INTERACTIVE CONTROL ENGINEERING
Principles and Applications

Volume 9 Model Predictive Control and Economic Optimization

CHAPTER 13

Summary and Outlook

Pages 311–325

Edited by John D. Hedengren

How to cite this chapter. "Summary and Outlook," in *Model Predictive Control and Economic Optimization*, John D. Hedengren, Ed., vol. 9 of *Interactive Control Engineering: Principles and Applications*. Advanced Process Solutions, LLC, 2026, ch. 13, pp. 311–325.

The page numbers in this file are those of the printed volume.

The volume online: contents, other chapters and citation records
https://apmonitor.com/apps/books/v09-mpc-optimization/

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

CHAPTER 13

Summary and Outlook



Open gaps 1 models 2 baselines 3 margins 4 forecasts 5 coordination 6 infeasibility

Notice that the gap of keeping the model right joins four plants, the balloon, the weld, the boiler and the data hall, more than any other gap. Illustration, not to scale.

IN PLAIN TERMS **What Eleven Plans Taught**

The main idea. Each chapter of this book put a plan above a plant: a controller that looks ahead, or an optimizer that picks the best place to run. Each plan was then set against a simpler rule on the same day.

The challenge. The plan won where it knew something the rule did not, such as a forecast, an order or a price. It lost when its picture of the plant was wrong, and it used up every spare margin it was given.

Why it matters. So a plan is worth its cost only after a fair test against the simple rule, tuned once, with every saving taken apart.

Eleven chapters put a model and an optimizer above eleven plants and measured each against the simpler control that Table 1.3 names for it. The results ran from a weld that a step test from the wrong regime left with millimeters of scrap to a reservoir plan that kept the town under its limit through a water year in which the rule curve flooded it. This chapter sets those results side by side, draws a short guide for deciding when a plan is worth its model, looks at where predictive control and economic optimization are heading, reports how the volume itself was made with AI agents and what they got wrong, and lists the problems that remain open.

13.1 What the Chapters Showed

Tables 13.1 and 13.2 complete Table 1.3 with the measurements. The first holds the four chapters that use prediction to regulate, and the second the seven that use it to decide what the regulators hold or which whole decisions to take. Each row names the simpler control, where the predictive or optimizing layer paid, where it did not, and the numbers with the table or section that reports them. Figure 13.1 sets one like pair of numbers from each row side by side.

The columns of the two tables make five points.

The plan paid for what it could see. Every clear win in the tables traces to information a feedback loop does not have. The canal's plan sees an offtake in all three reaches at once. The heating network's plan sees the CHP's spare heat in the evening, the reservoir's tree sees seven futures, and the building's plan sees the curtailment notice. Take the information away and the win goes with it. The aeration plan with its preview off used 8622 kWh over two days at a worst mean of 2.607 g N/m^3, against 8596 kWh at 2.205 g N/m^3 for the supervisor (Table 10.7). The building's plan with its preview off did on the notified day what the supervisor does (Section 9.6.6). This is the closing argument of Section 1.4.13 measured on eleven plants.

Table 13.1 Where Prediction Paid in Regulation, Chapters 2 to 5

Chapter	Simpler control	Where the plan paid	Where it did not	Measured result
2, hot-air balloon	Altitude PID. A three-deep cascade.	A smaller overshoot on a level-off.	The landing, where the cascade now touches down softer. The cruise, where the cascade swung less. A 120 s horizon, and a constant bias on an integrating output.	Touchdown 7.85, 1.27 and 1.39 m/s for PID, cascade and DMC (Table 2.6). Late swing 201.1, 0.8 and 6.8 m (Table 2.5). Hold offset 132.9 with the bias alone against 3.9 m with the drift ramp (Table 2.9).
3, friction stir weld	Two PI loops, ITAE-tuned for setpoint or for load.	DMC built on a traverse step test, gentler on the spindle.	The same code built on a plunge step test. Both DMCs on an unmeasured thin patch.	Sound weld 98.3, 96.9 and 91.6 % for the traverse DMC, the load-tuned PI and the plunge DMC (Table 3.4). Thin patch 315 and 410 against 109 °C s (Section 3.6).
4, drum boiler and turbine	Boiler-following, turbine-following and coordinated PI.	Offset-free MPC on sliding pressure, the only setting to pass the challenge.	The nominal MPC, offset under a fuel it cannot see. The tube, which holds the limit and gives up range.	Challenge passed on 11 of 11 seeds, every other setting on none. Final pressure gap 0.12 against 1.81 kg/cm^2 for coordinated PI (Table 4.5). Nominal offset 1.25 kg/cm^2 (Section 4.6).
5, irrigation canal	Distant downstream PI with order feedforward.	An unannounced offtake, spread over three reaches by one plan. Plans per gate that exchange.	Published orders, which feedforward timed to the wave meets better. Plans per gate that do not exchange.	Farm closure IAE 406 against 1194 (m s), worst 5.84 against 11.46 cm. Published order 396 against 383, 2.88 against 2.15 cm (Table 5.8). Worst after a surprise 2.90, 3.10 and 6.68 cm (Table 5.9).

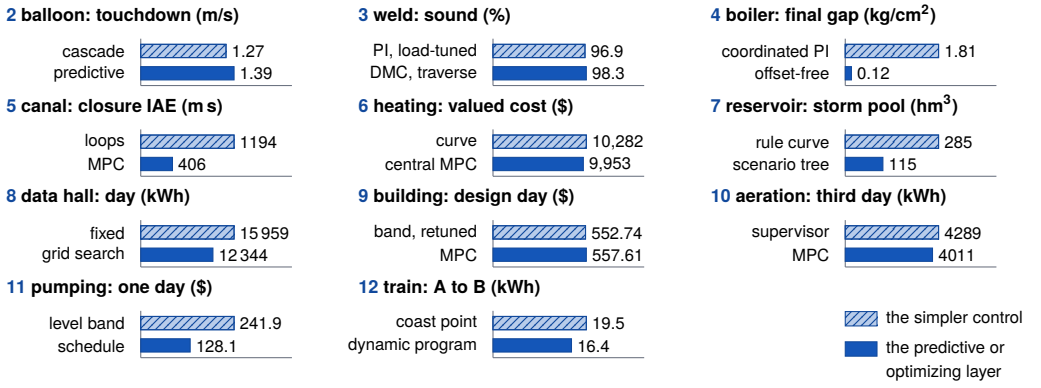


Figure 13.1 One like pair of numbers from each row of Tables 13.1 and 13.2, digit for digit, each panel on its own scale. The hatched bar is the simpler control of the row, the solid bar the predictive or optimizing layer. Higher is better for the weld, lower for the other ten. Notice that the plan lost on the balloon's landing and to the building's band retuned once, and that the largest gaps come where a first optimizer replaced a fixed rule.

Table 13.2 Where Optimization Paid in Targets and Whole Decisions, Chapters 6 to 12

Chapter	Simpler control	Where the plan paid	Where it did not	Measured result
6, district heating	Outdoor-compensated supply curve, a boost by hand.	Charging pipes and tank on the CHP's spare evening heat.	A boost timed by hand. A price that cannot settle when capacity binds.	Valued cost \$10,282 curve, \$10,484 boost, \$9,953 central MPC (Table 6.4). Cold snap passed by the curve on 0 and the distributed MPC on 5 of 7 weathers (Table 6.5).
7, flood reservoir	Rule curve, with pre-release on the forecast.	A scenario tree on the seven-member ensemble.	A season's value table, which neither routes the river nor sees the ramp.	Storm pool 115 against 285 hm^3, 123.7 against 93.6 GWh, SDP town 530 m^3/s (Table 7.4). Water year 7 of 7 against 0 of 7 seeds (Table 7.5).
8, data hall cooling	Fixed setpoints. A grid-search supervisor.	The supervisor against fixed setpoints. Continuous moves over the grid.	A horizon over the supervisor. A lag model identified at one valve position.	Day 15 959 against 12 344 kWh (Table 8.5). MPC \$1409.40, grid \$1415.88, retuned grid \$1411.08, cheapest on none of four cases (Table 8.8).
9, office building	Fixed setpoint. A priced band supervisor.	The band against the fixed setpoint. The plan on a curtailment day and on the challenge's peak.	Against the supervisor retuned once, on the design day and the heat wave. A flat tariff.	Design day \$615 fixed, \$561.19 band, \$557.61 MPC, \$552.74 band retuned (Tables 9.4, 9.5). Comfort 100 against 97.8 % with notice. Peak 204.4 against 234.8 kW.
10, aeration basin	Fixed oxygen. An ammonia PI supervisor.	A day-ahead plan of the eight-hour mean on the typical load curve.	The same plan with no preview. Nitrogen discharged. The graded day.	Third day 4011 against 4289 kWh (Table 10.6). Nitrogen 35.9 against 29.0 kg. No preview 8622 kWh at a worst mean of 2.607 (Table 10.7).
11, water network	Level-band pump rule with fixed valves.	An hourly integer schedule with modulated valves.	The schedule with fixed valves. Two pumps lost on a hot day.	\$128.1 against \$241.9 on one day (Section 11.6). Challenge \$515.3 against \$571.6, 11 of 11 against 0 of 11 seeds (Table 11.4).
12, electric train	Flat out. Four regimes with one coast point.	A dynamic program, with re-planning and the slip override.	A timetable shorter than the fastest run. Any element of the stack removed.	16.4, 19.5 and 28.2 kWh for DP, coast point and flat out (Table 12.4). Challenge 51.8 kWh passes, 58.8 with four regimes fails (Table 12.5).

The honest baseline is the simpler control retuned once. In the cruise the balloon's cascade swung 0.8 m against the predictive controller's 6.8 m, with less burner travel. On the design day the building's supervisor, retuned once, closed at \$552.74 against the plan's \$557.61, and it won the heat wave as well. The data hall's plan was the cheapest on none of four cases, and a grid sixteen times denser matched or beat it on every case but the job step (Section 8.6.8). The canal's loops with feedforward timed to the wave beat the plan on published orders. A comparison against the shipped tuning alone would have credited the plan with all of these.

The model decided more than the optimizer. The weld's two predictive controllers are the same code with three different numbers in the step response, and the one identified in the plunge finished the weld 91.6 % sound with 18.2 mm of flash. The balloon's linear step response over-predicts a large move by 29 % at 90 s, so a 120 s horizon cycled at 214.4 m whatever the move suppression (Table 2.8). The boiler's nominal MPC held fuel back against a pressure rise that never came, 1.25 kg/cm^2 below its setpoint, until the estimator carried a fuel bias. The data hall's lag model, identified at a 27 % valve, let its valve saturate for 5.0 min in the hot-weather case, where the grid search's did not (Table 8.9). In every case the program was solved correctly. The prediction was wrong.

Most of the money was in the target, not the trajectory. Against a fixed setpoint the data hall's grid search saved 22.7 % of a day's cooling energy, and the building's priced band took the design day from \$615 to \$561. The hourly pump schedule spent \$128.1 where the level band spent \$241.9, and the dynamic program ran the train on 16.4 kWh where flat out took 28.2 kWh. The dynamic plan above an optimizing supervisor added far less. It saved \$6.48 on the data hall's day and \$3.58 on the building's design day, and 6.5 % of the aeration energy where the plan could schedule against the permit's own eight-hour mean. The first optimizer above a simple rule is where the large savings sat.

An optimizer spends every margin it is given. The aeration plan rode its ceiling and exceeded 2.20 g N/m^3 on all 30 seed-days by at most 0.022 g N/m^3, and it discharged 24 % more nitrogen than the supervisor, because nothing in its objective priced the nitrogen below the ceiling. The heating network's plan passed the cold snap with its stores 14.9 MWh lower at the end. A continuous search of the data hall's cost, free of the plan's margins, was the cheapest on both days and rode the water budget for 180.4 and 305.5 min. Each saving in the tables has to be read beside what was spent to earn it.

13.2 Choosing a Method

The chapters suggest an order of questions before a model and an optimizer are put above a plant. Each names the chapter where its answer decided the result.

1. *What will the plan know that a loop does not?* A preview, a published order, a price, a forecast or a limit ahead. If the answer is nothing, retune the loop and add feedforward first. The weld's thin patch and the canal's published orders went to the simpler control (Chapter 3, Chapter 5).

2. *Is the value in where to operate or in how to get there?* When an active constraint moves with the weather or the tariff, a steady-state optimizer or a priced rule takes most of the saving (Section 1.4.7, Chapter 8, Chapter 9). A dynamic plan adds where storage lets energy be bought early or where the measure is itself a mean over hours (Chapter 6, Chapter 10).

3. *Is the model right in the regime the plan will act in?* Take the step test in that regime, keep the horizon inside the range where the linear response holds, and give an integrating output a drift term (Chapter 3, Chapter 2).

4. *What does the plan do about what it cannot measure?* A predictive controller has no integral term of its own. Without a disturbance model it keeps an offset (Section 1.4.5, Chapter 4).

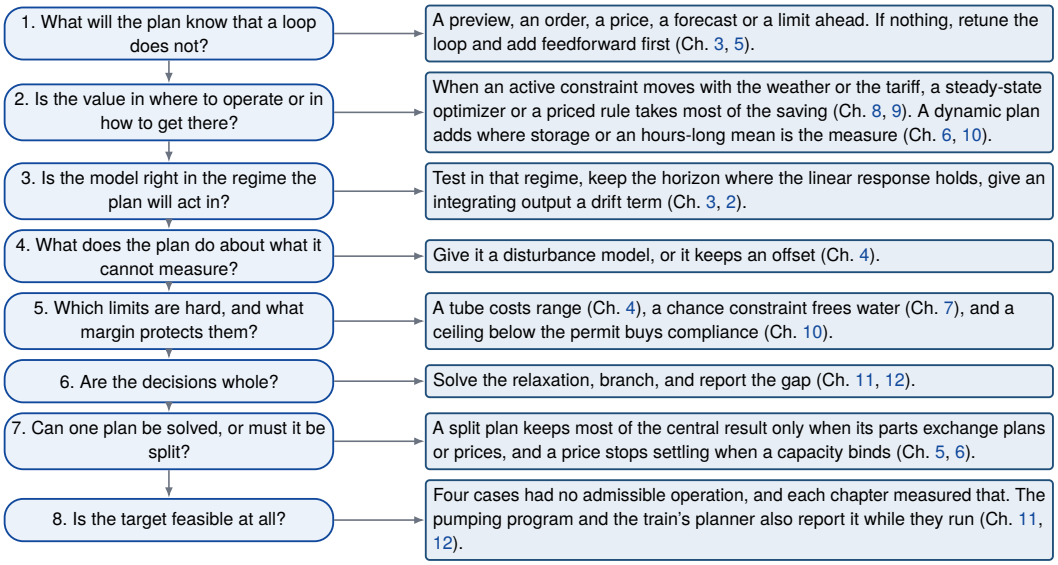


Figure 13.2 The eight questions of this section in order, with the list's answer and the chapters where it decided the result. Notice that what the plan will know and whether its model holds are settled before any optimizer is chosen.

5. *Which limits are hard, and what margin protects them?* A tube costs range, 70.53 kg/cm^2 of 75 usable on the boiler. A chance constraint frees water, 10.4 GWh in the reservoir's spring storm. A ceiling set below the permit costs a little energy and buys compliance on every seed (Chapter 10).
6. *Are the decisions whole?* Pumps and regimes are on or off. Solve the relaxation, branch, and report the gap (Section 1.4.10, Chapter 11, Chapter 12).
7. *Can one plan be solved, or must it be split?* A split plan keeps most of the central result only when its parts exchange plans or prices, and a price stops settling when a capacity binds (Chapter 5, Chapter 6).
8. *Is the target feasible at all?* The boiler's nameplate, a flood with the smallest pool, two pumps lost and a short timetable had no admissible operation. Each chapter measured that, and the pumping program and the train's planner report it while they run.

The comparison that decides whether to keep the plan is then run against the simpler control retuned once, on the same seed, instruments and forecast, as Chapters 5, 8, 9 and 10 do. The aeration plan's saving survives that test at 4.9 % below the best supervisor found with the day known. The building's design-day saving does not. Figure 13.2 draws the eight questions in order.

13.3 Where Practice Is Heading

Predictive control is established, and the work is upkeep. In a 2018 survey of the IFAC Industry Committee, 85 % of the members who answered named model predictive control as a technology with high impact in the future, against 78 % for PID [1]. A review of industrial practice reports that after decades of application, practitioners now focus on ease of commissioning, monitoring and automated maintenance, that many industries need better usability of existing technology more than better algorithms, and that continuous performance monitoring and automated re-identification are being used toward adaptive MPC [2]. The weld of Chapter 3 shows why. A constant output bias absorbed a gain error of 1.7 to 1.9 as an offset that decayed over about a minute, which is how a badly

identified controller can stay in service unnoticed. In the editor's view the next decade of industrial MPC is about keeping models right, not about new formulations.

The objective moves from tracking to cost. Economic MPC, which optimizes an operating cost directly instead of tracking a setpoint chosen elsewhere, has a tutorial review of its methods [3]. For buildings, a review states that MPC has been shown to reduce energy use notably while practical applications remain in early stages, and it calls for standardized performance assessment and methods for comparing building control algorithms [4]. The building of Chapter 9 is a small case of that call. Its plan beat the shipped supervisor and lost to the same supervisor retuned once.

Uncertainty, structure and schedules enter the plan. Stochastic MPC, which plans against a distribution of futures, has an overview of its algorithms and of the theoretical challenges still open [5]. Distributed MPC has a tutorial review of its architectures and future directions [6], and the integration of production scheduling with process control has a systematic review [7]. The reservoir's scenario tree, the canal's exchanged plans, the heating network's price and the pump schedule are small instances of each. In the editor's view these will reach practice first where a plant already publishes a forecast, an order book or a tariff, as these four do.

Computation stops being the limit, and learning enters. Custom methods that exploit the structure of the MPC problem compute the control action on the order of 100 times faster than a generic optimizer, solving a quadratic program with 450 variables in about 5 ms [8]. An open-source operator-splitting solver for quadratic programs detects infeasible problems, supports warm starts and is suited to embedded use [9]. In this volume the slowest boiler solve was 4.8 ms and the building's plan 15 ms. Learning-based MPC, reviewed in three categories, learns the prediction model from recorded data, infers the cost and constraints that give the best closed-loop performance, or uses MPC to add constraint satisfaction to a learned controller [10]. In the editor's view the model, not the solver, will remain the scarce resource, and learning will be judged by whether it keeps the model right in the regime where the plan acts.

Forecasts enter the operating rule. A viability assessment of forecast-informed reservoir operations reports that approved major deviations let the operating agency apply an ensemble forecast model in real-world operations in the water years 2019 and 2020. The model adds a variable buffer pool to the guide curve and recommends flood releases from a 15-day streamflow ensemble forecast. In 2020, the third driest year of a 127-year record, it gave 19 percent more water in storage by the end of winter. The assessment recommends writing the concept into the reservoir's control manual and moving to a further phase only when a defined set of forecast skill metrics improves [11]. The scenario tree of Chapter 7 plans on a seven-member ensemble in the same spirit, and the canal of Chapter 5 on published orders, the older form of the same idea.

Buildings demonstrate the plan but rarely adopt it. A technical report on grid-interactive buildings finds supervisory building control typically implemented as rules, such as setpoint schedules and economizer resets, with many tuning parameters set building by building. It reports early experiments at building sites in which predictive control saved up to 17 to 65 percent of energy against the installed control sequences, and finds it not widely adopted despite several field demonstrations, held back by legacy hardware, the cost of accurate models and operators' acceptance. It also calls for work on how much thermal flexibility can be used without predictive optimization at all [12]. Chapter 9 measured that question on one design day, where a band supervisor retuned once beat the plan.

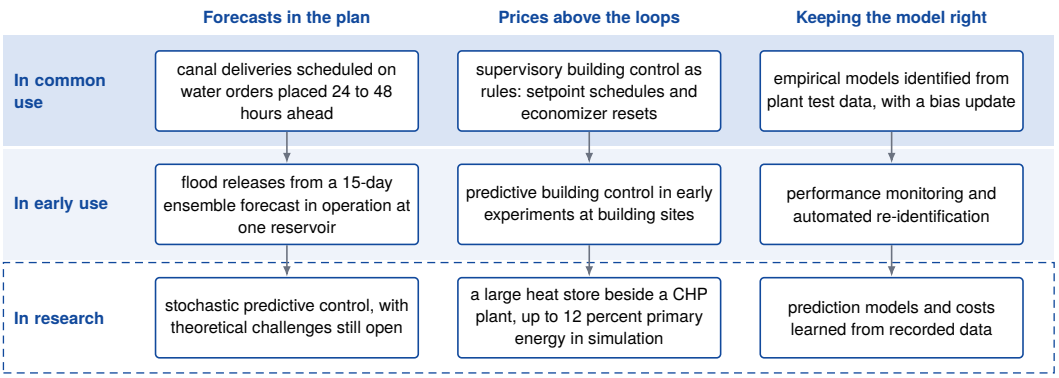


Figure 13.3 Developments in three threads. The top lane holds what the cited source calls typical or the way most industrial controllers work, the middle lane what it reports in operation at one site, in early experiments at sites or being used in practice, and the bottom lane what it reports from simulation or names as a need, in the editor's reading. Sources by thread: forecasts in the plan [5, 11, 14], prices above the loops [12, 15], keeping the model right [2, 10, 16]. No date is drawn, because the sources give none for these steps.

Networks split the plan and keep it working. A paper on automated irrigation reports pilot projects in which sensors, actuators, a communication network and identified models raised the distribution efficiency to better than 90 percent, against 70 percent under manual operation, and met better than 90 percent of water orders on demand. In its simulations of larger canals, decentralized control with feedforward could in most scenarios not be told apart from centralized control, which the paper notes cannot be implemented [13]. The canal of Chapter 5 found the same order on three reaches. After a surprise, plans per gate that exchange held the worst error to 3.10 cm against the central plan's 2.90 cm, and plans that did not exchange to 6.68 cm. Figure 13.3 sets these developments and those above in three threads.

> ⧖ HISTORY **1999 to 2003: Prediction Spreads Beyond the Refinery**
>
> The history boxes of Chapter 1 end in 1986, when limits became rows of a quadratic program. A survey published in 2003, from a snapshot of vendors' figures taken in mid-1999, estimated 4542 linear applications. Refining held 67 percent of those classified, utilities 14 and air and gas 10, and the count had roughly doubled from 1995 to 1999. Industry's research needs were mostly control engineering, not algorithms: which variables to control, when a model is good enough to stop the plant test, why a controller performs poorly, and how to estimate the benefits of a control system [16]. The plants of this volume sit in the small rows of that table, and in the editor's reading its questions are still theirs.

13.4 Engineering With AI Agents

An AI agent is a language model that works toward a goal in a loop. It plans, calls tools such as a compiler, a simulator or a browser, reads the results and continues [17]. In the editor's view, agents can now carry out much of the work that used to fill the hours of a control project. They derive a model from balances or fit one to plant data. They write the simulator and its tests, code the controller, tune it against stated criteria, script commissioning checks, write the documentation and watch running loops for degradation. The apps and chapters of this series were produced that way under the editor's direction, and the account that follows this section reports what the agents got wrong in this volume and how it was caught.

In the editor's view, this changes where an engineer's value lies. When a model, a controller and a report can each be produced in minutes, producing them is no longer the scarce skill. The work moves to five things.

Quality control. An agent's output arrives as a claim: this model fits, this tuning settles in four minutes, this code implements the equation. The engineer checks claims against the plant or the simulator. A reported number is measured again on a fresh run at the stated seed. The code is read for what it clamps, what it initializes and what it reads, because a display that shows the controller's model in place of the plant looks correct until the two differ. A tuning is tried on a disturbance the agent did not test. The most useful checks are independent of the author, made by a second agent or a person who has not seen the first answer, and they are checks that can fail.

Supervision. The engineer sets the scope of each task and decides what an agent may touch. Reading a historian and proposing a tuning is one level of access. Writing a setpoint, downloading a configuration or changing an interlock is another, and it passes through the plant's management of change with a named approver. Supervision also fixes the order of work: a baseline before a change, one change at a time and a comparison that shows what moved.

Creativity. An agent answers the question it is given. Choosing the question remains the engineer's task: which loop is worth improving, whether a disturbance calls for feedforward, a cascade or a change to the process, which step test the operation can tolerate, and which comparison would show that a predictive controller earns its place over a retuned PID loop. The experiment that separates two explanations is designed by someone who can imagine both.

Process insight. A model contains what its author put in. The equipment does more. A valve sticks, the dead time of an analyzer grows when its sample flow drops, a thermowell lags, a heat exchanger fouls and an operator intervenes at the same hour each night. The engineer who knows the unit recognizes a result that is too good, asks what the model left out and knows which measurement to distrust first.

Accountability. Responsibility does not transfer to a tool. Functional safety standards assign the activities of the safety lifecycle to persons and organizations of stated competence, and they require an assessment that includes someone independent of the design team [18]. An agent may draft the hazard analysis or the test procedure. A person owns the decision to accept it, to start up and to keep running.

These five are the professional form of the habit introduced in Section 1.7: check before trusting, and know enough to check.

13.4.1 How This Volume Was Made

Each of the eleven 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 a referee report was written on each paper. At least five of the eleven reports are self-reviews by the agent that wrote the paper. Five apps passed usability, design and technical-accuracy reviews before their papers, and four papers had a further referee pass on predictive material added later. Defects 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 the rules on what each agent could change, and statements that runs contradicted were left for the editor to decide.

The defect logs. The eleven logs written beside the papers hold 87 entries with a severity. Of these, 24 are wrong behavior or output of an app, 46 are documentation, 13 are cosmetic and 4 are other kinds, such as a test or a measurement that is not a defect. The six apps without a review before the paper account for 70 entries, and the five with one account for 17, because their reviews had already logged and fixed findings in each app's own record. Five cases show the errors that matter for a volume about plans and their payoff.

Savings credited to the wrong cause. The data hall's draft risked crediting the plan's horizon with its saving. The referee pass on the predictive layer had the section state that the plan is the cheapest on none of four cases and that the horizon's only content is the lag of its own moves. The building's Conclusion said half of the precooling saving from \$615 to \$561 was demand charge. It was \$22 of \$54, and the half-and-half split belongs to the \$24 between the floating band and precooling. The same paper split a price ratio of 7.5 as about three from the tariff and the rest from the chiller, a sentence inherited from the app's guide. The tariff gives 4.3 and the chiller 1.8. The canal's headline on an unannounced offtake could be read as a win in total error, and a per-reach copy of the plan scored 388 against the plan's 406 (Table 5.8). The text now credits the joint plan only with the worst error. Each was caught by taking a saving apart into its sources.

Numbers taken from the app instead of measured. The weld paper's first draft quoted the first-order-plus-dead-time constants stored in the app as the plant's. Differencing two runs, as the app says it does, gave a plunge time constant of 1.33 s and no dead time where the stored values said 1.92 s and 2.10 s. The chapter's own least-squares fit now reads 1.17 s and no dead time (Table 3.3). The data hall's draft said the binding constraint changes through the diurnal day, following the guide. Rerunning the search found none binding at any of 145 samples. The balloon paper said a higher gain diverges and puts the balloon in a field. Measured, the swing nearly triples and the loop keeps hunting without reaching the ground. All three came from the app's own prose, and the referee of each caught it by measuring.

A horizon result that could have been the terminal weight. The balloon's 214.4 m cycle at a 120 s horizon was first argued from a sweep of move suppression. The referee pointed out that the terminal weight sat on the last points of the horizon, exactly where the linear model is worst. With the terminal weight removed the swing at 120 s still ran from 142.8 to 293.7 m, and at 52 s a fortyfold

terminal weight moved it only from 1.6 to 4.4 m. The conclusion of Table 2.8 now rests on two factors, not one.

A cost booked twice. The aeration draft claimed that below the blower's surge line a higher oxygen setpoint costs less power. The referee read the source and found the vented air billed twice, hardest at low demand, which reversed the sign. The app was fixed and the claim withdrawn. An optimizer given that cost would have been rewarded for raising the oxygen at low load. The fixed-setpoint frontier did not move, because none of those days vents air.

Controllers that read the plant's truth. Five log entries in four apps record a controller or an operator aid that read values a real plant would not give it. The balloon's step test integrated the run's own parameters, air and true state, so that 17 of 34 parameters scaled by 1.25 before Run moved the identified response. The heating network's planners and observers were built on the run's hidden building data. The canal's gate inversion, lead and turnout readings used edited values, and the data hall's start-up registers and a manual helper read plant states. Only the balloon's and the heating network's fixes moved a control result. The balloon's is the one that matters for this volume. Before the fix the predictive controller landed at 0.43 m/s and after it at 0.71 m/s. Part of its margin over the cascade had come from knowing the plant it was landing.

The book round. The eleven chapter records list 197 rows of numeric verification, each a stated value or group of values set beside its reproduction. Nine records report that every rerun number reproduced to its printed digit or its rounding. The aeration record found two values that did not follow from the model and a flow of 4.93 printed as 5.0. All three are now corrected. The pumping record found the last digits of the challenge row moved by cents. The book round's chapter passes logged nine defect entries for this volume. The reservoir's challenge could be passed by raising the channel limit, with the town at 980 m^3/s, and the app now voids the verdict. The balloon's descent envelope never bound near the ground at the shipped slope, so the flare setting the chapter credits did not set the touchdown. After a repair was approved the landing guidance uses the profile directly, and the suggested landings became 1.27 and 1.39 m/s for the cascade and the predictive controller (Table 2.6). Both still stand up, and the predictive controller's earlier margin has reversed. The agent writing Chapter 1 first priced a soft limit from the steady multiplier, 0.04, and the linear program showed the last hour's multiplier to be the full price, 0.20.

Every error above was caught by a check that could fail: a saving taken apart, a number measured instead of copied, a second factor varied, a source line read against the balance, a hidden value moved to see what else moved. Most of them flattered the plan. The headlines that survived are the ones in Tables 13.1 and 13.2. Four working rules follow. Compare every plan against the simpler control retuned once. Take every saving apart into its sources before it is printed. Read what the plan reads and what it spends. Let a person decide what changes.

13.5 Gaps in Predictive Control and Economic Optimization

Six gaps belong to this volume's topic, and the gaps common to agent-era practice follow them in Section 13.5.1. 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 13.4 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. *Keeping the prediction model right in service.* The weld's plunge model, the data hall's lags identified at one valve position and the balloon's step response beyond its linear range each failed without any error in the optimizer, and the boiler's nominal model held an offset under a fuel it could not see until the estimator carried a bias. A routine is needed that detects from closed-loop

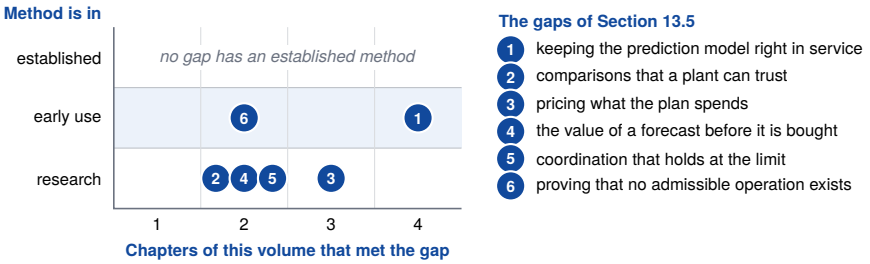


Figure 13.4 The six gaps of this section. Across is the number of chapters that each item names as evidence: four for gap 1 (Chapters 3, 8, 2 and 4), two for gap 2 (9 and 5), three for gap 3 (10, 6 and 8), two for gap 4 (10 and 7), two for gap 5 (6 and 5) and two for gap 6 (7 and 11). Up is the maturity of the method available, the editor's reading of the sources cited in each item and in the Open challenge box. Notice that the gap met by the most chapters, keeping the model right, is also the one closest to practice.

records which part of the model has drifted, re-identifies it in the regime where the plan acts, and does so without a step test the plant cannot tolerate. The Open challenge box at the end of this section treats this gap.

2. *Comparisons that a plant can trust.* The building's plan won against the shipped supervisor and lost to it retuned once, and the canal's best fixed loop setting could be chosen only after the run. A report on building controls cites savings measured against the installed control sequences and calls for work on how much can be had without predictive optimization [12], and the 2003 survey lists how to estimate the benefits of a control system among industry's open questions [16]. The gap is hard because the baseline a plant runs today is rarely the best it could run, and retuning it costs engineering that nobody budgets. Practice needs a standard protocol for an MPC business case, with the baseline retuned by a stated rule, the same information and weather, several seeds, and the saving taken apart into its sources.

3. *Pricing what the plan spends.* The aeration plan saved energy and discharged more nitrogen, the heating network ended with emptier stores, and a margin-free search rode the data hall's water budget for hours. A review of thermal network control warns that more frequent temperature cycling may speed fatigue in steel pipes, and finds a simple, accurate and concise way to state the flexibility available still lacking [15]. The gap is hard because the margins are owned by different people in different units, a permit, a pipe's life, a comfort band. An economic objective that prices the margins a plant cares about, and says which of them a saving consumed, is not yet routine.

4. *The value of a forecast before it is bought.* The aeration plan needed the typical load curve, and the reservoir's chance constraint let the wettest member go with nothing to say about the year in which that member comes. The forecast-informed reservoir assessment ties each further phase of operation to a set of forecast skill metrics still to be defined, and lists their development among its future research [11]. The gap is hard because a forecast's worth depends on the plant's own storage and limits, not on its skill alone. A method that states in advance what a given forecast is worth, and what a wrong one costs, would decide where a plan belongs.

5. *Coordination that holds at the limit.* The heating network's price did not settle in any of the 55 steps after the trip, and the canal's exchanged plans form a Nash point that costs 1.13 times the central optimum on average (Section 5.4.4). The irrigation paper finds that a global controller must cope with varying information rates and delays across the network, and that substantial further research is necessary [13]. The gap is hard because a coordinating price carries no meaning

once no plan fits the shared capacity. Distributed schemes that keep converging when a shared capacity binds, and that say how far from the central plan they are, remain open.

6. *Proving that no admissible operation exists.* The reservoir's infeasible flood rests on a search, not a proof, and the pumping program could only flag its relaxed plan as unable to hold the floor. A study of pump scheduling finds heuristics that cannot guarantee global optimality often applied, and branch and bound slowed from 1 to 43 s by two loops in a test network of 24 time steps, and it asks for tailored solvers [19]. The gap is hard because the hydraulics and the on-off decisions make the program nonconvex. Certificates of infeasibility for nonlinear plants with integer decisions, delivered fast enough to warn an operator, are not available.

> ⚠ **OPEN CHALLENGE** **Keeping the Model Right Where the Plan Acts**
>
> Gap 1 runs through four chapters, each with a model that failed in service while the optimizer solved correctly. The 2003 survey found most industrial controllers built on empirical models identified from plant tests, and industry asking when a model is good enough to stop the test and how to find the source of a problem when a controller performs poorly [16]. A report on building controls names the development and calibration of models accurate and robust enough as a significant challenge to wider use [12]. In the editor's view the gap is hard because the plan's own moves mask the drift it should detect. Progress would be a monitor that names the drifting part of the model from closed-loop records.

13.5.1 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 [20]. In the editor's view, code written by an agent has no comparable body of method yet. Practice needs a routine path from an agent's draft to code a plant can trust: independent derivation, tests of what the code reads and clamps, and trials with hardware in the loop.

2. *Trustworthy models from sparse or poor plant data.* Plant records are short, taken in closed loop, noisy and marked by unmeasured disturbances. A fitting routine returns a model from any record. What is missing is a dependable statement, delivered with the model, of where it can be trusted, and a method that says so when the data cannot identify a parameter.

3. *Safety cases and standards for agents in commissioning and operation.* Functional safety standards define a lifecycle of specified requirements, verification and assigned human roles [18]. In the editor's view, how an agent that writes configuration, runs commissioning checks or advises an operator is qualified, limited and audited under them is not yet settled.

4. *Security of agent access to control systems.* An agent with credentials to a control system is a new path into it. Industrial security standards state the requirements a control system must meet at each security level [21]. 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 [22]. 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 [23]. 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 plan paid for what it could see that a loop could not: a forecast, an order, a price or a limit ahead.
- Most of the saving sat in the first optimizer above a simple rule. A dynamic plan above that optimizer added much less.
- The model, not the solver, decided the result, and keeping it right in service is the open problem.
- An optimizer spends every margin it is given, so each saving is read beside what it spent and against the simpler control retuned once.

References

[1] T. Samad, M. Bauer, S. Bortoff, S. Di Cairano, L. Fagiano, P. F. Odgaard, R. R. Rhinehart, R. Sánchez-Peña, A. Serbezov, F. Ankersen, P. Goupil, B. Grosman, M. Heertjes, I. Mareels, and R. Sosseh, "Industry engagement with control research: Perspective and messages," *Annual Reviews in Control*, vol. 49, pp. 1–14, 2020.

[2] M. G. Forbes, R. S. Patwardhan, H. Hamadah, and R. B. Gopaluni, "Model predictive control in industry: Challenges and opportunities," *IFAC-PapersOnLine*, vol. 48, no. 8, pp. 531–538, 2015.

[3] 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.

[4] J. Drgoňa, J. Arroyo, I. Cupeiro Figueroa, D. Blum, K. Arendt, D. Kim, E. P. Ollé, J. Oravec, M. Wetter, D. L. Vrabie, and L. Helsen, "All you need to know about model predictive control for buildings," *Annual Reviews in Control*, vol. 50, pp. 190–232, 2020.

[5] A. Mesbah, "Stochastic model predictive control: An overview and perspectives for future research," *IEEE Control Systems*, vol. 36, no. 6, pp. 30–44, 2016.

[6] P. D. Christofides, R. Scattolini, D. Muñoz de la Peña, and J. Liu, "Distributed model predictive control: A tutorial review and future research directions," *Computers & Chemical Engineering*, vol. 51, pp. 21–41, 2013.

[7] M. Baldea and I. Harjunkoski, "Integrated production scheduling and process control: A systematic review," *Computers & Chemical Engineering*, vol. 71, pp. 377–390, 2014.

[8] Y. Wang and S. Boyd, "Fast model predictive control using online optimization," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 2, pp. 267–278, 2010.

[9] B. Stellato, G. Banjac, P. Goulart, A. Bemporad, and S. Boyd, "OSQP: An operator splitting solver for quadratic programs," *Mathematical Programming Computation*, vol. 12, no. 4, pp. 637–672, 2020.

[10] L. Hewing, K. P. Wabersich, M. Menner, and M. N. Zeilinger, "Learning-based model predictive control: Toward safe learning in control," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 3, no. 1, pp. 269–296, 2020.

[11] Lake Mendocino FIRO Steering Committee, "Lake Mendocino forecast informed reservoir operations final viability assessment," UC San Diego, Tech. Rep., Dec. 2020. [Online]. Available: https://cw3e.ucsd.edu/FIRO_docs/LakeMendocino_FIRO_FVA.pdf

[12] A. Roth and J. Reyna, "Grid-interactive efficient buildings technical report series: Whole-building controls, sensors, modeling, and analytics," U.S. Department of Energy, Washington, DC, Tech. Rep. NREL/TP-5500-75478, 2019.

[13] I. Mareels, E. Weyer, S. K. Ooi, M. Cantoni, Y. Li, and G. Nair, "Systems engineering for irrigation systems: Successes and challenges," in *IFAC Proceedings Volumes (16th IFAC World Congress)*, vol. 38, no. 1, 2005, pp. 1–16.

[14] U.S. Bureau of Reclamation, *Canal Systems Automation Manual, Volume 1*, 1st ed. Denver, CO: U.S. Department of the Interior, Bureau of Reclamation, 1991. [Online]. Available: https://www.usbr.gov/tsc/techreferences/mands/mands-pdfs/CanalSysAuto1.pdf

[15] A. Vandermeulen, B. van der Heijde, and L. Helsen, "Controlling district heating and cooling networks to unlock flexibility: A review," *Energy*, vol. 151, pp. 103–115, 2018.

[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] L. Wang, C. Ma, X. Feng, Z. Zhang, H. Yang, J. Zhang, Z. Chen, J. Tang, X. Chen, Y. Lin, W. X. Zhao, Z. Wei, and J. Wen, "A survey on large language model based autonomous agents," *Frontiers of Computer Science*, vol. 18, no. 6, p. 186345, 2024.

[18] International Electrotechnical Commission, "Functional safety – Safety instrumented systems for the process industry sector – Part 1: Framework, definitions, system, hardware and application programming requirements," IEC 61511-1:2016, Edition 2.0, 2016.

[19] R. Menke, E. Abraham, P. Parpas, and I. Stoianov, "Exploring optimal pump scheduling in water distribution networks with branch and bound methods," *Water Resources Management*, vol. 30, no. 14, pp. 5333–5349, 2016.

[20] 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.

[21] 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.

[22] 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.

[23] L. Bainbridge, "Ironies of automation," *Automatica*, vol. 19, no. 6, pp. 775–779, 1983.

13.6 Exercises

13.1 Discussion · 8 min

A vendor offers a day-ahead MPC for the office building of Chapter 9 and quotes the design day, \$557.61 against \$561.19 for the shipped supervisor. Using that chapter and Chapter 8, say what comparison you would ask for before buying, and what the two chapters suggest it would show.

13.2 Discussion · 7 min

The aeration plan of Chapter 10 used 277 kWh a day less than the ammonia supervisor at nearly the same worst eight-hour mean, and discharged 24 % more nitrogen. Is that a better way to run the works? Say what you would change in the plan's objective or limits, and what each change would cost.

13.3 Discussion · 8 min

An agent hands you a step-response MPC for a new plant. Its model came from one step test taken during start-up, its horizon is twice the settling time of the step, and its disturbance model is a constant output bias. Using Chapter 3, Chapter 2 and Chapter 4, list what you would check before the controller is switched on, and say which chapter shows the failure each check prevents.

13.4 Discussion · 7 min

The heating network of Chapter 6 and the canal of Chapter 5 each split one plan among several agents. Compare how the two schemes behaved, say when a split plan is worth having, and propose what a split plan should do when its coordination fails.

13.5 Discussion · 9 min

Section 13.4.1 lists errors the agents made while building this volume. Pick three and name the check that caught each. Then write a procedure of five to eight steps for an agent that will propose an

economic MPC for a running plant, stating what it may read, what it may write and what a person must approve.

13.6 Discussion · 6 min

The reservoir of Chapter 7, the water network of Chapter 11 and the train of Chapter 12 each meet a case that no operation can satisfy. Discuss what an optimizing controller should tell the operator in such a case, and why a plan that quietly does its best is not enough.