77.07 and 38.01.

The reconciliation reproduces every structural statement of the published example of Narasimhan and Bhatt (Example 1 and Example 2 of the preprint) [10]. With all six streams measured the reconciled errors of F1 and F5, F2 and F6, and F3 and F4 are equal, and with only streams 1, 2 and 5

measured the projection leaves the single balance F1 = F5 with F2 non-redundant. The example's Table 2 prints, for 1000 simulated samples, root sums of squared errors before and after reconciliation (labeled RMSE). Taking the printed "before" values as the error scales and the printed standard deviations as the weights, independent errors predict the four "after" values of F1, F3, F4 and F5 within 1.6 % but F2 and F6 5.9 % high. The printed "before" value of F6 is itself 0.862 times what its stated standard deviation of 0.1 gives, 6.2 standard errors low. A single error correlation of −0.14 between F2 and F6, which is 4.4 standard errors from zero for 1000 independent pairs, reproduces all six printed "after" values within 1 %. The gap therefore comes from the example's simulated errors rather than from the method, and the workbench reproduces the method.

13.7.3 Runtime

Every wall time below is a bound over repeated runs on the development host. A record of one million rows and six signals (55.8 MB of text) parses in under 0.5 s, one decimated chart draws in under 10 ms, a PCA fit and score takes under 0.2 s, an EWMA under 10 ms and an alarm evaluation under 15 ms. In a desktop browser at 1366 × 768 a shipped case opens in under 0.4 s on a fresh page and in under 30 ms once its record is fetched, and training, serial elimination and an alarm setting each take under 0.1 s.

13.8 Technology in Practice

The workbench has no instruments of its own. It reads records, and the devices that made each record decide what its tools can find. Figure 13.6 draws two of the records as a plant would take them.

What Measures It

The water record's station flowmeter is a simulated meter that the record script multiplies by 1.03 from 4 h, because the simulator's meters have no bias option. A magnetic flowmeter suits most water-based liquids, and its reading is linear in flow with one proportionality constant [20]. An error in that constant would scale every reading by one factor, which is the form of the shift the record declares. The alarm log's thirteen limits are soft alarms, raised when a continuous measurement crosses a programmed limit, as distinct from hard alarms raised by process switches. A distributed control system time-stamps and archives its alarms for later analysis [20].

> **TECHNOLOGY** **Sensor: A Magnetic Flowmeter**
>
> A conductive liquid moving through a magnetic field induces a voltage across the pipe, and two electrodes opposite each other in the pipe wall pick it up. The voltage is proportional to the mean velocity, so the meter is linear in volumetric flow with one proportionality constant, and its open bore leaves very little pressure loss. It needs a reasonably conductive liquid and a full pipe, with the electrodes horizontal so that a passing bubble does not break contact. The induced voltage is often a millivolt or less and is easily buried in noise from stray currents unless the flowtube is grounded. The coils are driven with alternating polarity, because a steady field would polarize the ions at the electrodes and corrupt the reading [20].

What Moves It

The workbench moves nothing. The final elements are in the records: loop A's steam valve carries its simulator's own 3 % stiction deadband, and the stiction screen of Section 13.4 needs the valve position beside the controller output. On a plant that position comes from the valve's positioner. A process switch has a deadband for a related reason: without one, a variable that sits at the trip point cycles the switch between trip and reset and raises repeated nuisance alarms [20]. The deadband of the alarm tool does the same in software.

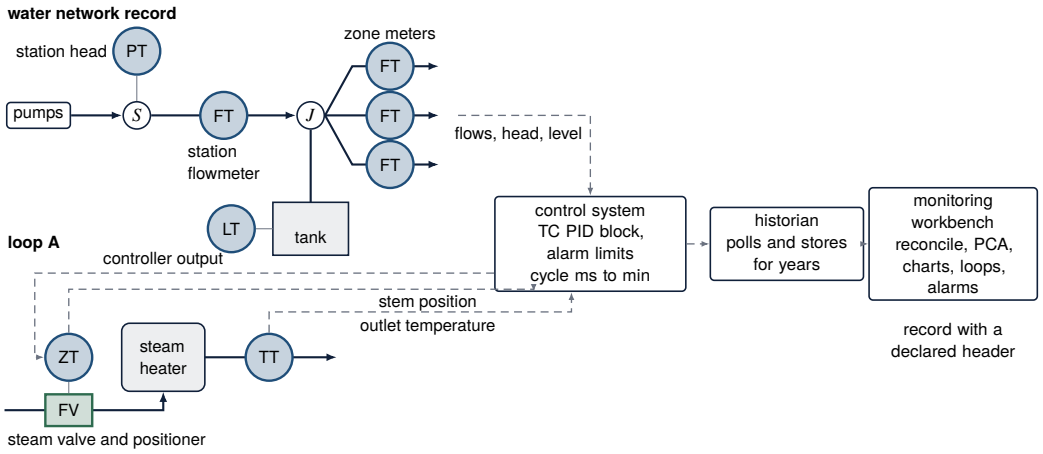


Figure 13.6 Two of the chapter's records as a plant would take them. Top: the water network's station flowmeter between the station header *S* and the town junction *J*, the station head, the three zone meters and the tank level of Figure 13.1. Bottom: loop A, a steam heater's outlet-temperature loop whose valve positioner reports the stem position that the stiction screen needs. The control system runs the loop and its alarm limits, the historian keeps what it polls, and the workbench reads a record. Notice that the workbench sits at the end of the chain and acts on no device.

TECHNOLOGY **Final Element: A Valve Whose Positioner Reports Its Stem**

Packing friction and the force of the process on the plug keep a pneumatic valve's stem from following its actuator pressure exactly. A positioner compares the stem position with the command and adjusts the actuator pressure until they agree, a small control loop of its own. An electronic positioner also reports diagnostic data such as supply and actuator pressure, position error and total stem travel. Its plot of actuator pressure against stem position, the valve signature, shows packing friction as the offset between the opening and the closing traces. A signature recorded after a rebuild can be compared with a later one to judge wear without taking the valve apart [20].

Where the Controller Runs

The workbench runs in a browser on a stored record: on a record of one million rows a PCA fit and score take under 0.2 s. The loops that made the records run in plant controllers with cycle times from milliseconds to minutes [21], and a data historian polls the control system at regular intervals and keeps the data, typically for years [20]. The shipped records are sampled from every 0.1 s to every 15 min, and a record can be far coarser than the engine that made it: the Tennessee Eastman engine steps every 1 s and its records keep one row in 3 min.

IN PRACTICE Rate Targets and Quick Fixes for Alarms

The guidance of the Health and Safety Executive starts by measuring: how many alarms there are, how many occur in normal operation and in an upset, and how many stand. Its targets are a long-term average of no more than one alarm every ten minutes in normal operation, and no more than ten displayed in the first ten minutes after a major upset. Priorities rest on the consequence of a missed response, with about three levels in proportions such as 5, 15 and 80 percent. Quick fixes include removing alarms that have no defined operator response, tuning the settings of nuisance alarms and adjusting the deadbands of repeating ones [6].

What Is Modernizing It

The first development is the electronic valve positioner, which measures actuator pressure as well as stem position, keeps a count of total stem travel and can predict when the packing will wear out and raise a maintenance alarm [20]. Where such data reach the historian, excess packing friction can be read from the valve's own signature instead of being inferred from the loop record. The second is machine learning. Fault diagnosis can be posed as classification, with sensor readings as the inputs to a classifier that predicts the fault or its absence, and an open benchmark on simulations of the Tennessee Eastman plant now tests eleven strategies for carrying such a classifier from one operating mode to another [22]. The principal-component monitor of Section 13.4 is trained on normal operation alone, while a classifier needs labeled examples of every fault it is to name.

13.9 Limitations

Reconciliation is linear and steady, so component, energy and dynamic balances are absent, and the water record needed three derived streams before its biased meter could be identified at all, because the junction's tank and zone streams form an equivalent set within one balance. The PCA is static or lagged, with no kernel, recursive or moving-window form, no partial least squares, and contributions that point where a fault shows. Its limits and the chart limits assume independent samples, which a closed-loop record violates, and the measured false-alarm rate of 1.25 detections per normal day is a property of these defaults on this plant, not a general rate. The loop indices are screening tools that need a measured valve position and a declared delay, and the Harris index of a limit cycle depends on the autoregressive estimator. The alarm metrics treat each tag alone, with no shelving, state-based suppression or priority, and their thresholds are teaching values. Every record is generated, so the workbench has not been tested on a plant's own historian. Each record is also stylized. A real station meter would drift rather than shift in one block, while the declared step shows how a gross error spreads into a balance and how elimination removes it. The cross-check of the workbench's Jacobi solver against the storage-ring simulator's decomposition waits for that simulator.

OPEN CHALLENGE A Monitor Trained in One Operating Mode

A fault classifier learned from historical data assumes that new data follow the same probability distribution as its training data, which is seldom verified in practice. The Tennessee Eastman plant has six modes of operation, set by the product mass ratio and the production rate, and each mode changes the statistics of the measured series, so a model trained in one mode may not generalize to another. A benchmark built on simulations of the plant tests methods that adapt a classifier with unlabeled data from the new mode [22]. Progress would be a monitor that keeps its false-alarm rate when the plant changes mode, without a labeled fault record for every mode.

13.10 Conclusion

This chapter studied a browser workbench for monitoring and diagnosing plant records, with five tools acting on records generated by other simulators of the collection with every change declared. On the water-network record a 3 % meter shift raises the global test to 313.67 and every measurement test above its limit, and serial elimination names the station meter with a gross error of 2.35 L/s,

after which the balances close at 0.135. On the Tennessee Eastman records Q detects the cooling water step at 48.05 h while T^2 never does, a contribution plot reaches the feed only 1.8 h after the feed ratio step, and the D feed temperature step is not detectable at the measured false-alarm rate of 1.25 detections per normal day over 61 seeds. The loop records are told apart by the valve plot (93.9 % against 0.4 % held) and not by their indices, and rationalization takes an alarm hour from 494 annunciations to 14. Independent checks disagree with two cells of a published run-length table and trace a gap in a published example to its simulated errors. The methods are linear and assume independent samples, and every record is generated. A fuller workbench would add bilinear and dynamic reconciliation, recursive and kernel monitoring, a stiction test without the valve position, state-based alarm suppression and a recorded plant historian. The workbench shows that a statistic above its limit is the start of a diagnosis and not its answer.

KEY POINTS

- A statistic above its limit is the start of a diagnosis and not its answer.
- One gross error spreads into every stream that shares a balance with it: a 3 % shift on one flowmeter puts all seven measurement tests above their limit.
- Limits assume independent samples, so on a closed-loop plant Q lies above its 99 % limit on 4.3 % of samples on normal days, and one fault is not detectable at that false-alarm rate.
- The loop indices do not tell a sticking valve from an aggressive loop, while the valve plot does (93.9 % against 0.4 % held).

Where else this applies. The drifting analyzer that its own loop holds at the setpoint is the kappa loop of Volume 5, Chapter 4, and the sticking steam valve of loop A belongs to Volume 2, Chapter 4.

References

[1] R. S. Mah, G. M. Stanley, and D. M. Downing, "Reconciliation and rectification of process flow and inventory data," *Industrial & Engineering Chemistry Process Design and Development*, vol. 15, no. 1, pp. 175–183, 1976.

[2] S. Narasimhan and C. Jordache, *Data Reconciliation and Gross Error Detection: An Intelligent Use of Process Data*. Gulf Professional Publishing, 1999.

[3] J. V. Kresta, J. F. MacGregor, and T. E. Marlin, "Multivariate statistical monitoring of process operating performance," *The Canadian Journal of Chemical Engineering*, vol. 69, no. 1, pp. 35–47, 1991.

[4] T. J. Harris, "Assessment of control loop performance," *The Canadian Journal of Chemical Engineering*, vol. 67, no. 5, pp. 856–861, 1989.

[5] International Society of Automation, "Management of Alarm Systems for the Process Industries," ANSI/ISA-18.2-2016, Research Triangle Park, NC, USA, 2016.

[6] Health and Safety Executive, "Better alarm handling," Health and Safety Executive, HSE Information Sheet Chemicals Sheet No 6 (CHIS6), 2000, information current at 3/00. [Online]. Available: https://www.hse.gov.uk/pubns/chis6.pdf

[7] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Computers & Chemical Engineering*, vol. 17, no. 3, pp. 245–255, 1993.

[8] R. S. H. Mah and A. C. Tamhane, "Detection of gross errors in process data," *AIChE Journal*, vol. 28, no. 5, pp. 828–830, 1982.

[9] C. M. Crowe, Y. A. Garcia Campos, and A. Hrymak, "Reconciliation of process flow rates by matrix projection. Part I: Linear case," *AIChE Journal*, vol. 29, no. 6, pp. 881–888, 1983.

[10] S. Narasimhan and N. Bhatt, "Deconstructing principal component analysis using a data reconciliation perspective," *Computers & Chemical Engineering*, vol. 77, pp. 74–84, 2015, preprint arXiv:1505.00314.

[11] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1-2, pp. 100–115, 1954.

[12] S. W. Roberts, "Control chart tests based on geometric moving averages," *Technometrics*, vol. 1, no. 3, pp. 239–250, 1959.

[13] D. Brook and D. A. Evans, "An approach to the probability distribution of cusum run length," *Biometrika*, vol. 59, no. 3, pp. 539–549, 1972.

[14] J. M. Lucas and M. S. Saccucci, "Exponentially weighted moving average control schemes: Properties and enhancements," *Technometrics*, vol. 32, no. 1, pp. 1–12, 1990.

[15] W. F. Guthrie, "NIST/SEMATECH e-Handbook of Statistical Methods (NIST Handbook 151)," National Institute of Standards and Technology, 2020, section 6.3.2.3.1, Cusum Average Run Length. [Online]. Available: https://www.itl.nist.gov/div898/handbook/

[16] J. E. Jackson and G. S. Mudholkar, "Control procedures for residuals associated with principal component analysis," *Technometrics*, vol. 21, no. 3, pp. 341–349, 1979.

[17] W. Ku, R. H. Storer, and C. Georgakis, "Disturbance detection and isolation by dynamic principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 30, no. 1, pp. 179–196, 1995.

[18] N. F. Thornhill, B. Huang, and H. Zhang, "Detection of multiple oscillations in control loops," *Journal of Process Control*, vol. 13, no. 1, pp. 91–100, 2003.

[19] M. A. A. Shoukat Choudhury, N. F. Thornhill, and S. L. Shah, "Modelling valve stiction," *Control Engineering Practice*, vol. 13, no. 5, pp. 641–658, 2005.

[20] 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.

[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, Gaithersburg, MD, Special Publication 800-82r3, 2023.

[22] E. Fernandes Montesuma, M. Mulas, F. Ngolè Mboula, F. Corona, and A. Souloumiac, "Benchmarking domain adaptation for chemical processes on the Tennessee Eastman process," arXiv preprint 2308.11247v2, 2024, published in Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Communications in Computer and Information Science, Springer, 2026, pp. 307–322, doi:10.1007/978-3-032-25305-7_21. [Online]. Available: https://arxiv.org/abs/2308.11247

13.11 Exercises

13.1 ○ Concept check · 2 min

On the water-network record of case 5 only the station flowmeter carries the declared 3 % shift. On the window from 6 to 30 h, all seven measurement tests still exceed their limit of 2.68. Why?

(a) The declared standard deviations of the zone meters are too small, so noise inflates their tests.

(b) The main break at 10 h adds a second gross error, which lands in the zone 3 meter.

(c) The Šidák correction pulls the limit below the noise level that every test carries.

(d) The error enters both balances, and reconciliation spreads it into every stream that shares a balance with the meter.

13.2 ○ Concept check · 2 min

On the cooling water step of case 3, trained with the defaults, Q detects at 48.05 h while T^2 never detects. Which statement explains the difference?

(a) The step moves samples off the plane of the 11 retained components, which Q measures, while their scores stay inside the T^2 limit.

(b) T^2 is built from the 39 discarded eigenvalues, so it is blind to the temperatures that the step moves.

(c) The T^2 limit is set at 99.9 % confidence and the Q limit at 99 %, so T^2 needs a larger move.

(d) T^2 must stay above its limit for three samples and Q for one, so T^2 responds later than Q.

13.3 ○ Concept check · 1 min

Over 61 normal days, Q lies above its 99 % limit on 4.3 % of samples rather than 1 %. What does the chapter give as the main reason?

(a) The Jackson-Mudholkar formula underestimates the residual quantile of any data set.

(b) The plant runs in closed loop, so its samples are correlated and its operation wanders between days.

(c) The two-day training window is too long and mixes in drift, which inflates the limit.

(d) The run-of-three rule triples the share of samples that are counted above the limit.

13.4 Theory and calculation · 8 min

The water-network tests of Section 13.4 use a Šidák-corrected normal limit and a χ^2 global test at $\alpha = 0.05$. The window means from 6 to 30 h ($n = 97$ samples) are, in L/s: pump curves 75.54, station meter 77.85, tank inflow -2.37, zone meters 29.36, 24.07 and 21.53, and trunk leakage 2.86.

(a) Compute the Šidák limit z_c for $m = 7$ and $m = 6$ redundant measurements, and the Bonferroni limit $\Phi^{-1}(1 - \alpha/(2m))$ for $m = 7$. *Checkpoint:* 2.68, 2.63 and 2.69.

(b) The cumulative distribution of $\chi^2(2)$ is $1 - e^{-x/2}$, and a $\chi^2(1)$ variable is the square of a standard normal one. Find the 0.95 quantile of each. *Checkpoint:* 5.99 and 3.84.

(c) With the station meter removed, the station flow can be estimated by two independent routes: the pump curves ($\sigma = 1.7$ L/s per sample) and the sum of the junction outflows (tank 1.8, trunk 0.86 and zone meters 0.1, 0.093 and 0.076 L/s). Find each route's estimate and standard error over the window, combine them by inverse-variance weights, and compute the station meter's gross error. *Checkpoint:* 75.54 L/s $\pm$ 0.173 and 75.45 $\pm$ 0.203; combined 75.50 L/s $\pm$ 0.132; gross error 2.35 L/s.

13.5 Theory and calculation · 6 min

The product-flow chart of case 4 (XMEAS 17) is trained with mean 22.9494 and standard deviation 0.1169 m^3/h, with the defaults $\lambda = 0.2$, $L = 2.962$, $k = 0.5$ and $h = 5$ of Equation (13.5).

(a) Compute the EWMA limit $L\sqrt{\lambda/(2-\lambda)}$. *Checkpoint:* 0.987.

(b) At the sample before the first signal, $C^- = 3.25$ and $w = -0.80$. The next sample has $z = -2.46$. Find the product flow it reads, and the new C^- and w. Do the charts signal? *Checkpoint:* 22.662 m^3/h; $C^- = 5.21$, $w = -1.132$; both signal.

(c) The in-control run length of Shewhart rule 1 is $1/p$, with p the two-sided probability beyond 3σ. Compute it, and convert it and the EWMA's 499.3 samples into hours at the record's 3 min sampling. *Checkpoint:* 370.4 samples, 18.5 h; 25.0 h.

13.6 Theory and calculation · 6 min

A detection on the Monitor tab is a run of three consecutive samples above a limit. The Tennessee Eastman records are sampled every 3 min.

(a) If samples were independent and each lay above the limit with probability p, a run of three starts at a given sample with probability about $(1 - p)p^3$. How many runs of three start per day at the nominal $p = 0.01$, and how many years pass between them on average? *Checkpoint:* 4.75×10^{-4} per day, about 5.8 years.

(b) Repeat with the measured share $p = 0.043$ of normal days. *Checkpoint:* 0.037 per day.

(c) Over 61 normal days the defaults raised 1.25 detection episodes per day. By what factor does that exceed (b), and what does the factor say about the record? *Checkpoint:* About 34.

13.7 App challenge · 8 min

Use scenario 5 · The water network: find the meter that lies. The Reconcile tab opens with the declared balances, the window 6 to 30 h and $\alpha = 0.05$. The record is a fixed file and the tools are deterministic, so every value below repeats to the digits the app shows.

(a) Press Reconcile. Read the global test γ and its limit, and the test z of the station flowmeter, the pump curves and the zone 1 meter. *Checkpoint:* $\gamma = 313.67$ against $\chi^2(2)$ 5.99; z 17.71, 13.20 and 11.68.

(b) Press Serial elimination. Read which stream is removed, its estimated gross error and the new γ. *Checkpoint:* Station flowmeter, 2.35; γ 0.13 against 3.84.

(c) Set the Kind of Trunk main leakage (law) to unmeasured and press Serial elimination again. Read the tiles and the elimination step, and explain the tie. *Checkpoint:* $\gamma = 177.21$ against $\chi^2(1)$ 3.84, one reduced balance; the pump curves are removed, tied with the station flowmeter.

13.8 App challenge · 8 min

Use scenario 3 · A cooling water step: the two statistics disagree. On the Monitor tab keep the defaults (training 0 to 48 h, cumulative variance 50 %, lags 0, confidence 99 %, three in a row). The record is a fixed file, so the values repeat exactly.

(a) Press Train. Read k, both limits, both first detections and the share of samples above each limit after training. *Checkpoint:* $k = 11$; 25.22 and 42.59; Q 48.05 h, T^2 none; 12.5 % and 0.8 %.

(b) Set Lags (dynamic PCA) to 2 and train again. Read k and both detections. *Checkpoint:* $k = 17$; Q 48.05 h, T^2 48.10 h.

(c) Return Lags to 0, set Confidence to 99.9 % and train. What happens, and what does this say about raising the limit to cut the false alarms of case 1? *Checkpoint:* Neither statistic detects.

13.9 App challenge · 7 min

Use scenario 8 · An alarm flood: rationalize it. The Alarms tab opens with no deadband, filter or delays. The metrics are counts over a fixed record and repeat exactly.

(a) Read the hour's annunciations, the number of ten-minute periods in flood, and the chattering tags and the annunciations they carry. *Checkpoint:* 494; 6; 5 tags with 473.

(b) Set Deadband 2 %, Filter τ 5 s, On-delay 15 s and Off-delay 60 s. Read the same metrics. Does the setting meet the case's target? *Checkpoint:* 14 annunciations, no flood, no chattering tag; yes.

(c) Set Deadband 1 %, Filter τ 0 s, On-delay 10 s and Off-delay 30 s. Read the total and the busiest period. *Checkpoint:* 16, with one flood of 11 between 30 and 40 min.

13.10 Application area · 4 min

A water utility reads its station flowmeter for production reports. The meter of case 5 reads 3 % high from 4 h on.

(a) With the gross error of 2.35 L/s that elimination estimates, how much water does the station overstate per day? *Checkpoint:* About 203 m^3 per day.

(b) With the trunk leakage declared unmeasured, the pump curves and the station meter tie. What must a utility measure or model so that a biased station meter can be named?

13.11 Application area · 4 min

In the digester record, the laboratory samples every 4 h and reports 1 h later, and the analyzer samples every 15 min and reports 20 min later. The six paired differences (analyzer less laboratory) at 0, 4, 8, 12, 16 and 20 h are 0.05, 0.67, 2.60, 3.37, 4.35 and 5.93.

(a) Fit a least-squares line to the differences against time. Compare its slope with the declared drift. *Checkpoint:* 0.2944 kappa/h against 0.3 kappa/h.

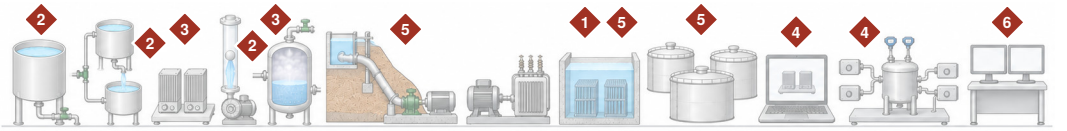
(b) An engineer pairs each laboratory result with the analyzer value on the screen when the result arrives. How far apart in time were the two samples taken, and how much bias does the drift alone add to the pair? *Checkpoint:* 30 min; 0.15 kappa.

13.12 Review the agent · 5 min

An AI assistant reads the loop table of case 7 and writes: “Loop A has a decay ratio of 0.904 and a Harris index of 0.0095, against 0.790 and 0.341 for loop B, so loop A is the more badly tuned of the two. Halve its controller gain, and leave loop B alone.” Find the flaw, and say how you would check it in the app.

CHAPTER 14

Summary and Outlook



Open gaps 1 slow and integrating tests 2 operating-point models 3 instrument in the model
4 physics-aware structure 5 untestable events 6 correlated monitoring

Notice that the fuel pool carries two gaps and that the laboratory plants on the left met the gaps of operating point and instrument, while the large plants met the gap of events that cannot be tested. Illustration, not to scale.

IN PLAIN TERMS **What Twelve Plants Taught About Models**

The main idea. Every chapter of this book built a model of a plant, from a water tank to the world oil market. The model came from counting what flows in and out, from pushing the plant and recording its answer, or from both.

The challenge. A curve read from one test belongs to that test and can mislead on the next. Some events, such as a fuel pool boiling, can never be tried, so their models are never checked against the real thing.

Why it matters. Checking each model on a second record, and over its whole range, catches most of the errors these chapters found.

Twelve chapters built models of twelve plants, from a tank of water to a world oil market. Each chapter had to decide which form of model its plant needed and where the constants would come from: a balance, a test, a record, or a mix of the three. This chapter sets those decisions side by side with the results they produced, draws a short guide for choosing a method, looks at where modeling and identification practice is heading, reports how the volume itself was made with AI agents and what they got wrong, and lists the problems that remain open.

14.1 What the Chapters Showed

Table 1.2 of Chapter 1 listed, for each chapter, the inventory that sets the dynamics, the model form and how the chapter obtains it. Tables 14.1 and 14.2 add the results. Each row names the difficulty the plant posed, the model that worked and how it was obtained, a model or reading that failed, and the measured result with the table or section of the chapter that reports it.

Figure 14.1 draws one like pair of those results from each row. The columns of the two tables make five points.

The balance gave the form, and a test gave or checked the numbers. The tank's gain and time constant both moved with the level, and the balance said in advance that their ratio would not. Every step test then started at the same slope. The two tanks returned one gain at every splitter setting, as the balances said they would, while the dynamics changed. The twin of Chapter 4 reproduced the gain matrix of its own energy balances. In each case the balance told the experimenter what to expect and the test confirmed or corrected the constants.

A step test reads a model that belongs to the method and the window. None of the plants of Chapter 3 or Chapter 4 has a transport delay, yet both fits returned one. The two-tank fit read 18.5 to

Table 14.1 Which Model Each Plant Needed, Chapters 2 to 7

Chapter	Difficulty	Model that worked, and its source	What failed	Measured result
2, gravity-drained tank	Gain and time constant both grow as $\sqrt{h}$. An integrator with the outlet shut.	The mass balance linearized at each level, checked by 1 % step tests and a drain test.	A step that moved the level 21 % at 2.04 m: its fitted time constant sat 7.0 % above the tangent value.	Fitted gain up 2.14 times and time constant 2.12 times from 2.04 to 9.88 m, against 2.20 predicted, and a drain in 19.65 against 19.79 s (Section 2.4).
3, two tanks and a splitter	The splitter sets the order of the plant. Its gain does not move.	Linearized balances, two lags whose pole and zero cancel with the splitter open. Two-point fits at three settings.	A tuning from the open-splitter fit run with the splitter shut. A predictive model fitted shut and run half open.	Gain 0.0323 m/% at every setting, apparent dead time 18.5, 4.0 and –0.6 s (Table 3.3). Overshoot 58.7 % (Table 3.5). Half-open IAE 8.68 fixed, 1.00 scheduled, 2.65 for the PID (Table 3.8).
4, two heaters	Four lags. A thermistor slower than the heater it measures.	Energy balances of heaters and sensors, a 2×2 gain matrix from manual steps.	The fitted time constant read as the heater's. A PID loop that meets the sensor lag only as error.	Gains 0.599, 0.050, 0.100 and 0.300 °C/%, $\tau_p = 141$ and $\theta_p = 18$ s on a 140 s sensor (Table 4.3). Loop 2 upset 0.20 under the predictive controller against 1.68 °C (Table 4.7).
5, ball in a tube	An integrator behind a valve whose gain falls from 30.4 to 1.0 mm/s per degree over its stroke.	A momentum balance about hover with a valve map, a drop test and step tests on the twin.	The valve without its orifice. A 200 ms sensor.	Usable band 1.19° without the orifice and 22.34° with it (Section 5.4). IAE 1880 with the 200 ms sensor against 72 at 66 ms (Table 5.6). No number yet from the rig.
6, flash vessel	The level reads water plus the vapor below the surface. It swells while the vessel loses water.	Mass and energy balances closed by steam tables. A controller that expects the swell.	Single-element control, which reads swell as inventory. An inventory estimate under foaming water.	Swell 3 points on a 25 % demand rise (Section 6.4.2). IAE 399, 232 and 152 % s for single-element, three-element and predictive (Table 6.4).
7, hydro unit	Water inertia: opening the gate first cuts the power.	The water column linearized to a right-half-plane zero. A computed stability map checked against the engine.	The design plan's zero at $2/T_w$. A governor tuned on the grid and run on the island.	Dip 4.38 MW, boundary gain 11.73 mapped and 11.70 in the engine (Table 7.4). Zero 0.675 rad/s at 80 MW (Section 7.4.1). Island 56.009 to 62.479 Hz (Table 7.5).

Table 14.2 Which Model Each Plant Needed, Chapters 8 to 13

Chapter	Difficulty	Model that worked, and its source	What failed	Measured result
8, grid frequency	One event, three quantities, each set by a different group of parameters.	The swing equation in closed form: rate of change from inertia alone, offset from droop and load damping alone.	Droop alone after a unit trip, which leaves the grid at 59.429 Hz (Table 8.5).	Rate of change −0.150, −0.299 and –0.746 Hz/s at the same settled 59.857 Hz. Settled 59.941, 59.857 and 59.679 Hz at one rate (Table 8.4).
9, spent-fuel pool	Self-regulating with the cooling trains running, integrating without them. Hours per time constant.	A full step test for the running pool. The two balances marched forward to the limits for the stopped one.	The first hour read as an integrator. The present rate extrapolated to boiling.	$K = -0.0551\,^{\circ}\mathrm{C}/\%$, $\tau = 3.19$ h, $\theta = 0.16$ h, and an integral time 2.6 times too long from the first hour (Section 9.5). Boiling forecast within 0.10 against 1.12 h early (Section 9.6).
10, oil market	One tank of stocks with a short-run and a long-run gain.	Balances linearized, gains of \$12.5 and \$1.34 per million barrels a day (Section 10.4), a factor of 9.3 (Section 10.9).	A PI that holds \$100 on the short-run picture.	Price \$98.3 on average while the share falls from 30 to 15.27 %. The economic predictive controller at a mean of \$62.1 keeps 32.64 % (Table 10.5).
11, heater with three balances	Nested model forms. Only the products UA and mc_p are identifiable.	The convective form fitted on one step test and judged on a second, untouched one.	A radiation term that improved its own record and reported a radiative share in data generated without radiation.	Objective 141.9 to 134.2 on the fitted step, 283.8 to 429.0 °C on the held-out step (Table 11.5). Share 70.1 % (Table 11.6).
12, identification workbench	Several inputs that move at once. Paths that may be absent.	Transfer-function matrices with exact structural zeros, ranked on free-run error over held-out data.	A zero imposed on a strong path. A FIR model on an integrating level.	Six gains within 0.9 % (Table 12.5). Validation 0.1093 to 0.1085 for a true zero, 0.3892 for a false one (Table 12.6). Tank 4.386 FIR, 0.062 state space (Table 12.4).
13, monitoring workbench	Faults hidden in records of normal operation. Correlated samples.	Balances for reconciliation, a principal component model of normal data, the valve plot.	T^2 on the cooling water step. Loop indices for a valve fault.	Station meter named at 2.35 L/s, test 313.67 to 0.135 (Section 13.5). Q at 48.05 h, T^2 never (Table 13.3). False detections 1.25 a day (Section 13.5.3).

–0.6 s as the splitter opened, and the heater fit read a time constant that belongs to the thermistor. The first hour of the pool's response looks like an integrator and gives an integral time 2.6 times too long. The oil market's short-run gain is 9.3 times its long-run gain. A fitted constant is true of the plant only together with the method and the window that produced it (Section 1.4.6).

Where a test is unsafe or impossible, the balance carries the analysis. No one steps a penstock to find its zero, boils a spent-fuel pool to time it or runs a decade of oil prices twice. The hydro chapter derived its zero and checked a stability map against the engine to about 0.26 %. The pool chapter marched its balances to boiling and beat extrapolation by an hour. The flash vessel's swell came from balances closed by steam tables and a declared bubble closure, not from a fit. The same chapters show the cost: a balance is only as good as its weakest correlation, and the ball's twin still waits for its rig.

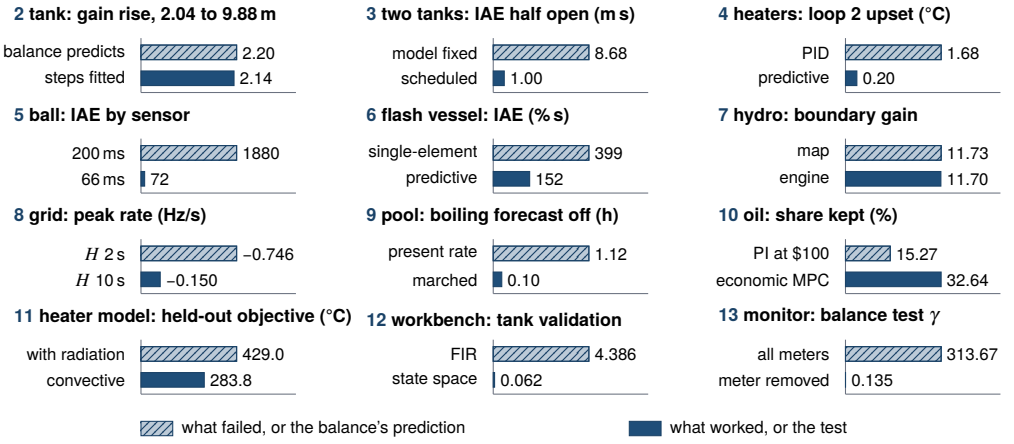


Figure 14.1 One like pair of numbers from each row, as the "Measured result" column of Tables 14.1 and 14.2 prints them. Each panel has its own scale from zero, and the grid panel's axis is reversed so that its negative rates grow to the right. In most panels the hatched bar is the model, reading or setting that failed and the solid bar the one that worked. In the draining-tank and hydro panels the hatched bar is the balance's prediction and the solid bar the test or the engine, and in the grid panel the two bars are the two inertias. Notice that the bars of every failure panel differ plainly, while a balance and its test agree so closely that their bars nearly match.

More parameters fit better and predict worse. The radiation term of Chapter 11 lowered the objective on its own record and raised it from 283.8 to 429.0 °C on the held-out record, and it reported a 70.1 % radiative share in data that contained none. A structural zero on a strong path in Chapter 12 raised the validation error from 0.1093 to 0.3892, and the FIR model, with hundreds of coefficients, was the worst of five families on the integrating level. In both chapters a second record decided the structure and the objective of the first record did not (Section 1.4.9).

A statistic flags, a balance names. The workbench of Chapter 13 named the faulty meter by eliminating measurements until the balances closed. Its statistical monitor detected the cooling water step in Q and never in T^2, and on normal days it raised 1.25 detections a day where its limit assumed far fewer, because the samples were not independent. A statistic says that the record departs from normal data. A model of the plant says where and why.

14.2 Choosing a Method

The chapters suggest an order of questions for a new model. Each question names the chapter where its answer decided the result, and Figure 14.2 draws the same order.

1. *Which inventory sets the dynamics, and can its balance be written?* Write it first, even when the constants will come from data. It gives the form, the sign of each gain and the groups of parameters a test can identify, as UA and mc_p in Chapter 11.

2. *Does the operating point move the constants?* Linearize at several points (Section 1.4.3). Gain and time constant that move together, as in Chapter 2, leave the initial slope fixed. A gain that moves by a large factor, as the ball's valve gain does, needs a model at each operating point or a schedule.

3. *Is the plant self-regulating or integrating over the horizon that matters?* Run the test long enough to tell the two apart. The pool of Chapter 9 is both, depending on which pumps run.

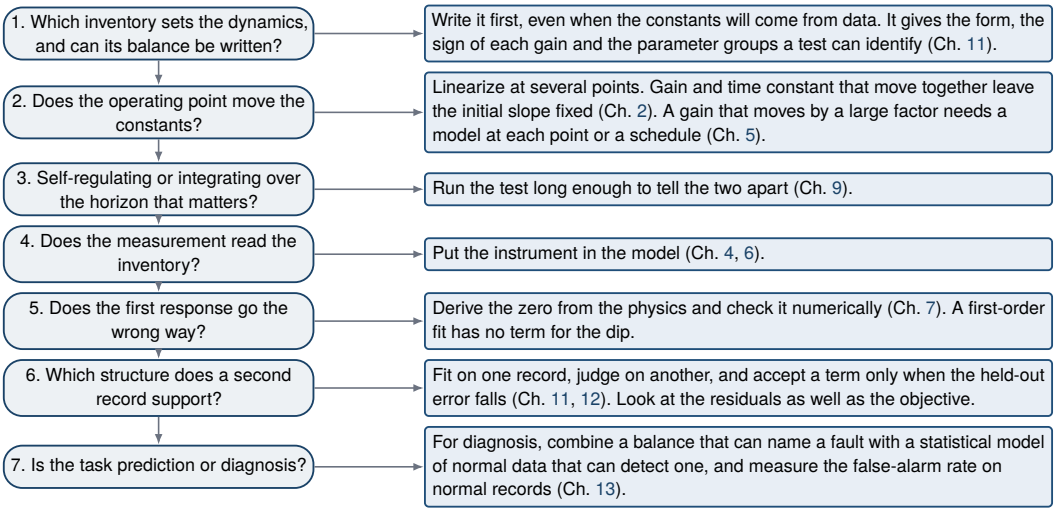


Figure 14.2 The seven questions of this section in order, with the list's answer and the chapters where it decided the result. Notice that the balance comes first and the second record comes late: the form of a model is settled before any data are fitted, and its structure is settled only after a fit has been judged on data it has not seen.

4. *Does the measurement read the inventory?* A thermistor, a swelling level and a thermowell add dynamics or reverse the first response. Put the instrument in the model (Chapter 4, Chapter 6).
5. *Does the first response go the wrong way?* Derive the zero from the physics and check it numerically, as Chapter 7 does. A first-order fit has no term for the dip (Section 1.4.7).
6. *Which structure does a second record support?* Fit on one record, judge on another, and accept a term only when the held-out error falls (Chapter 11, Chapter 12). Look at the residuals as well as the objective.
7. *Is the task prediction or diagnosis?* For diagnosis, combine a balance that can name a fault with a statistical model of normal data that can detect one, and measure the false-alarm rate on normal records (Chapter 13).

A predictive controller showed what each model was worth. It cut the two-tank IAE where its model matched the splitter and scored 3.3 times the PID's where it did not. It held the heater twin's second loop within 0.20 °C because it carried the conduction between the heaters, which is a matched-model result on a simulator. Volume 9 treats predictive control in full.

14.3 Where Practice Is Heading

Identification meets machine learning. Identification with kernels and regularization links system identification to machine learning and function estimation [1]. Common deep networks are nonlinear ARX models, and recurrent networks of the LSTM kind are nonlinear state-space models, so both fit into the established code and practice of identification [2]. For plants that are not linear, a guide for practicing engineers walks through the choices a user must make and the ways the nonlinear problem differs from the linear one [3]. In the editor's view the names of the model families will keep changing, while the questions of Chapter 11 and Chapter 12, which structure a second record supports and which parameter a record cannot determine, will not.

Physics and data are combined. Hybrid models join parametric submodels from first principles with nonparametric submodels fitted to data. A review of their use in process systems engineering found that they do not lead to better results automatically, although a reasoned combination of knowledge can improve model-based operation and design [4]. Physics-informed machine learning trains networks on data together with the physical laws they must obey [5]. Sparse regression can select the fewest terms of a governing equation that represent the data, trading accuracy against complexity to avoid overfitting [6]. Language models have been used to propose the skeleton of an equation from domain knowledge, with the parameters then fitted to data. On four benchmark problems this found more accurate equations than symbolic regression baselines, most of all outside the range of the training data [7]. The radiation term of Chapter 11 is the warning for all of these methods: a physically plausible term can fit noise and report a physical quantity that is not there.

> ⧖ HISTORY **1976 to 2003: The Questions Industry Asked About Models**
>
> The timeline of Chapter 1 ends its testing thread in 1976, with the first predictive control applications and their impulse-response models identified from plant test data. A survey of industrial predictive control published in 2003 found that the number of applications had about doubled from 1995 to 1999, and that most of these controllers still used empirical dynamic models identified from test data, so shorter plant tests would have a large positive impact for their users. It found the research needs that industry perceived to be mostly engineering issues rather than algorithms, among them when a model is good enough to stop the identification test and when the added expense of a nonlinear controller can be justified [8]. In the editor's reading, the twelve chapters of this volume answered both questions for one plant at a time, and neither has a routine answer yet.

Twins run beside plants, and monitors watch them. A digital twin is a virtual representation of an asset, kept current by data and simulators, for prediction, monitoring, control and decisions [9]. Data-driven monitoring and fault diagnosis has a large literature of its own [10]. The app of Chapter 4 runs its twin beside the board it models under one controller, and Chapter 5 shows the other side: a twin whose numbers have not yet met their rig. In the editor's view a twin is only as current as its last comparison with the plant, and that comparison, with its date and its residuals, belongs in the twin's record.

Models are corrected from routine data. The performance of a predictive controller depends closely on its model, and a process drifts from the model identified at commissioning as feedstocks, instruments, equipment and operating conditions change. A step test to re-identify it disturbs closed-loop operation, while routine setpoints may stay constant for several months. One study re-identifies the model from routine operating data with a regularized regression and, in numerical experiments on a pilot-scale distillation column, corrected mismatch in static gain and in transport delay without injecting an excitation signal [11]. The predictive controller of Chapter 3 lost to the PID when its model no longer matched the splitter. In the editor's view such a method can be trusted only as far as it states which constants the routine record could not determine.

Learned models are given a physical shape. Relying on data alone does not guarantee that an identified model keeps the physical properties of the system. One study builds monotonic behavior into a neural network model, trains it on 250 points from a two-heater laboratory kit with heater inputs between 10 and 50 percent, and runs it in a predictive controller whose input reaches 65 percent, outside the training range [12]. The kit is the board of Chapter 4. In the editor's view the constraint does the work that a balance did in this volume: it fixes the sign of each gain before the data arrive.

Property models run fast enough for the loop. Chapter 6 closes its balances with the industrial formulation of steam properties, which the steam power industry uses as its standard for contracting

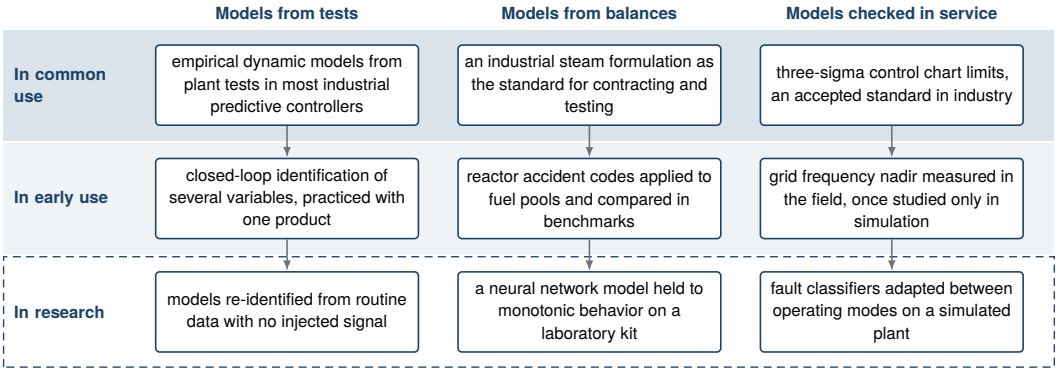


Figure 14.3 Developments in three threads. The top lane holds what the cited source calls standard or common practice, the middle lane what it reports in service at some sites or on field data, and the bottom lane what it reports from simulation, from a laboratory kit or a benchmark, in the editor's reading. Sources by thread: models from tests [8, 11], models from balances [12, 13, 15] and models checked in service [16–18]. Notice that each thread moves from a model fixed once toward one that is corrected or checked while the plant runs.

and testing [13]. A guideline from the same international association describes a spline-based table look-up intended for fast property calculations in simulations of non-stationary processes and in real-time process optimization, where the usual equations of state cost too much computing time. In flow simulations it cut computing times by factors between 6 and 10 against the industrial formulation used directly [14]. In the editor's view this brings the full balance of a steam vessel closer to the controller that needs it, so that the model of record and the model in the loop can be the same. Figure 14.3 sets developments of modeling practice in three threads.

14.4 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 [19]. 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 [20]. 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.

14.4.1 How This Volume Was Made

Each of the twelve apps was built and tested by agents. An agent then wrote the paper on it, keeping verification notes that trace every number to a run, and each paper received a referee report. At least three of the twelve reports are self-reviews by the agent that wrote the paper. Defects found along the way went into a log for each app with a severity, a cause, a suggested fix and a status. For this book one agent per chapter reran the numbers its exercises use. The editor set the scope of each round and what each agent could change, and statements that runs contradicted were left for the editor to decide.

The defect logs. The twelve logs hold 117 entries with a severity. Of these, 30 record wrong behavior of an app (one of them records the fix to another), 60 are documentation, 17 are cosmetic or display and 10 are other kinds, such as content and requested changes. The cases below are the ones that concern models.

A correlation used with the wrong definition. The pool's evaporation correlation takes the density of dry air per unit volume of moist air. The code added the vapor's own density. At the design surface temperature of 35 °C the rate came out 13.9 % low, and at the 95 °C cap that drives the loss-of-cooling case it came out 3.3 times high. The paper agent found it by measuring the rate in the running model at five temperatures and comparing it with the published correlation the code cites. The fix was first deferred, with the departure stated in the paper. Once applied it moved boiling after a loss of cooling from 16.88 to 14.74 h, while the design point still closed. A check at the design point alone would have passed the error.

Records that were not what they claimed. In the two-heater twin the exported temperature carried a fresh draw of sensor noise, independent of the one the controller acted on, so the two differed by up to about 0.6 °C at the default noise. A model fitted to that export fits the wrong pair of numbers. In the model-selection app the first-order form exported a heat loss in degrees per second under a column headed in watts, visible only in the export. Both were found while the papers were written, and the papers' runs avoided both.

Held-out data inside the training. The identification workbench estimated parameter spread with a bootstrap that drew residual blocks from the validation and test parts of each record. Held-out data thus shaped the training refits. The blocks now come only from training stretches. In the same app every exported Python model stopped with an error at line 57. The test that would have caught it was skipped whenever a Python environment was not named. A check that needs no Python now runs in every build.

A derivation against the plan. The design plan for the hydro app placed the turbine's zero at $2/T_w$, which is the magnitude of the pole. The derivation, checked with the framework's eigenvalue routine, puts it near $1/T_w$, at 0.675 rad/s at 80 MW. The app follows the derivation, and the referee report records the reason.

The controller's model was the plant. Section 1.4.9 separates the simulation model, the identified model and the controller's model. In the oil market the economic predictive controller linearized the simulator with the run's own parameters, every hidden shock in force and the true state. At one meeting, 23 of 31 hidden changes moved its request while the price it could observe stayed the same. Its model is now one frozen design and it starts from its own estimate, and the paper's results for it were measured again. In the tank and two-tank apps the controllers started from the run's own balances. There no measured run moved when the read was closed. Three logs record such reads.

A rig claim with no rig. The referee of the levitation paper found a draft that described the hardware of the surveyed rigs beyond what their publications state, and numbers that could not be regenerated from the shipped engine. The claims were softened and the numbers removed, and the chapter says plainly that none of its results has met the hardware.

The book round. The twelve chapter records list 206 rows of numeric verification, each a stated value set beside its reproduction, and every value an exercise uses reproduced. Three records found drift elsewhere: a count of checks that had grown since the paper, a recorded alarm file that no longer regenerates byte for byte, and a gain in a limitations paragraph that did not recompute, 3.68 where the fit gives 3.545 °C/%, since corrected in the paper and the chapter. The levitation Conclusion said a dead-band cycle held "whatever the tuning" while its own results showed retuning moving the cycle between ± 1.52 and ± 6.35 mm, and it was aligned with the results. The agent writing Chapter 1 had an ARX delay index off by one sample until a noise-free test showed which index fitted to machine precision.

Every error above was caught by a check that could fail: a correlation evaluated across its range, a record compared with what the controller saw, a test that actually ran, a derivation checked numerically, a hidden value moved to see what else moved. Four working rules follow. Check a model over its whole range, not at its design point. Fit on one record and judge on another. Know what each recorded column is. Read what a controller's model contains.

14.5 Gaps in Modeling and Identification

Six gaps belong to this volume's topic, and the gaps common to agent-era practice follow them in Section 14.5.1. 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 14.4 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. *Tests for slow and integrating plants.* The pool's time constant is hours, and its first hour reads as an integrator. A survey of industrial predictive control found no tool to determine whether the collected data are adequate to represent the process dynamics, so that in practice the plant is generally over-tested [8]. The gap is hard because a long lag and an integrator trace nearly the same curve early in a test, and here a long enough record means hours of a plant held off its normal duty. A routine method is needed that says, while a test runs, whether the record can yet separate a long lag from an integrator and what the tuning would lose by stopping now.

2. *Models that follow the operating point.* The tank's constants move with the level, the ball's valve gain falls from 30.4 to 1.0 mm/s per degree and the two-tank predictive model lost its advantage when the splitter moved. The Open challenge box after this list takes the gap further.

3. *The instrument as part of the model.* A thermistor set the heater's apparent time constant, and foaming water misled the flash vessel's inventory estimate with its residual small. A report on steam-generator level control found unplanned reactor trips due to that loop to be significant contributors to plant unavailability, with the measured level reacting at first in reverse to a change of inventory, and named the mismatch between steam and feedwater flow as the only true indication of an inventory change [21]. The gap is hard because the instrument and the process reach the record as one response, and a fit has no reason to split it. Methods are needed that identify an instrument's dynamics apart from the process and notice when the physics of the measurement changes.

4. *Structure selection that knows the physics.* The radiation term fitted noise and reported a physical share, and a structural zero imposed on a strong path in Chapter 12 raised the validation error from 0.1093 to 0.3892. A review of digital twins notes that sparse regression recovers a model's terms from many candidates in little computing time but performs poorly on data corrupted with noise, and that the errors of a purely data-driven model cannot be bounded [9]. The gap is hard because a wrong term and a right one both lower the objective on the record they were fitted to. Before a term is fitted, a check of identifiability and of the information a record holds about it should be as routine as the fit itself.

5. *Models for events that cannot be tested.* Load rejection on a penstock, boiling in a pool and a decade of oil prices are predicted from balances and correlations with no record of the event. The evaporation defect shows how far such a prediction can move. A status report on spent fuel pools found that the experiment database consists largely of tests made for reactors, with only one integral test targeted at a pool accident, that the available codes are intended primarily for reactor accidents, and that benchmarks between codes helped identify their limitations [15]. A standard for models and simulations requires a record of each model's domain of validation and allows extrapolation outside it only within the model's limits [22]. The gap is hard because the event that would validate the model is the event the plant exists to avoid. A practice for validating these models, against bench rigs, partial events and independent codes, is still thin.

6. *Monitoring with correlated data and rare faults.* The monitor's limit assumed independent samples and raised 1.25 detections a day, and a small feed temperature step stayed below what that rate allows. A handbook of statistical methods gives the standard three-sigma chart a false alarm every 371 points on average and notes that such properties are generally stated as if the

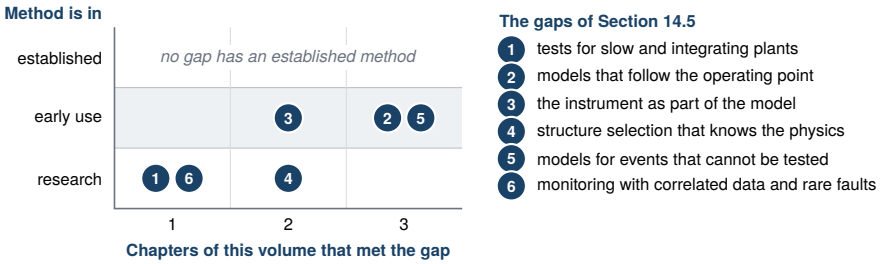


Figure 14.4 The six gaps of this section. Across is the number of chapters that each item names as evidence: one for gap 1 (Chapter 9), three for gap 2 (2, 3 and 5), two for gap 3 (4 and 6), two for gap 4 (11 and 12), three for gap 5 (7, 9 and 10) and one for gap 6 (13). Up is the maturity of the method available, the editor's reading of the sources cited in each item and in the Open challenge box. Notice that the gaps met by the most chapters, models that follow the operating point and models of events that cannot be tested, have methods in early use, while judging a test while it runs, choosing a structure by its physics and monitoring correlated data are still research.

mean and spread were known [16]. A benchmark of fault diagnosis on the same simulated process as Chapter 13 notes that learned models assume training and test data from one distribution, which is seldom verified in practice, and that a model trained in one operating mode may not generalize to another [18]. The gap is hard because correlated samples, estimated limits and a change of mode all move the false-alarm rate away from its design value at once. Detection that accounts for autocorrelation and links each statistic to a balance that can name the fault remains open.

OPEN CHALLENGE **A Model That Knows When It No Longer Applies**

Three chapters met gap 2: a tank whose constants follow its level, a valve whose gain follows its stroke and a predictive model undone by a splitter. A roadmap of nonlinear identification notes that the real world is nonlinear and time-varying, and that nonlinear identification needs a more tedious experiment, a more involved model choice and a harder estimation [3]. A survey of industrial practice found no validation method to tell whether a model had deteriorated after commissioning [8]. In the editor's view the gap is hard because routine operation seldom moves a plant far enough to show which constant has changed. Progress would be a model that flags when it leaves its validated range and refits from routine records.

14.5.1 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 [23]. 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 [20]. 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 [24]. 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 [25]. 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 [26]. 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 balance gave each model its form, and a test gave or checked its numbers.
- A fitted constant is true of the plant only together with the method and the window that produced it.
- Where a test is unsafe or impossible the balance carries the analysis, and it is only as good as its weakest correlation.
- More parameters fit better and predict worse, so a second record decides the structure.
- A statistic says that a record departs from normal data, and a model of the plant says where and why.

References

[1] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: A survey," *Automatica*, vol. 50, no. 3, pp. 657–682, 2014.

[2] L. Ljung, C. Andersson, K. Tiels, and T. B. Schön, "Deep learning and system identification," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 1175–1181, 2020.

[3] J. Schoukens and L. Ljung, "Nonlinear system identification: A user-oriented road map," *IEEE Control Systems Magazine*, vol. 39, no. 6, pp. 28–99, 2019.

[4] M. von Stosch, R. Oliveira, J. Peres, and S. Feyo de Azevedo, "Hybrid semi-parametric modeling in process systems engineering: Past, present and future," *Computers & Chemical Engineering*, vol. 60, pp. 86–101, 2014.

[5] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021.

[6] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 113, no. 15, pp. 3932–3937, 2016.

[7] P. Shojaee, K. Meidani, S. Gupta, A. B. Farimani, and C. K. Reddy, "LLM-SR: Scientific equation discovery via programming with large language models," arXiv:2404.18400, International Conference on Learning Representations (ICLR), 2025.

[8] 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.

[9] A. Rasheed, O. San, and T. Kvamsdal, "Digital twin: Values, challenges and enablers from a modeling perspective," *IEEE Access*, vol. 8, pp. 21 980–22 012, 2020.

[10] S. J. Qin, "Survey on data-driven industrial process monitoring and diagnosis," *Annual Reviews in Control*, vol. 36, no. 2, pp. 220–234, 2012.

[11] Y. Kono and M. Koizumi, "Model life extension for continuous process: non-invasive correction of model-plant mismatch with regularization," in *2023 European Control Conference (ECC)*, 2023, pp. 1–8.

[12] N. T. Nguyen and J. C. Tique, "Physics-constrained Taylor neural networks for learning and control of dynamical systems," arXiv preprint 2410.02258, 2024.

[13] International Association for the Properties of Water and Steam, "FAQs about IAPWS," Web page, https://iapws.org/faqs/faq2, 2026, accessed 21 November 2026.

[14] ——, "Guideline on the fast calculation of steam and water properties with the spline-based table look-up method (SBTL)," IAPWS, Stockholm, Sweden, Guideline G13-15, 2015. [Online]. Available: https://iapws.org/technical-guidance/release/SBTL

[15] OECD Nuclear Energy Agency, Committee on the Safety of Nuclear Installations, "Status report on spent fuel pools under loss-of-cooling and loss-of-coolant accident conditions: Final report," OECD Nuclear Energy Agency, Paris, Report NEA/CSNI/R(2015)2, 2015. [Online]. Available: https://www.oecd-nea.org/upload/docs/application/pdf/2021-02/csni-r2015-2.pdf

[16] W. F. Guthrie, "NIST/SEMATECH e-Handbook of Statistical Methods (NIST Handbook 151)," National Institute of Standards and Technology, 2020, section 6.3.2, What Are Variables Control Charts? [Online]. Available: https://www.itl.nist.gov/div898/handbook/

[17] J. H. Eto, J. Undrill, P. Mackin, R. Daschmans, B. Williams, B. Haney, R. Hunt, J. Ellis, H. Illian, C. Martinez, M. O'Malley, K. Coughlin, and K. H. LaCommare, "Use of frequency response metrics to assess the planning and operating requirements for reliable integration of variable renewable generation," Lawrence Berkeley National Laboratory, Berkeley, CA, Report LBNL-4142E, Dec. 2010.

[18] E. Fernandes Montesuma, M. Mulas, F. Ngolè Mboula, F. Corona, and A. Souloumiac, "Benchmarking domain adaptation for chemical processes on the Tennessee Eastman process," arXiv preprint 2308.11247v2, 2024, published in Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Communications in Computer and Information Science, Springer, 2026, pp. 307–322, doi:10.1007/978-3-032-25305-7_21. [Online]. Available: https://arxiv.org/abs/2308.11247

[19] 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.

[20] 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.

[21] A. M. H. Basher and J. March-Leuba, "Development of a robust model-based water level controller for U-tube steam generator," Oak Ridge National Laboratory, Oak Ridge, TN, Tech. Rep. ORNL/TM-2001/166, Sep. 2001.

[22] National Aeronautics and Space Administration, "Standard for models and simulations," Office of the NASA Chief Engineer, Washington, DC, NASA Technical Standard NASA-STD-7009B, 2024, approved 2024-03-05, standards.nasa.gov.

[23] 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.

[24] 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.

[25] 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.

[26] L. Bainbridge, "Ironies of automation," *Automatica*, vol. 19, no. 6, pp. 775–779, 1983.

14.6 Exercises

14.1 Discussion · 8 min

An agent hands you a first-order-plus-dead-time model of a new heated tank, fitted by the two-point method to one heater step: gain, time constant and a dead time of about one tenth of the time constant. The tank has no transport delay and its thermometer sits in a thermowell. Using Chapter 3 and Chapter 4, list what you would ask before the model is used to tune a loop, and say what each question could reveal.

14.2 Discussion · 8 min

An agent fits the radiative form of Chapter 11 to one step test, reports a 5.4 % lower objective than the convective form and recommends the radiative model "because it contains more physics". Write the reply. Say what evidence the recommendation lacks, what the chapter measured, and which experiment would settle the question on a physical board.

14.3 Discussion · 8 min

The evaporation defect of Section 14.4.1 made the rate 13.9 % low at the pool's design temperature and 3.3 times high at the 95 °C cap. Explain why tests at the design point would not have caught it, why the error mattered most where it was largest, and propose a routine check for any empirical correlation an agent codes into a model.

14.4 Discussion · 7 min

A utility asks for a model of a hydroelectric unit to tune its governor, and no field step test on the gate is allowed. Using Chapter 7 and Section 14.2, describe how the model should be built, which results must be checked numerically, and what the derivation in the chapter caught that a plan written from memory got wrong.

14.5 Discussion · 7 min

The app of Chapter 4 runs a twin beside the board it models, while the twin of Chapter 5 has not yet met its rig. An agent proposes to commission a new rig's controller on its twin alone. Discuss what evidence the twin needs first, what record should be kept with it, and which defects of Section 14.4.1 show why the record matters.

14.6 Discussion · 7 min

An agent deploys a principal component monitor on a unit with the limit set for a 1 % false-alarm rate per sample. In the first week it raises more than one alarm episode a day on normal operation. Using Chapter 13, explain the likely cause, what the agent should measure before changing anything, and how a balance could make the monitor more useful to the operators.

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 $\bar{F} = 1.0\,\mathrm{m}^3/\mathrm{min}$, $K_L = 8\,\mathrm{m}$ per $\mathrm{m}^3/\mathrm{min}$, $\tau_L = 16\,\mathrm{min}$.

1.3 (b). The outflow no longer depends on the level, so the tank integrates the imbalance of $0.02\,\mathrm{m}^3/\mathrm{min}$ over its area of $1\,\mathrm{m}^2$.

1.5 (a) $\tau_n = 4\,\mathrm{min}$, $\zeta = 1.25$, no overshoot. (b) $K_p = 3$, $\tau_p = 9\,\mathrm{min}$, $\theta_p = 1\,\mathrm{min}$.

1.7 (d). The model equals $-2 + 3/(1+s)$: an instant path of gain -2 and a lag of gain $+3$.

1.9 $K_p = 71/350 = 0.203$. Residuals $+0.086$, -0.129 and $+0.057$.

Chapter 2. Gravity-Drained Tank: A Moving Time Constant

2.1 (a). The initial slope is $K_u/\tau = c/(\rho A)$, which holds neither the level nor the outlet position.

2.3 (c). With the outlet shut the valve can only add liquid, so every low noise sample ratchets the level up until such samples become rare.

2.5 (a) 19.79 s, equal to τ at 12 m. (b) 2.65, 5.80 and 9.90 s.

2.7 (a) 0.050 m/s. (b) 8.62 m, as the linear model predicts. (c) 3.01 m, while the linear model predicts 0.96 m.

2.9 (a) 6.32 m s, $\hat{d} = -8.82\,\%$. (b) 48.42 m s, 10.30 m, 0.90 m short of the setpoint. (c) 8.64 m s. The estimate carries the mismatch as an inlet offset that the plan cancels.

2.11 The table's upset changes the discharge path, while a surge drum receives inlet flow upsets that a tight loop passes downstream in full. The app has no downstream unit and no inlet flow disturbance, so it cannot show the trade.

Chapter 3. Two-Tank Level Control With a Splitter Valve

3.1 (d). The splitter transfer function of Equation (3.12) has a bare s in its numerator, so its steady gain is zero and its step response is a transient.

3.3 (b). The fitted gain is a secant across a large step on a curve whose slope, $2c_1^2\bar{u}/(100\,c_2)^2$, rises with the pump.

3.5 (a) 0.2089 m. (b) 47.86 %, with TK-1 at 0.916 m, 92 % of its rim. (c) 0.0117 m/s, then 0.250 m.

3.7 (a) 0.823 m and 0.0323 m/%. (b) Shut 42.3 and 90.0 s, open 13.7 and 42.3 s. (c) Shut $\hat{\tau}_p = 71.6\,\mathrm{s}$, $\hat{\theta}_p = 18.5\,\mathrm{s}$, open 42.9 s and $-0.6\,\mathrm{s}$.

3.9 (a) $e = 0.0223\,\mathrm{m}$, pump 42.06 %. (b) 0.4777 m and 42.06 %. (c) 0.3104 m, and 88 % of the fall is recovered.

3.11 (a) The steady gain holds, the time constant and the apparent dead time do not. (b) The bypass removes the ceiling on load rejection and a lag, and it removes the buffer that absorbed upstream variation.

Chapter 4. A Two-Heater Digital Twin Beside Its Device

4.1 (c). The thermistor lag $\tau_S = 140\,\mathrm{s}$ is the slowest pole, and the faster heater dynamics appear as the apparent dead time.

4.3 (a). With SP constant, $e_k - e_{k-1} = -(y_k - y_{k-1})$ and the same holds for second differences, so Equations (4.8) and (4.10) collapse onto one expression.

4.5 (a) 20 s and 14.3 s. (b) 141 and 18 s direct, 142.5 and 34.5 s across, ratio 0.13. The dead time is the heater lags, and for the cross response also the conduction path, folded into a delay.

4.7 (a) 50.96 °C and 25.99 °C, gains 0.599 and 0.100 °C/%. (b) $t_{28} = 65$ s, $t_{63} = 159$ s, $\tau_p = 141$ s, $\theta_p = 18$ s. Heat reaches heater 2 only by conduction through the board and leaves it to the room five times faster.

4.9 (a) 55.08 °C, 0.8 %. (b) 0.20 °C, against 1.68 °C under PID. (c) About one second, against about 43 under PID. The plan's model carries the conduction from heater 1, and that model is the twin itself.

4.11 (a) Half of all readings at the cutoff. At 1 °C below it, a $5\sigma_v$ margin, about 3×10^{-7} of them. (b) So the alarm cannot be dismissed while the board is still hot. The latch protects a transistor, and the twin has none.

Chapter 5. Air-Levitated Ball in a Tube

5.1 (b). Nothing on the right of Equation (5.1) depends on z, so the height integrates the mismatch between the air speed and the hover speed.

5.3 (d). The servo takes a new angle only when the command moves more than δ (Equation (5.4)), and at half the gain that takes twice the error.

5.5 (a) 295.1 Pa, 209.7 Pa and 0.29 Pa. (b) 78.0 Pa (26 %), then 287.7 Pa (97.5 %). (c) $a = 0.0950$ at 23.96°, then $a = 0.0495$ at 17.57°.

5.7 (a) 592 mm/s, $u_{eq} = 0.744\,\mathrm{m/s}$, $C_{eff} = 24.6$. (b) 14.3 to 14.5 and 19.6 to 19.8 mm/s per degree. (c) 17.0 against 16.88. Each step measures a chord of the valve curve on its own side of hover.

5.9 (a) 76 and 72 mm s, 1578 and 361°. (b) About 1880 mm s, 6 contacts with the retainer, 1103°. (c) 174 mm s and 89.5°.

5.11 (a) No: 0.48 m/s is 19 % below 0.592 m/s. $u_{eq} = 0.631\,\mathrm{m/s}$, $C_{eff} = 34.2$. (b) $C_{eff} = 41.0$. The drop test identifies u_{eq}, which fixes only m/C_{eff}.

Chapter 6. Flash Steam Level Control

6.1 (b). The vessel loses water from the moment of the step. The rise is vapor below the surface, released by bulk boiling as the pressure falls.

6.3 (d). The limits never bind, so the gain comes from prediction.

6.5 (a) $x = 4.30$ %, 1.75 kg/s. (b) 1.75 kg/s, 0.643 kg/(s bar). (c) 26.5 s. (d) 7.50 bar.

6.7 (a) 55.00, 51.89, 3.11 at 90 s. 58.03, 51.78, 6.30 at 100 s. (b) 2.9 kg lost, against 74 kg implied. (c) 53.84, 50.65 %, 7.52 bar, 1.77 kg/s. The steam still exceeds the fixed 1.75 kg/s of feedwater.

6.9 (a) −1.37 % mean, 1.48 % largest. (b) 0.19 %, with travel rising from 46 to 205 %. (c) After a setpoint change the vessel balances at the old valve position, after the blowdown it needs a new one.

6.11 The valve command against the stem position, and the feedwater flow, which hunts. The stem travels 269 instead of 62 %, 4.3 times as much wear.

Chapter 7. Hydroelectric Governor and Penstock

7.1 (a). The flow cannot jump, so a larger gate area lowers the head $h = (q/G)^2$ until the column has accelerated.

7.3 (c). Holding the power makes the flow rise as the head falls, which is negative damping on the tank's slow mode.

7.5 (a) 287.5 m^2, 314.2 s. (b) 128.1 s, 147 m^2, margin 1.95. (c) 153.9 s, and both criteria call the tank unstable.

7.7 (a) 75.62 MW, 0.7 s after the move, a dip of 4.38 MW. (b) 1.4 s and 11.0 s. (c) 62.30 Hz and 62.00 Hz, both below the 62.33 Hz steady state.

7.9 (a) 1.79 pu and 1.37 pu: the head limit is broken. (b) 1.28 pu and 1.60 pu: the speed limit is broken. (c) 1.29 pu and 1.44 pu: both limits are met.

7.11 (a) $(0.2/60)/0.05 = 6.67$ % of stroke, about 4.9 MW by the slope, against the measured 4.65 MW. (b) The power then carries the droop's share of 100 MW, reaching 6.78 MW at 900 s after an overshoot to 7.29 MW, slowly and with a surge tank wobble.

Chapter 8. Grid Frequency: Droop Offset and Restoration

8.1 (d). Droop is proportional action that belongs to the plant, and AGC supplies the only integral action.

8.3 (b). At nominal frequency both Δf terms vanish, so the command must equal the imbalance.

8.5 (a) $a_3 = 0.2H$, $a_2 = 1.4H + 0.1D$, $a_1 = 2H + 0.7D$, $a_0 = D + 1/R$, and 75.97 against 21. (b) 3.21 s. (c) Both lie just below the stability limit for their droop.

8.7 (a) –0.299 Hz/s, 59.779 Hz, 0.143 Hz. (b) –0.299 Hz/s, 59.644 Hz, 0.321 Hz, IAE 8.87 Hz s. (c) The rate did not move. The offset grew 2.25 times, the ratio of the stiffnesses.

8.9 (a) 59.144 Hz, 0.0 %, 0.000 Hz, 0.200 pu. (b) 5.3 %, 0.290 pu, 60.173 Hz. (c) Once the ramp limit binds, more gain buys no power and winds up the demand, which overshoots.

8.11 (a) 55.6 % for the 4 % unit, 44.4 % for the 5 % unit. (b) The division is indeterminate.

Chapter 9. Spent-Fuel Pool Inventories and Time to Limit

9.1 (c). The train duty is $nk(T - T_{cw})$, which rises by 423.0 kW/K of pool temperature with two trains at 44.20 % and is zero when no train runs.

9.3 (a). The second term of Equation (9.4) raises the surface 18.1 mm as the pool warms 5.68 °C.

9.5 (a) $K_I = 0.0129\,°\mathrm{C/h}$ per percent, 26.8 %/°C and 3.19 h. (b) 36.0 %/°C and 8.64 h, 1.34 times the first-order gain. (c) 456.5 %/°C, 0.32 h, 0.08 h. The gain cannot be entered, since the station stops at 200 %/°C.

9.7 (a) –0.590 °C, –0.0551 °C per percent. (b) 3.4 h and 0.2 h, so $\theta \approx 0.2$ h, $\tau \approx 3.2$ h. (c) 1.429 °C/h. Evaporation grows on the way up.

9.9 (a) 38.5 m^3, and 6.830 m, 170 mm low at 132 h. (b) 0.8 mm, 46.8 m^3. (c) 0.21 m^3/h, 1.9 mm/h. The level must fall before the loop knows anything.

9.11 (a) $4.63 - 3.82 + 3.57 = 4.38\,\mathrm{m}^3$, about 0.050 m^3/h over 88 h. (b) The mass balance. It needs a makeup meter and the evaporation computed from the pool temperature and the building air.

Chapter 10. Oil Price Versus Market Share in a Quota Loop

10.1 (c). The price held above the reference moves the slow states of the fringe and the consumers, and the PI keeps the price only by making room for them.

10.3 (a). A stock target is a price target filtered by a sixty-month moving average of the stocks themselves, so it has no anchor at the normal stocks.

10.5 (a) \$12.5 per Mb/d. (b) 0.745 Mb/d per dollar, \$1.34 per Mb/d, 9.3. (c) The slowest mode has a time constant of about seven years, so after seven years about a third of its part of the recovery is still to come.

10.7 (a) \$66.88, \$71.99, \$73.43, lowest \$66.87 at 0.27 years. (b) \$83.76 and \$76.40. (c) \$8.12 and \$1.58 per Mb/d for the rise, \$8.76 and \$1.40 for the cut. The price law is convex, so equal and opposite changes of stock move the price more upward than downward.

10.9 (a) 6.12 (the meeting of year 6.10), 8.12, and \$61.06 at year 9.25. (b) 5.10. (c) Each cheater is a step disturbance at the plant input, which the integral cancels by cutting the quota until the floor binds. A slow integral cannot cut fast enough to keep up with a new cheater every year.

10.11 (a) Year 4.33, \$86.3 and \$86.1, \$43 billion. (b) Give the inventory one controller, or separate the two by a split range or a gap between their setpoints. A reserve ceiling of \$95, above anything the cartel holds, keeps the reserve out of the fight.

Chapter 11. Model Selection on a Temperature Control Lab

11.1 (b). Equation (11.6) divided by UA depends only on $\tau = mc_p/(UA)$ and $K_p = \alpha_1/(UA)$.

11.3 (d). With the generator's lag cut to 20 s the two agree within 0.57 % of the range.

11.5 (a) $UA = 0.010\,152\,\mathrm{W/K}$, $\tau = 197\,\mathrm{s}$, $K_p = 0.676\,°\mathrm{C/\%}$, about 13 % high (the chapter's 12.8%). (b) $\Delta\alpha_1 = UA\,\Delta T_a/Q_1 = 0.0012\,\mathrm{W/\%}$, about 18 %. A second heater level gives a second relation.

11.7 (a) 2767.9 °C, then 197 s, 0.676 °C/% and 141.9 °C. (b) 1876.4 °C, 167 s, 0.833 °C/% both times. (c) $U = 16.91$, $\alpha_1 = 0.0069$, and the same 197 s, 0.676 °C/% and 141.9 °C as (a).

11.9 (a) 111.9 against 197.1 s, a gap of 14.4 %. (b) 31.9 against 35.3 s, 0.57 %. Shortening the missing sensor lag removes most of the gap, so the gap is the form's missing state, taken up by drifting parameters.

11.11 (a) 0.886: the loop is about 11 % less aggressive than designed. (b) Toward 15 °C, below any plausible room, because the forms have no sensor state. Seed T_a from the first reading.

Chapter 12. Multivariable Identification With Exact Zeros

12.1 (a). Persistence has no model and still scores within 0.003 of the best model one step ahead. Only the free run tests what the model predicts from the inputs alone.

12.3 (c). A fitted gain is a property of the model structure as well as of the plant.

12.5 (a) 0.1, 1 and 28.3; the lags differ by a factor of about 3200, close to $e^8 = 2981$. (b) 0.0995 s; overshoot 62.1 against 53.4 %, period 0.632 against 0.641 s.

12.7 (a) 0.512, 104 s, 10.9 s; 0.0789, 181 s, 52.6 s. (b) 0.0232, 0.0299, 0.0165. (c) 0.0853, a factor of 3.7; 0.154 and 0.0165; 0.542 and 4.79 s. Heat from heater 2 that reached sensor 1 late is now credited to heater 1, which moves together with it, so the direct path rises earlier and higher.

12.9 (a) PI 347 and 0.344 °C; both MPC forms 392 and 0.137 °C. (b) 217, 372 and 132 °C s. (c) 342 and 0.0610 at 0.1; 598 and 0.133 at 10. Move suppression prices the moves that the PI makes freely, and the DMC's output bias assumes the present error persists while the load is still rising through a 104 s lag.

12.11 Under feedback the valve moves in response to the disturbance, so every estimator here can be biased toward the inverse of the controller. Add external excitation such as setpoint steps, fit with ARX refinement off, and require the families to agree.

Chapter 13. Process Monitoring Workbench

13.1 (d). Reconciliation distributes one balance's misclosure over every stream in that balance, so one gross error moves every adjustment.

13.3 (b). The limits assume independent samples from one stationary normal operation, which a closed-loop record is not.

13.5 (a) $2.962/3 = 0.987$. (b) 22.662 m^3/h. $C^- = 5.21 > 5$ and $|w| = 1.132 > 0.987$, so both signal on the same sample. (c) $p = 0.0027$, 370.4 samples or 18.5 h, and the EWMA averages 25.0 h between false signals.

13.7 (a) 313.67 against 5.99, then 17.71, 13.20 and 11.68. (b) Station flowmeter, gross error 2.35 L/s, $\gamma = 0.13$ against 3.84. (c) 177.21 against 3.84 with one reduced balance. The pump curves are removed and reported tied with the station flowmeter, because the two are the only measurements of one flow and no test can tell them apart.

13.9 (a) 494 annunciations, all 6 periods in flood, 5 chattering tags carrying 473. (b) 14, no flood, no chattering: below a tenth of 494 (49.4) with no flood, so the target is met. (c) 16 with one flood of 11 in the period from 30 to 40 min, so the target is missed.

13.11 (a) 0.2944 kappa/h, within 2 % of the declared 0.3. (b) 30 min apart, and $0.3 \times 30/60 = 0.15$ kappa of extra bias from the drift.

Chapter 14. Summary and Outlook

14.1 A good answer asks where the dead time comes from, since the plant has no transport delay: lags in series (a second tank, a sensor) read as dead time, as in the two-tank fits (18.5 to –0.6 s with no delay at all). It asks whether the time constant belongs to the tank or to the thermowell, as the 141 s fit of Chapter 4 belonged to the 140 s thermistor. It asks for the step size against the operating range, the window and whether the step settled, for a second step in the other direction or at another operating point, and for a balance that predicts the gain and time constant so the fit can be checked against it.

14.3 The design point is where a model is calibrated and where its tests cluster, and a 13.9 % error in a term that carries about one percent of the heat load moves little there. The wrong density mattered more as the vapor pressure grew, and the loss-of-cooling case runs at the cap, so the error set the latent absorption and both clocks. Boiling moved from 16.88 to 14.74 h once fixed. A routine check evaluates the coded correlation against the source's own table or worked example across the whole range the model will visit, with units and definitions read from the source, and records the comparison beside the code.

14.5 The twin needs a comparison with the hardware on the tests that set its constants: the drop test, the hover map and step tests at several operating points for the ball, with residuals and a date. The record keeps the parameter table with the source of each value (nominal, measured, fitted), the comparison runs and what the recorded columns mean. The two-heater export's independent noise draw and the first-order loss column in the wrong unit show that a record can mislead a fit, and the levitation referee's removal of a hardware claim shows that nominal values must not be reported as measured.

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.

actuator saturation The state in which an actuator sits at a limit, so a larger command moves the process no further. *Chapter 1*.

Akaike information criterion A score that adds a charge for each fitted parameter to the fit error, so a more complex model must earn its extra parameters. *Chapter 1*.

alarm flood A burst of alarms arriving faster than operators can read and act on them, as happens during a large upset. *Chapter 13*.

anti-windup Logic that stops a controller's integral term from growing while the actuator is held at a limit. *Chapter 1*.

area control error The signal a grid control area steers to zero, made of its frequency error and its error in scheduled power exchange. *Chapter 8*.

ARX model A sampled model that predicts the next output from past outputs and past inputs, fitted with one linear least-squares solve. *Chapter 1*.

autocorrelation How much each residual resembles the one before it, near zero when the residuals are pure noise. *Chapter 1*.

automatic generation control The central control that slowly moves generator setpoints to bring grid frequency and power exchanges back to schedule. *Chapter 8*.

balance equation A statement that the rate at which a quantity accumulates equals what flows in, minus what flows out, plus what is generated. *Chapter 1*.

bumpless transfer A switch between manual and automatic control that does not jolt the actuator. *Chapter 1*.

cascade control Two loops in series in which an outer controller sets the setpoint of an inner, faster loop. *Chapter 6*.

closed-loop identification Fitting a model to data recorded while a controller was acting, so that the input and the noise move together. *Chapter 12*.

control chart A plot of a measured statistic against time with limits, usually three standard deviations from the mean, that flag an unusual change. *Chapter 13*.

control layers The levels of a plant's control system, from field devices and PID loops up to predictive control, optimization and people. *Chapter 1*.

control valve A valve moved by an actuator so that a controller can set the flow through it. *Chapter 2*.

damping factor A number that says how quickly a second-order response settles, with values below one giving overshoot and ringing. *Chapter 1*.

data historian A computer that collects plant readings at regular intervals and keeps them, typically for years, for later analysis. *Chapter 1*.

data reconciliation Adjusting measured flows by the least amount that makes the balances close, which also exposes a meter in error. *Chapter 13*.

dead band A range of input change over which a device does not respond at all. *Chapter 5*.

dead time An interval after a change during which the output shows nothing of it, as when fluid travels down a pipe. *Chapter 1*.

decay heat Heat that spent nuclear fuel keeps producing after the reactor stops, falling slowly over months and years. *Chapter 9*.

deviation variable A variable measured from its value at the operating point, so that the steady state becomes zero. *Chapter 1*.

digital twin A model run beside its physical device on the same inputs, so that the two can be compared as they run. *Chapter 4*.

discrete-time model A model that steps from one sample to the next, written as a difference equation rather than a differential equation. *Chapter 1*.

droop A governor setting that makes each generator raise its output in proportion to a fall in frequency, which leaves a steady offset. *Chapter 7*.

eigenvalue A number of a linear model's state matrix that gives one mode of its response, equal to one pole of its transfer function. *Chapter 1*.

fault detection Spotting from data that a sensor, valve or process has stopped behaving as it does in normal operation. *Chapter 13*.

feedforward Control action computed from a measured disturbance before that disturbance has moved the controlled variable. *Chapter 5*.

final control element The device that a controller moves to change the process, such as a valve, a pump, a heater or a gate. *Chapter 1*.

finite impulse response model A model that writes the output as a weighted sum of past inputs, with

one weight for each sample of delay. *Chapter 12*.

first-order response The response of a single lag to a step, which starts at its steepest and covers 63.2 percent of its change in one time constant. *Chapter 1*.

first-order-plus-dead-time model A model with a gain, one time constant and a dead time, abbreviated FOPDT, the form that most tuning rules need. *Chapter 1*.

flash vessel A vessel in which hot water drops in pressure and part of it boils into steam. *Chapter 6*.

frequency nadir The lowest frequency a grid reaches after it loses a generator, before the governors stop the fall. *Chapter 8*.

gain scheduling Changing a controller's settings with the operating point, so the loop behaves alike where the plant's gain differs. *Chapter 3*.

identifiability Whether the data can fix a model parameter on its own, rather than only a group of parameters together. *Chapter 1*.

inertia constant The kinetic energy stored in the rotating machines at rated speed divided by their rating, expressed in seconds. *Chapter 8*.

integrating process A process with no self-regulation, whose output keeps rising or falling as long as its flows are out of balance. *Chapter 1*.

integrator windup Growth of a controller's integral term while the actuator sits at a limit, which later causes a large overshoot. *Chapter 3*.

inverse response A response that first moves the wrong way before it settles in the right direction. *Chapter 1*.

Kalman filter An estimator that blends a model's prediction with noisy measurements, weighting each by its expected error. *Chapter 2*.

least squares Choosing a model's parameters to make the sum of squared differences between the data and the model as small as possible. *Chapter 1*.

Levenberg–Marquardt method An iterative search for least-squares parameters that blends small cautious steps with larger steps aimed at the minimum. *Chapter 12*.

linearization Replacing a nonlinear model by its straight-line approximation near an operating point. *Chapter 1*.

load damping The fall of electrical load as frequency falls, which by itself slows a decline of frequency. *Chapter 8*.

loop interaction The effect of one control loop's actions on the controlled variable of another loop. *Chapter 4*.

mass-equivalent level The level a vessel's liquid would show at a fixed reference density, so it follows the mass of liquid and ignores bubbles. *Chapter 6*.

measurement noise Random scatter in a sensor's reading that carries no information about the process. *Chapter 2*.

model discrimination Choosing among rival model structures by how well each predicts data it was not fitted to. *Chapter 11*.

model predictive control Control that uses a model to predict the plant over a horizon and plans moves that respect limits, abbreviated MPC. *Chapter 2*.

model structure The form of a model, such as its order, dead time and terms, as distinct from the numbers inside it. *Chapter 1*.

model validation Testing a fitted model on data it was not fitted to, and checking that its residuals look like noise. *Chapter 1*.

moving horizon estimation Estimating a process state by fitting a model to a window of the latest measurements, repeated at each sample. *Chapter 2*.

natural period The time scale of a second-order model, which together with the damping factor sets how fast it responds and rings. *Chapter 1*.

non-interacting tanks Tanks in series in which the outflow of each tank does not depend on the level of the tank below it. *Chapter 1*.

offset A steady difference between setpoint and measurement that a controller without integral action leaves behind. *Chapter 3*.

OPEC An organization of oil-exporting governments whose meetings set production targets for its members. *Chapter 10*.

operating point The steady state about which a model is linearized and at which its gain and time constant hold. *Chapter 1*.

orifice A restriction through which the flow grows with the square root of the head or pressure drop across it. *Chapter 1*.

overfitting A fit that follows the noise of its own data and so predicts new data worse than a simpler model would. *Chapter 11*.

penstock The pipe that carries water from a reservoir or surge tank down to a hydroelectric turbine. *Chapter 7*.

per-unit system Expressing quantities as fractions of chosen base values, so that machines of any size share one scale. *Chapter 8*.

PID control Feedback control that acts on the present error, on its accumulated sum and on its rate of change. *Chapter 1*.

plant test A planned set of input moves on a running plant, recorded so that a model can be fitted to it. *Chapter 12*.

primary frequency control The fast, automatic response of generator governors to a change of grid frequency. *Chapter 8*.

principal component analysis A method that finds the few combinations of many correlated measurements that carry most of their normal variation. *Chapter 13*.

process gain How far the output finally moves for each unit change of the input. *Chapter 1*.

process model Equations that describe how the outputs of a process respond to its inputs and disturbances. *Chapter 1*.

rate of change of frequency How fast grid frequency moves in the first moments after a sudden imbalance between generation and load. *Chapter 8*.

relative gain array A table of ratios that shows how strongly the loops of a multivariable plant would interact under control. *Chapter 4*.

residual The difference between a measurement and the model's prediction of it. *Chapter 1*.

right-half-plane zero A zero of a transfer function with a positive real part, the mark of an inverse response and a limit on control speed. *Chapter 1*.

sample period The time between two successive readings or two executions of a controller. *Chapter 1*.

secondary frequency control The slower control that returns grid frequency to its setpoint after primary control has stopped its fall. *Chapter 8*.

second-order response The response of a model with two lags, or with two stores that trade their contents, which may overshoot and ring. *Chapter 1*.

self-regulating process A process whose own outflow or losses grow with its output until it settles at a new steady state. *Chapter 1*.

spent-fuel pool A deep pool of water that cools and shields used nuclear fuel after it leaves the reactor. *Chapter 9*.

state-space model A model written as first-order differential equations in a set of state variables, with inputs and outputs. *Chapter 1*.

statistical process monitoring Watching plant data with statistics learned from normal operation, to flag when operation is no longer normal. *Chapter 13*.

steady state A condition in which no variable changes with time, so every derivative in the balances is zero. *Chapter 1*.

steam tables Tables or equations that give the properties of water and steam, such as density and enthalpy, at each pressure and temperature. *Chapter 6*.

step test A test that holds a plant steady, steps one input and records the outputs until they settle. *Chapter 1*.

stiction Static friction that makes a valve stem stick and then jump, so small changes of signal do not move it. *Chapter 2*.

subspace identification Fitting a state-space model directly from input and output data with matrix methods rather than an iterative search. *Chapter 12*.

swell and shrink The rise of a boiling vessel's level when steam demand rises, and its fall when demand falls, against the change in its water. *Chapter 6*.

swing equation The momentum balance on a grid's rotating machines, which ties the rate of change of frequency to the power imbalance. *Chapter 8*.

system identification Building a model of a dynamic system from measured input and output data. *Chapter 1*.

Temperature Control Lab A small board with two transistor heaters and two temperature sensors used to practice modeling and control. *Chapter 4*.

Tennessee Eastman process A simulated chemical plant widely used as a shared benchmark for control and monitoring methods. *Chapter 13*.

thermistor A temperature sensor whose electrical resistance changes strongly with temperature. *Chapter 4*.

thermowell A closed metal tube that holds a temperature sensor in a process stream and adds a lag to its reading. *Chapter 1*.

time constant The time a first-order response takes to cover 63.2 percent of its final change. *Chapter 1*.

time to limit The time a pool takes to reach a temperature or level limit if nothing changes, computed by marching its balances. *Chapter 9*.

transfer function The ratio of a linear model's output to its input in Laplace terms, which holds its gain, poles, zeros and delay. *Chapter 1*.

transient droop A governor feature that adds a temporary droop after a change and lets it decay, to allow for the inertia of the water column. *Chapter 7*.

two-point method Reading a time constant and a dead time from the times at which a step response covers 28.3 and 63.2 percent of its change. *Chapter 1*.

velocity form A digital PID that computes only the change of its output at each execution. *Chapter 1*.

water starting time The time the head takes to accelerate the water in a penstock to its rated flow, a measure of the column's inertia. *Chapter 7*.

Index of Control Topics

actuator saturation, 248
AI agents
  engineering practice, 344
  open problems, 348
  reproduction, 346
  truth audit, 346
  verification record, 345
  while studying, 18
air levitation
  history, 114
alarm management, 317, 318
  alarm rate, 332
  chattering alarm, 323
analyzer drift, 320
ARX model, 9
  multi-input, 289
automatic generation control, 182
  history, 184
averaging level control, 26, 218

back-calculation, 34, 64, 92, 122, 192, 217
balance
  model form, 338
balance equations, 4, 87, 140, 237
  energy, 185, 208
  mass, 208
  momentum, 115, 163
bootstrap, moving-block, 296
bumpless transfer, 16, 33, 64, 92, 122, 146, 191, 244

cascade control, 146
closed-loop identification, 303
closure law
  commissioning tests, 176
code verification, 173
colored noise, 186
competing controllers, 244, 249
conservation laws, 4
constraint back-off, 249
control charts, 317
  average run length, 322
control layers, 14
control loop performance, 317
control valve
  up-and-down test, 46

damping factor, 6
data export, 17
data reconciliation, 317
dead band, 117, 216
dead time, 7
  apparent, 55, 62, 91, 216, 338
  fractional, 292
  lumped lags, 120, 216
deviation variables, 5, 289
digital twin, 84, 278, 343
  extrapolation, 279
  history, 260
  learning-based control, 107
  teaching laboratory, 130
discrete-time model, 9
drag law, 115
droop, 169, 182
  load sharing, 183
drop test, 122
dynamic matrix control, 297

eigenvalues, 167
evaporation, 210
  correlation, 345

fault detection, 315, 332
feedforward, 146
  from a computed disturbance, 218
  static map, 122
final control element, 12
finite impulse response model, 290
first-order response, 6
first-order-plus-dead-time model, 7, 91, 215
forecast
  preview in MPC, 150
free-run simulation, 285
frequency bias, 169
frequency measurement, 196

gain scheduling, 55, 66
GEKKO export, 297
gross error detection, 317
  serial elimination, 321

Harris index, 322
hybrid model, 343
hydropower
  flexibility, 176

IAPWS-95
  research needs, 154
IAPWS-IF97
  in practice, 154
identifiability, 11, 258
  parameter groups, 268
  practical, 260
inertia
  low-inertia grids, 198
infeasible balance, 149
input design, 272
integral action
  offset removal, 182
integral control, 248
integrating disturbance, 242
integrating process, 6, 31, 112, 144, 181, 204
  frequency, 182
  identification, 292
  inventory, 208
  with lag and dead time, 120
integrator windup, 64, 126
interacting tanks, 57
inverse response, 8, 160, 166
  swell and shrink, 135, 144
  water column, 170

Kalman filter, 17, 146
  extended, 266
  input disturbance, 35, 65, 298

laboratory kits
  history, 86
leak detection
  mass balance, 228
least squares, 9, 122
level control, 53, 218, 234
  three-element, 135, 137
level measurement
  narrow range, 226
  swell, 136
Levenberg-Marquardt method, 293
limit cycle, 126
  from excess loop gain, 171
linearization, 5, 30, 61, 120, 166, 214, 241
load frequency control, 183
loop interaction, 84
loss of cooling
  operating experience, 226

Markov parameters, 295
measurement
  delayed, 240
measurement lag, 88
measurement noise, 28, 58, 88, 117, 142, 165, 212, 272
mechanical starting time, 164
method of lines, 165
minimum-variance benchmark, 322
model
  from first principles, 340
model building
  procedure, 341
model discrimination, 258
model fitting
  absolute-deviation objective, 261
  Nelder-Mead simplex, 264
  two-point method, 9, 62, 91
model mismatch, 94
model predictive control, 93, 147
  control hierarchy, 308
  economic, 235
  in the apps, 16
  model, 342
  model maintenance, 79, 343
  model mismatch, 34, 74, 150, 245
  move suppression, 298
  multiple-model, 66
  on an identified model, 297
  performance monitoring, 328
  soft constraint, 94
  successive linearization, 244
model structure, 10, 258
  information criterion, 10, 278
  nested models, 263
model validation, 10, 284
  chronological split, 287
  domain of validation, 276
  hold-out, 269, 341
  in service, 348
  leakage, 346
moving horizon estimation, 17, 35, 66, 95, 265

neural network
  physical constraints, 343
non-interacting tanks, 57
normalized root-mean-square error, 290

offset
  proportional control, 71, 125, 148, 181, 189
offset-free control, 65, 93

on/off control, 216
operating point
  gain and time constant, 23
  valve gain, 120
overfitting, 270, 341

parameter correlation, 270
parameter estimation, 95, 264
per-unit system, 184
PID control
  anti-windup, 16, 34, 92, 192
  derivative action, 34, 125, 148
  derivative and noise, 105
  derivative kick, 93
  derivative on measurement, 91
  velocity form, 16, 32, 63, 91, 121, 146, 169, 190, 217, 244
PID tuning
  before tuning, 77
  Hovey rule, 168
  internal model control, 68
  SIMC rule, 299
  SIMC rules, 120, 217
  Tyreus–Luyben rule, 67
  Ziegler–Nichols rules, 217
plant order, 53
plantwide control, 78
pole-zero cancellation, 61
primary frequency control, 185
  in service, 198
principal component analysis, 317
  contribution plot, 322
  T^2 and Q statistics, 322
process gain
  short-run and long-run, 242
process model, 1
  milestones, 2
process monitoring
  residuals, 341
protection layer, 148

quantized actuator, 242

rate limit, 165, 192
regression, 9
relative gain array, 90, 299
relay autotuning, 36, 67, 95
  switching band, 67, 96
residual analysis, 10, 269
  autocorrelation, 10
right-half-plane zero, 8, 160, 166
Routh stability criterion, 189

safety interlock, 96, 267
sample period, 117
  as dead time, 126
sampled-data control, 90, 242
saturation
  of the plant, 248
scenarios, 17
  challenge, 37, 74, 127, 151, 196, 225, 249
  completion rules, 17, 97, 123, 171, 267
scorecard, 17
second-order response, 6, 302
secondary frequency control, 190
self-regulating process, 23, 204, 234
sensor fault
  bias, 150
setpoint ramp, 122
simulation engine, 15
simulator framework, 14
soft sensor, 279
speed measurement
  toothed wheel, 174
spent-fuel pool
  accident research, 228
stability boundary, 167
state estimation
  inventory, 146
state-space model
  identified, 290
statistical process monitoring, 315
steam tables
  history, 138
step test, 9, 90, 215
  history, 3
  on a nonlinear plant, 31, 62, 125
  on an integrating plant, 122
  window, 338
stiction
  detection, 48
structural zero, 284, 292
subspace identification, 295
system frequency response model, 184
system identification, 1, 260
  history, 3, 286, 343
  machine learning, 342
  multivariable, 284
  nonlinear, 309
  open problems, 347

temperature control, 84

time to limit, 204
    marched, 218
    present-rate extrapolation, 218
time-scale separation, 190
transfer function
    zero at the origin, 62
transfer function matrix, 289
transient droop, 168

ultimate gain, 96, 171

valve authority, 112
valve characteristic, 117
    equal-percentage, 212
    installed, 129
valve stiction, 113, 142, 317
variable projection, 292

water starting time, 163

zero-order hold, 292

Index of Applications

air levitation, 112
alarm flood, 323
area control error, 183

ball in a tube, 112
blower, 115
boil-off, 210
breakeven price, 239
bulk boiling, 141

cartel and competitive fringe, 235
cheating on a quota, 238
closure law, 165
collapsed level, 142
component cooling water, 207
composite stiffness, 189
control valve
    positioner, 331
    split range, 153
    stiction, 46, 317
converter-interfaced generation, 184
cooling train, 207

dashpot governor, 162, 169
data historian, 12, 277
decay heat, 210
decision calendar, 242
demand destruction, 239
distillation column
    Wood-Berry model, 300
drum boiler, 137

elastic penstock, 165
emissivity, 263
evaporation
    mass-transfer correlation, 210

flap valve, 115
    butterfly and damper, 128
flash quality, 141
flash tank, 137
flash vessel, 135
float regulator, 25
flowmeter, 319, 330
forced-circulation loop, 138
Francis turbine, 162
frequency nadir, 194
Fukushima Daiichi, 207

gas chromatograph, 306
gate servo, 165
generator trip, 195
geothermal separator, 137
gravity-drained tank, 3, 23, 57
grid frequency, 181

headrace tunnel, 162
heat exchanger, 303
    effectiveness-NTU, 211
    fouling, 211
heat-recovery coil, 138
heat-recovery steam generator, 137
heated tank, 3
hobby servo, 115
hydro governor, 160
hydroelectric power, 160
    history, 162

IAPWS-IF97, 140
inertia constant, 184
instrument, 12
islanded operation, 164
isochronous governor, 183

kraft digester, 320

level control, 23
level transmitter, 12, 56
    guided-wave radar, 45
    magnetostrictive, 77
    wet leg, 152
load damping, 184
load rejection, 169
loss of cooling, 207
lumped capacitance, 263

makeup valve, 212
    in practice, 226
market share, 239
mass-equivalent level, 142
microgrid, 184

natural convection, 263

oil inventories, 237
  five-year range, 251
  statistics, 251
oil market, 234
oil price
  forecasting, 252
OPEC, 234
  founding agreement, 237
orifice discharge, 27
orifice plate, 115
  series restriction, 120
overfill, 47, 58
  history, 25
overspeed, 168

penstock, 162
plant historian, 315
plant test, 284
  practice, 307
positioning stage, 302
power grid, 181
price elasticity, 239
price war, 236, 244
pump
  variable-frequency drive, 77

radiation heat loss, 88, 263
rate of change of frequency, 189
  relay setting, 198
renewable variability, 186
residual demand, 242

shale oil, 237
shielding depth, 207
skimmer weir, 209
spare capacity, 251
spent-fuel pool, 204
  high-density racks, 206
  history, 206
splitter valve, 53
steam tables, 140
  fast calculation, 343
steam turbine, 185
storage limit, 248
strategic petroleum reserve, 239
  origin, 244
submerged vapor, 141
surge tank, 165
  Thoma stability area, 168
surge vessel, 26, 55, 78, 302
swell and shrink, 144
swing equation, 185
synchronization, 169

tanks in series, 53
Temperature Control Lab, 84, 258, 287
Tennessee Eastman process, 317
thermal expansion, 209
thermistor, 88, 261
  calibration, 275
  self-heating, 104
thermowell, 12
three-way valve, 56
tie-line bias control, 183
time-of-flight sensor, 115
  principle, 128
Torricelli outflow, 27, 57
  history, 57
transistor heater, 87, 261
  pulse-width modulation, 105
turbine governor, 182
  valve, 197

under-frequency load shedding, 187

waste heat recovery, 135
water clock, 25
water distribution, 319
water hammer, 162
  Joukowsky relation, 168
Web Serial, 87, 267
wicket gates, 162
  servomotor, 175

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.