LESSON BOOK



Time-Series Forecasting for Energy

What is forecastable, how to know a forecast is good, and the decisions it is for

Sixteen lessons · 363 pages

Ricardo Marquez, Ph.D.

ricardo-marquez.com

Time-Series Forecasting for Energy

From the forecasting problem to risk-aware fleet dispatch

Track A lessons, written up as a book

Ricardo Marquez

October 26, 2026 (one chapter per lesson, sixteen lessons)

How to read this book

Each chapter of this book expands one lesson of "Track A", a sequence on time-series forecasting built around energy problems rather than generic examples. The progression is

what is forecastable → baselines → features → models → validation → probabilistic forecasts → fleet forecasting → operational decisions.

The running example throughout is an aggregator's fleet of residential thermostats, electric vehicles and batteries, and the quantity most often forecast is not load but *flexibility*: how much the fleet's consumption could be changed on request. The emphasis is on the physical and operational logic of each method, not on statistics for its own sake. Where the field uses a shorthand term, a *Jargon* box expands it once and the text then uses the expanded meaning. Every symbol is defined in the notation chapter that follows and is used with the same meaning in every chapter. Each chapter closes with the conceptual questions the lesson raised and a set of exercises.

Contents

How to read this book **ii**

Notation and vocabulary **vii**

1 The forecasting problem **1**
1.1 What a forecast is . . . 1
1.2 The target variable in an energy system . . . 3
1.3 Issue time and horizon . . . 5
1.4 The information set . . . 8
1.5 The forecast as a conditional statement . . . 10
1.6 The forecasting loop and a worked example . . . 12
1.7 What to retain . . . 15
1.8 Exercises . . . 16

2 Structure of energy time series **17**
2.1 The decomposition of a load series . . . 17
2.2 Calendar structure . . . 20
2.3 Weather dependence . . . 24
2.4 Persistence and the autocorrelation function . . . 27
2.5 Fleet aggregation and common-mode structure . . . 31
2.6 Nonstationarity, control and the consequences for modeling . . . 33
2.7 What to retain . . . 37
2.8 Exercises . . . 37

3 Baseline forecasts **39**
3.1 The role of a baseline . . . 39
3.2 Naive and seasonal baselines . . . 41
3.3 Baselines that use domain knowledge . . . 44
3.4 Error sources and failure modes . . . 46
3.5 Measuring improvement over a baseline . . . 48
3.6 Baselines for fleet flexibility . . . 51
3.7 What to retain . . . 54
3.8 Exercises . . . 54

4 Feature engineering **56**
4.1 The feature vector and the supervised table . . . 56
4.2 Lag and rolling features . . . 59
4.3 Calendar and weather features . . . 63
4.4 State and control features for a controllable fleet . . . 67
4.5 Target transformations and feature selection . . . 69
4.6 Leakage, latency and feature hygiene . . . 71

4.7 What to retain . . . 74
4.8 Exercises . . . 74

5 Regression and tree models 76
5.1 Supervised learning of the forecast function . . . 76
5.2 Linear models and regularization . . . 79
5.3 Decision trees and ensembles . . . 82
5.4 Gradient boosting . . . 85
5.5 Choice of model class . . . 88
5.6 Diagnostics after fitting . . . 92
5.7 What to retain . . . 95
5.8 Exercises . . . 95

6 Autoregressive models and residual dynamics 97
6.1 The residual as the starting point . . . 97
6.2 Autoregressive models . . . 99
6.3 Autoregressive residual correction . . . 102
6.4 Stationarity, differencing and the ARIMA family . . . 105
6.5 The autocorrelation and partial autocorrelation functions . . . 108
6.6 Model comparison, multi-step forecasting and horizon . . . 110
6.7 What to retain . . . 115
6.8 Exercises . . . 116

7 Recurrent networks 117
7.1 From explicit lags to a learned state . . . 117
7.2 The recurrent network . . . 119
7.3 Vanishing gradients and the long short-term memory cell . . . 122
7.4 The gated recurrent unit . . . 126
7.5 Architectures for forecasting . . . 129
7.6 When to use a recurrent model . . . 132
7.7 What to retain . . . 136
7.8 Exercises . . . 136

8 Attention and transformers 138
8.1 The bottleneck of a single recurrent state . . . 138
8.2 The attention mechanism . . . 140
8.3 Transformers . . . 144
8.4 The temporal fusion transformer . . . 146
8.5 Interpretation and model choice . . . 149
8.6 Data requirements and the fleet case . . . 152
8.7 What to retain . . . 154
8.8 Exercises . . . 154

9 Forecast validation 157
9.1 Time order and the validation split . . . 157
9.2 Error metrics . . . 161
9.3 Slicing by regime and horizon . . . 163
9.4 Event-level and decision-aware validation . . . 166
9.5 Leakage, revisions and drift . . . 168
9.6 Validation protocol and a worked comparison . . . 172
9.7 What to retain . . . 178
9.8 Exercises . . . 178

10 Probabilistic forecasting **181**
10.1 From a point forecast to a distribution . . . 181
10.2 Calibration and sharpness . . . 183
10.3 Learning a predictive distribution . . . 185
10.4 Sources of uncertainty and their propagation . . . 188
10.5 Evaluating probabilistic forecasts . . . 190
10.6 From distribution to decision . . . 194
10.7 What to retain . . . 200
10.8 Exercises . . . 201

11 Fleet forecasting **202**
11.1 The fleet problem . . . 202
11.2 Local, cohort and global models . . . 203
11.3 Bottom-up, top-down and the statistics of aggregation . . . 206
11.4 Hierarchical forecasting and reconciliation . . . 210
11.5 Composition and availability . . . 213
11.6 Heterogeneity and fleet design . . . 216
11.7 What to retain . . . 221
11.8 Exercises . . . 221

12 Flexibility and counterfactual forecasting **223**
12.1 The missing world . . . 223
12.2 Two architectures and the baseline problem . . . 226
12.3 Response as a function of action and state . . . 230
12.4 Causal estimation at fleet scale . . . 233
12.5 Uncertainty and the interventional distribution . . . 238
12.6 Architecture, validation and the worked event . . . 240
12.7 What to retain . . . 246
12.8 Exercises . . . 246

13 Forecasting under distribution shift **249**
13.1 The changed data-generating process . . . 249
13.2 Sources of shift in a fleet . . . 252
13.3 Recognizing unfamiliar inputs . . . 255
13.4 Monitoring and diagnosis . . . 259
13.5 Regimes, actions and the optimizer . . . 264
13.6 Operating under shift . . . 268
13.7 What to retain . . . 272
13.8 Exercises . . . 274

14 Forecast to optimization **275**
14.1 Information versus action . . . 275
14.2 Reliability as a constraint . . . 277
14.3 Cost, quantiles and portfolios . . . 281
14.4 Multi-period dispatch and model predictive control . . . 284
14.5 Stochastic and robust formulations . . . 287
14.6 Designing the interface and closing the loop . . . 291
14.7 What to retain . . . 297
14.8 Exercises . . . 297

15 Production forecasting systems **300**
15.1 From model to system 300
15.2 The data pipeline 303
15.3 Features and serving 307
15.4 Monitoring 311
15.5 Failure handling and retraining 313
15.6 Learning from operation and the control loop 318
15.7 What to retain 324
15.8 Exercises 324

16 Capstone: 75 MW on the hottest day **327**
16.1 The case and the decision 327
16.2 The baseline and response models 330
16.3 Fleet representation and uncertainty 334
16.4 The hottest day and the optimizer 337
16.5 Closing the loop and validating the design 341
16.6 The pushback round and the end-to-end design 346
16.7 What to retain 352
16.8 Exercises 353

Notation and vocabulary

The same symbols are used throughout. Time t is the instant at which a forecast is *issued*; h is the *horizon*, the time between issue and target; power is in kilowatts (kW) or megawatts (MW); temperature is in degrees Celsius.

Symbols

Symbol	Unit	Meaning
t	time	Forecast issue time: the moment the prediction is made.
h	time	Forecast horizon; the target time is $t + h$.
y_{t+h}	–	The target: the quantity being predicted.
$\hat{y}_{t+h}$	–	A point forecast of the target.
$\mathcal{I}_t$	–	Information set: everything genuinely known at issue time t.
$p(y_{t+h} \mid \mathcal{I}_t)$	–	The probabilistic forecast: a distribution over the target given the information.
u	–	A planned control action (setpoint change, start, duration, cohort); the target may depend on it.
P_t	kW, MW	Load (power) at time t.
$P^{\text{base}}(t)$	MW	Baseline load: what the fleet would consume without intervention.
$P^{\text{ctrl}}(t \mid u)$	MW	Controlled load: what it would consume under action u.
$F(t)$	MW	Flexibility, $P^{\text{base}} - P^{\text{ctrl}}$; a counterfactual quantity.
T_t, T_{out}	°C	Outdoor temperature at t.
$\hat{T}_{t+h\mid t}$	°C	Weather forecast for $t + h$ as issued at t.
CD_t, HD_t	°C	Cooling and heating degrees, $\max(T_t - T^*, 0)$ and $\max(T^* - T_t, 0)$ about a base T^*.
RH_t	%	Relative humidity.
N_{avail}	–	Number of devices available (connected, enrolled, not opted out).
D_t	–	Duty cycle: fraction of the fleet's compressors running.
X_t	–	Feature vector built from $\mathcal{I}_t$: lags, weather, calendar, state.
$\boldsymbol{\beta}$	–	Coefficients of a linear model, $\hat{y} = X_t\boldsymbol{\beta}$.
λ	–	Regularization weight (ridge, lasso).
η	–	Learning rate of a boosted ensemble.

Symbol	Unit	Meaning
$T_m(X)$	–	The m-th tree of an ensemble.
ε_t	–	Noise or residual at t.
e_t	–	Forecast error, $y_t - \hat{y}_t$.
ϕ_k, θ_k	–	Autoregressive and moving-average coefficients.
p, d, q	–	ARIMA orders: AR lags, differences, MA lags; $(P, D, Q)_s$ their seasonal counterparts at period s.
ACF(k), PACF(k)	–	Autocorrelation and partial autocorrelation at lag k.
$\mathbf{h}_t$, $\mathbf{c}_t$	–	Hidden state and cell state of a recurrent network at step t (bold, to distinguish from the horizon h).
$\mathbf{Q}, \mathbf{K}, \mathbf{V}$	–	Query, key and value matrices of an attention layer; d_k the key dimension.
MAE, RMSE, MAPE, Bias	–	Mean absolute error, root-mean-square error, mean absolute percentage error, mean error.
Skill	–	$1 - \mathrm{RMSE}_{\mathrm{model}}/\mathrm{RMSE}_{\mathrm{baseline}}$.
τ	–	Quantile level; $\hat{y}_\tau$ the forecast τ-quantile.
$L_\tau(y, \hat{y})$	–	Pinball (quantile) loss at level τ.
CRPS	–	Continuous ranked probability score of a distributional forecast.
$y_{i,t}$	–	The series of device (or cohort) i at time t.
x_t, w_t	–	Latent physical state and disturbance in the state-space view, $x_{t+1} = g(x_t, u_t, w_t)$, $P_t = h(x_t)$.

Vocabulary

Issue time, target time, horizon
: When the forecast is made, when the predicted quantity occurs, and the gap between them.

Information set
: Everything that actually exists at issue time; a forecast may use nothing outside it.

Data leakage
: Training or evaluating with information that would not exist at issue time, for example an observed future temperature in place of the forecast that was available.

Point forecast, probabilistic forecast
: A single number versus a distribution (or a set of quantiles) for the target.

Baseline (model)
: A simple forecast a sophisticated model must beat: persistence, seasonal persistence, similar-day average, weather regression.

Baseline (load)
: In demand response, the counterfactual load without intervention. The two meanings coexist in this book; the context makes clear which is meant.

Persistence
: The forecast that the next value equals the last one; seasonal persistence uses the value one cycle earlier.

Skill score
: Error reduction relative to a baseline.

Trend, seasonality, autocorrelation
: Slow change; repeating daily, weekly or annual cycles; dependence between a series and its own past.

Exogenous variable
: A predictor outside the series itself: weather, calendar, device state.

Stationarity
: Constancy of a series' statistical properties in time; differencing is the usual way to obtain it.

Feature engineering
: Turning raw timestamps, loads and weather into the predictor vector X_t: lags, rolling statistics, degree variables, cyclic encodings.

Regularization
: A penalty on model complexity added to the training error (ridge, lasso) to improve generalization.

Gradient boosting
: An ensemble of small trees fitted sequentially, each to the residual of the ones before.

Recurrent network
: A model whose hidden state summarizes the history it has seen; LSTM and GRU are gated variants that control what is remembered.

Attention
: A mechanism that weights past observations by learned relevance to the current prediction; the basis of transformers.

Rolling-origin validation
: Evaluating a forecaster by repeatedly training up to a date and testing on the period after it, never on a random row split.

Calibration, sharpness
: Whether stated probabilities match observed frequencies; how narrow the forecast distribution is.

Pinball loss
: The loss whose minimizer is a quantile.

Global model, cohort model, local model
: One model for all series, one per group of similar series, one per series.

Hierarchical forecasting, reconciliation
: Forecasting at several aggregation levels and adjusting the forecasts so they add up.

Common-mode uncertainty
: Error shared by every device in a fleet (for example a weather forecast error); aggregation does not average it away.

Counterfactual
: What would have happened under a different intervention; never observed alongside what did happen.

Chapter 1

The forecasting problem: target, horizon and information set

1.1 What a forecast is

Suppose it is 2:00 PM on a hot afternoon. An aggregator operates a fleet of residential thermostats and needs to know how much of that fleet's load it can curtail between 5:00 and 6:00 PM. The natural first move is to start building a forecasting model. That move is premature, because the question as stated does not yet name a quantity. At least three different numbers are hiding inside it, and they have different values, different inputs and different error consequences. The chapter's question is therefore

what exactly is being predicted, for what moment, and from what information?

An *aggregator* is a company that enrolls many small customer-owned devices — thermostats, water heaters, batteries, electric-vehicle chargers — and operates them together as one resource it can sell to a utility or a market. To *curtail* the fleet is to reduce its electrical consumption on request, which for a thermostat fleet means raising cooling setpoints so that compressors stop. The *setpoint* is the indoor temperature a thermostat is trying to hold.

Jargon: issue time, target, horizon, information set, data leakage, counterfactual

The *issue time* t is the moment at which the forecast is made and after which no new information may enter it. The *target* y_{t+h} is the quantity being predicted, at the *target time* $t + h$; the gap h is the *horizon*. The *information set* $\mathcal{I}_t$ is everything that genuinely exists and is accessible at the issue time. *Data leakage* is the use, in training or in evaluation, of anything outside $\mathcal{I}_t$ — most commonly an observed future weather value in place of the forecast that was actually available. A *counterfactual* quantity is one defined by what would have happened under an action that was not taken; it is never observed alongside what did happen.

Key idea

The fundamental object of forecasting is not a number $\hat{y}$ but the conditional distribution $p(y_{t+h} \mid \mathcal{I}_t)$: a statement about a named target, at a named time, given only what was genuinely known when the statement was made.

This section fixes the three things a forecasting problem must name before any model is chosen, writes the general formulation they imply, and distinguishes the deterministic from the probabilistic version of it.

1.1.1 Three targets hiding in one question

"How much load can we curtail from 5:00 to 6:00 PM" resolves into three distinct quantities. The first is the *baseline load* $P^{\text{base}}(t)$: what the fleet would consume at that hour if the aggregator did nothing. The second is the *controlled load* $P^{\text{ctrl}}(t \mid u)$: what it would consume under a specific intervention u. The third is the *available flexibility* $F(t) = P^{\text{base}}(t) - P^{\text{ctrl}}(t \mid u)$, the reduction the intervention achieves. With the illustrative numbers used throughout this chapter, $P^{\text{base}}(17{:}00) = 120\,\text{MW}$, $P^{\text{ctrl}}(17{:}00 \mid u) = 75\,\text{MW}$, and the flexibility is 45 MW.

Only the third number answers the operational question, and it is the only one of the three that is never directly measured. Section 1.2 treats each in turn. The immediate point is structural: a forecasting problem is not defined by the data available or by the model class, but by a choice of target, and that choice is made by the person asking the question, not by the modeler.

1.1.2 The triple that defines a forecasting problem

Every forecasting problem is specified by three things, and changing any one of them produces a different problem:

(i) the **target** y_{t+h} — which physical quantity, for which population of devices, at which temporal resolution;

(ii) the **horizon** h and the issue time t — how far ahead, measured from when;

(iii) the **information set** $\mathcal{I}_t$ — what is genuinely available at t.

Two problems that share a target but differ in horizon need different predictors and often different model classes (Section 1.3). Two problems that share a target and a horizon but differ in what is known at issue time — for example a day-ahead forecast made with and without a numerical weather prediction — have different attainable accuracy (Section 1.4).

A fourth question sits beside the triple and determines how the forecast should be summarized: which decision will consume it. That question does not change the distribution $p(y_{t+h} \mid \mathcal{I}_t)$, but it changes which feature of the distribution is reported and which errors are expensive. Section 1.5 takes it up.

1.1.3 The general formulation

At the issue time t the forecaster holds an information set, which for the aggregator contains the recent fleet load, the recent and current outdoor temperature, a weather forecast for the hours ahead, the state of the devices, the calendar, and the fleet's historical behavior,

$$\mathcal{I}_t = \{\, P_t,\ P_{t-1},\ \ldots,\ T_t,\ \hat{T}_{t+1:t+H|t},\ \text{device state},\ \text{calendar},\ \text{history} \,\}. \tag{1.1}$$

The target is y_{t+h}. The deterministic version of the problem is to produce a single number,

$$\hat{y}_{t+h} = f(\mathcal{I}_t), \tag{1.2}$$

and the probabilistic version is to produce a distribution,

$$p(y_{t+h} \mid \mathcal{I}_t). \tag{1.3}$$

Equation (1.3) is the more complete problem and equation (1.2) is a summary of it. The question it answers is not "what will the load be" but "given everything known now, which values of the future load are plausible, and with what probabilities". The remainder of the book is organized around this object: Chapters 2 to 8 build models that estimate its center, Chapter 9 tests them, Chapter 10 estimates its spread and shape, and Chapter 11 studies what happens to it when thousands of series are forecast at once.

Principle 1.1. *A forecasting problem is defined before any model is chosen, by naming the target, the issue time and horizon, and the information set. Model selection is downstream of that definition, and an undefined target cannot be forecast well by any method.*

1.2 The target variable in an energy system

The three candidate targets of Section 1.1.1 are different functions of the same fleet. This section defines each, identifies which is a conventional forecasting problem and which is not, and states what makes a target well defined.

1.2.1 Baseline load

The baseline load $P^{\text{base}}(t)$ is the fleet's consumption in the absence of intervention. It is a conventional load forecast: a function of the recent load, the weather and the calendar,

$$P^{\text{base}}(t+h) = f(P_t,\ P_{t-1},\ T_{\text{out}},\ \text{hour},\ \text{day type},\ \ldots), \tag{1.4}$$

and it answers the question "what will happen naturally". Every technique in this book applies to it directly, because the quantity is observed whenever no intervention occurs, so training data and verification data both exist in quantity.

The term *baseline* carries two meanings in this field and both appear in this book. Here it means the counterfactual load without intervention, which is the demand-response usage. In Chapter 3 it will mean a simple reference forecast that a sophisticated model must beat. The context separates them; where confusion is possible, this book writes "baseline load" for the first and "baseline model" for the second.

1.2.2 Controlled load

Now suppose the aggregator raises thermostat setpoints. The controlled load $P^{\text{ctrl}}(t \mid u)$ is what the fleet consumes under that specific action. The conditioning argument u is not a scalar; it is a description of the intervention,

$$u = \{\,\text{setpoint change},\ \text{start time},\ \text{duration},\ \text{device cohort}\,\}, \tag{1.5}$$

where a *cohort* is a subset of the fleet selected to receive the command, for instance all thermostats in one climate zone. The forecast is accordingly

$$P^{\text{ctrl}}(t+h \mid u) = f(\mathcal{I}_t,\ u). \tag{1.6}$$

This is a different problem from equation (1.4), because the answer depends on what the aggregator decides to do. Two features follow. The target is not a property of the world alone, so a model fitted to historical load without recording which commands were sent is fitting a mixture of the two quantities. And the data available for fitting is much thinner: the baseline load is observed on every non-event day, while $P^{\text{ctrl}}(\cdot \mid u)$ is observed only on the days on which that particular kind of u was actually dispatched.

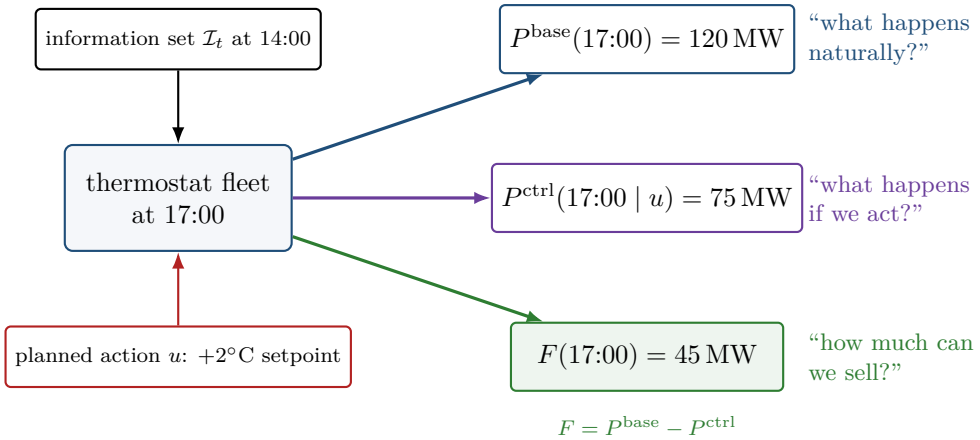
1.2.3 Available flexibility as a counterfactual

The quantity the aggregator sells is the difference,

$$F(t) = P^{\text{base}}(t) - P^{\text{ctrl}}(t \mid u) = 120 - 75 = 45\,\text{MW}, \tag{1.7}$$

a load reduction of 45 MW. Figure 1.1 draws the three targets together: a single box on the left for the fleet, with the information set entering it from above and the planned action u entering from below, and three target boxes on the right for the baseline load, the controlled load and their difference, each labeled with the question it answers and its illustrative value. The figure makes visible that only the two upper targets are powers that could in principle be metered; the third is constructed from them.

That construction is the difficulty. Flexibility is a *counterfactual* quantity: if the aggregator dispatches the fleet it observes P^{ctrl}, and the P^{base} for the same devices under the same weather at the same instant then does not exist to be measured. Figure 1.2 shows the two worlds against time. The horizontal axis is the clock from 16:00 to 19:00 and the vertical axis is the fleet's power; the two curves coincide before 17:00, because nothing has been done yet, and separate afterward, with the controlled trajectory drawn solid because it is the one observed and the baseline trajectory dashed because it is not. The shaded gap during the event hour is F. Since the dashed curve is never recorded, every flexibility measurement is really a flexibility *estimate*, obtained by substituting a modeled baseline for the missing world, and the error of that model enters the quantity being sold. Chapter 3 builds the estimators, and the rebound visible after 18:00 in the figure — the controlled curve overshooting the baseline as warmed buildings cool back down — is a second reason the two worlds cannot simply be spliced.



The upper two targets are powers that could be metered;
the third is their difference and is never measured directly.

Figure 1.1: The three candidate targets inside one operational question. The fleet is shown with the information set entering from above and the planned action u from below; each target box states the quantity, its illustrative value from Section 1.1.1, and the question it answers. Only P^{ctrl} is observed when the fleet is dispatched, and only P^{base} when it is not.

1.2.4 What makes a target well defined

A target is well defined when it is reproducible: two analysts given the same data and the same specification compute the same series. For the quantities above that requires four things to be stated. The *population*: which devices, enrolled on what date, including or excluding those that have opted out. The *resolution*: a 5-minute average power is a different series from an hourly average, and a peak is different from either. The *units and sign convention*: this book takes $F > 0$ to mean a load reduction. And, for the control-dependent targets, the *action* u in the sense of equation (1.5).

Omitting any of the four produces a target that looks like a single series but is a mixture. A common instance is a flexibility series assembled from events of varying duration: a 15-minute

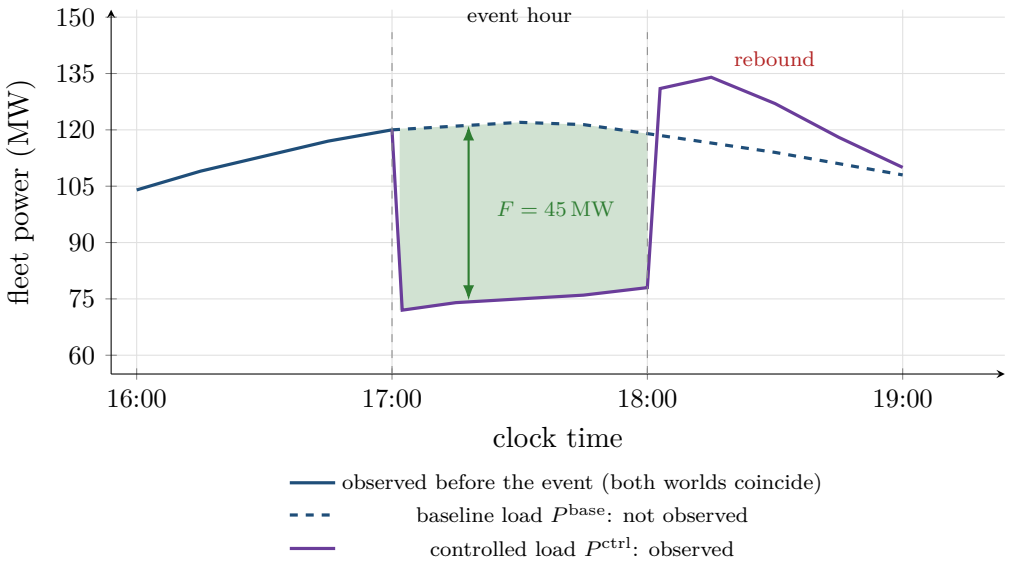


Figure 1.2: Flexibility as a counterfactual. Fleet power against clock time for a one-hour curtailment beginning at 17:00. The two worlds coincide before the event; afterward the controlled trajectory is observed and the baseline trajectory is not. The shaded area is F of equation (1.7), averaging 45 MW over the event hour. Trajectories are illustrative, drawn through the stated values of Section 1.1.1.

curtailment and a three-hour curtailment of the same fleet yield very different reductions, and averaging them produces a number that describes no event the fleet ever performed. A second instance is a fleet whose enrollment grew over the sample, so that the target's trend is partly an accounting artifact rather than a physical change.

1.3 Issue time and horizon

The same target at two different horizons is two different statistical problems. This section separates the issue time from the target time, works through three horizons for one target, and draws the consequence for model selection.

1.3.1 Issue time and target time

Fix $t = 14{:}00$ and a target at 17:00, so $h = 3$ hours. The rule is that the model may use only information available at 14:00. Figure 1.3 draws the arrangement: the horizontal line is the clock, the left region shaded blue is time that has already happened at the issue instant and the right region shaded red is time that has not, the arrow spanning them is the horizon, and the two labeled items below the axis are a temperature at 16:30 in two versions. The observed version $T_{16:30}$ is marked unavailable: at 2:00 PM it does not exist. The forecast version $\hat{T}_{16:30|14:00}$ is available, because it is a weather forecast issued at 2:00 PM for 4:30 PM.

The distinction is purely about when a value came into existence, and it is enforced not only at prediction time but also during training and evaluation. A model trained on $T_{16:30}$ learns the relationship between load and the temperature that actually occurred; operationally it will be fed $\hat{T}_{16:30|14:00}$ instead, which is a different and noisier variable. Section 1.4.3 quantifies the gap.

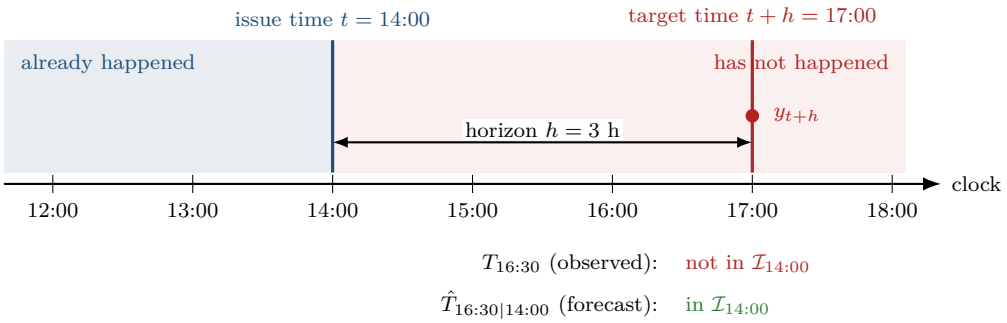


Figure 1.3: Issue time, horizon and target time. The clock runs left to right; the blue region is time that has already elapsed at the issue instant and the red region is time that has not. The two entries below the axis are the same physical temperature in two versions: the observed value, which does not yet exist at 14:00, and the forecast value issued at 14:00, which does.

1.3.2 The horizon changes which information matters

Take the fleet load as the target and vary only the horizon. Five minutes ahead, the load is dominated by the load now, because almost nothing in the physical system has had time to change: the same compressors are running, the buildings have the same indoor temperatures, and

$$P_{t+5\,\mathrm{min}} \approx f(P_t,\ P_{t-5\,\mathrm{min}},\ \ldots) \tag{1.8}$$

is an accurate model. The simplest such model, $\hat{y}_{t+h} = y_t$, is called *persistence*, and at this horizon it is a strong competitor.

Three hours ahead, persistence is weak and the drivers are the weather forecast, the time of day and the fleet's thermal state,

$$P_{t+3\,\mathrm{h}} = f(\text{recent load},\ \hat{T}_{t+3\mathrm{h}|t},\ \text{hour},\ \text{thermal state}). \tag{1.9}$$

A day ahead, the instantaneous state has almost no value — the fleet will have passed through an entire evening and night — and what remains informative is the weather forecast, the day of week, the season and the enrolled population's behavior. Figure 1.4 shows the shift as three stacked bars, one per horizon; the vertical axis is the share of predictive information attributable to each of three groups of inputs, and the three segments of each bar are the recent load, the weather forecast, and the calendar and behavior terms. The shares are illustrative, but the ordering is general: the recent-load segment shrinks monotonically with the horizon while the weather and calendar segments grow.

1.3.3 There is no universal best forecasting model

The practical consequence is that a claim of the form "gradient boosting is the best load forecasting method" is not well posed until the horizon is fixed. Table 1.1 gives the three horizons with the information that dominates each, the decision each supports in an aggregator's operation, and the model class that typically wins. The rows are not three tunings of one model; they are three problems with different predictors, different achievable accuracy and different reference baselines.

The same reasoning applies to the target. A method that forecasts P^{base} well need not forecast F well, because the second is a difference of two forecasts and inherits the errors of both. For the rest of the book, every statement about accuracy is a statement about a specific target at a specific horizon, and comparisons are made within a row of Table 1.1, never across rows.

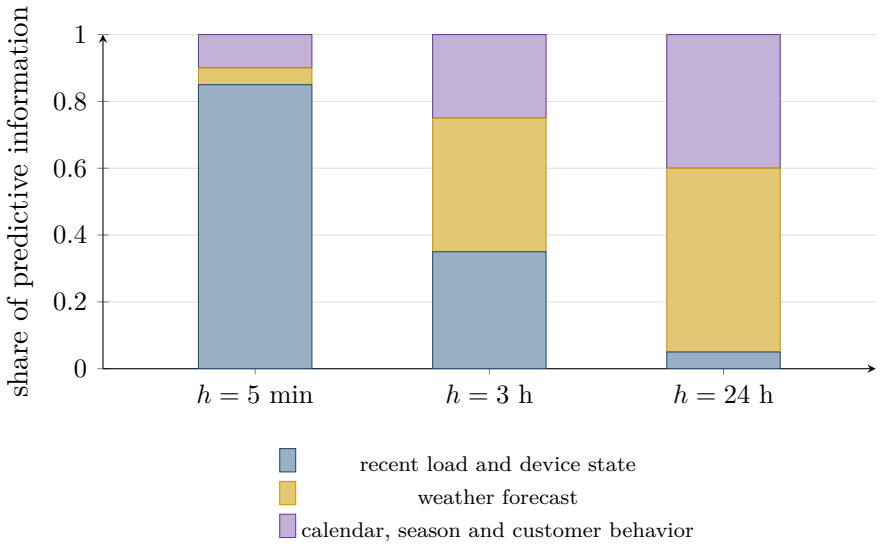


Figure 1.4: How the dominant information changes with the horizon, for one fixed target. Each bar divides the predictive information into three groups of inputs; shares are illustrative and are the ones used in Section 1.3.2. The recent-load share falls from 0.85 at five minutes to 0.05 at a day, and the weather share rises from 0.05 to 0.55.

Table 1.1: The same target at three horizons: three different problems.

Horizon	Dominant information	Decision supported	Typical winning model
5 min	current load, compressor state, indoor temperatures	real-time correction of a dispatch already running	persistence and short-lag autoregression
3 h	recent load plus the weather forecast and the fleet's thermal state	whether to call an event this afternoon, and how large	regression or gradient boosting on lag and weather features
24 h	weather forecast, day type, season, enrolled population, behavior	day-ahead capacity offer to the utility	weather-driven model with calendar features, reported probabilistically

1.4 The information set

The information set is the formal statement of what a forecast is allowed to use. This section enumerates its contents for the aggregator, separates forecast weather from observed weather, shows what happens when the separation is violated, and notes that in practice the set is smaller than the clock suggests.

1.4.1 Contents of the information set at issue time

Equation (1.1) lists the groups abstractly; Table 1.2 instantiates them for the aggregator at 14:00, with the nature of each entry in the last column. The grouping matters because the four groups behave differently. Measured history is exact but stale the instant it is recorded. Device state is partly measured and partly inferred, since a thermostat reports its indoor temperature but not the thermal mass of the building around it. Exogenous forecasts are themselves uncertain and carry their own error into the chain. The calendar is known exactly and arbitrarily far ahead, which makes it the only input whose quality does not degrade with the horizon.

The information set also grows. Figure 1.5 draws $\mathcal{I}_{13:00}$, $\mathcal{I}_{14:00}$ and $\mathcal{I}_{15:00}$ as nested rectangles over a common time axis: each one contains all of its predecessors and extends further to the right, and the target at 17:00 is marked outside all three. Two properties are visible. A forecast for 17:00 issued at 15:00 can be no worse than one issued at 14:00, because it conditions on a superset of the information, which is why operational forecasts are reissued as time passes. And the target itself never enters any of the sets drawn, which is what makes the problem a forecasting problem rather than an interpolation.

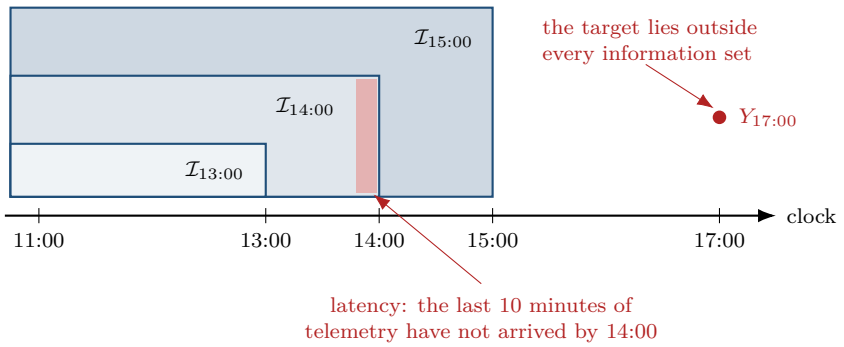
Table 1.2: The aggregator's information set $\mathcal{I}_{14:00}$, by group.

Group	Contents at 14:00	Nature
Measured history	fleet load $P_t, P_{t-1}, \ldots$; outdoor temperature T_t and its recent path; past events and their verified response	exact but subject to latency
Device state	indoor temperatures T_{in}, setpoints T_{set}, compressor on/off, duty cycle D_t, devices available N_{avail}	partly measured, partly inferred
Exogenous forecasts	$\hat{T}_{t+1:t+H\|t}$: temperature, humidity RH, cloud cover, wind, for every hour ahead	uncertain; error grows with h
Calendar and known plans	hour, day of week, holiday, season; scheduled maintenance; the aggregator's own planned action u	known exactly, arbitrarily far ahead

1.4.2 Forecast weather and observed weather are different variables

Weather enters an energy forecast in two incompatible forms. The observed temperature T_s at a past time $s \le t$ is a measurement and belongs to $\mathcal{I}_t$. The temperature at a future time $t + h$ does not belong to $\mathcal{I}_t$ in any form; what belongs is the forecast $\hat{T}_{t+h|t}$, a prediction issued by a weather service at or before t. These are not two names for one quantity. The forecast has its own error, which grows with h, varies by variable — temperature is forecast far better than cloud cover or wind — and is correlated across the hours of a day, because a numerical weather prediction run that is too warm at noon is usually too warm at 3:00 PM as well.

The consequence is that an energy forecast at horizon h is a composition of two forecasts, and its error has a component inherited from the weather forecast. If the load responds to



$\mathcal{I}_{13:00} \subset \mathcal{I}_{14:00} \subset \mathcal{I}_{15:00}$: each issue time conditions on a superset of the one before.

Figure 1.5: The information set as a growing family. Each rectangle is the information available at one issue time, drawn over a common clock axis; later sets contain earlier ones. The red sliver inside $\mathcal{I}_{14:00}$ is the telemetry that exists physically but has not been delivered by 14:00 (Section 1.4.4). The target at 17:00 belongs to none of the sets.

temperature with sensitivity $\partial P/\partial T$ and the weather forecast has temperature error standard deviation σ_T, then under independence the inherited term adds in quadrature,

$$\sigma_P^2 \;=\; \sigma_{\text{model}}^2 \;+\; \left(\frac{\partial P}{\partial T}\right)^2 \sigma_T^2. \tag{1.10}$$

For the fleet of this chapter, a plausible sensitivity is 4 MW per degree Celsius, and a three-hour-ahead temperature forecast might have $\sigma_T = 1.5°\text{C}$; the inherited term is then 6 MW, which is larger than the error of many well-built load models. Weather forecast error is not a detail of the input pipeline; at useful horizons it is often the dominant term.

1.4.3 Data leakage

Suppose a modeler fits

$$P_{17:00} = f(T^{\text{obs}}_{17:00}) \tag{1.11}$$

using the temperature that actually occurred, and reports excellent accuracy on held-out data. The model is not wrong as a description of the relationship between load and temperature. It is unusable, because at 14:00 the quantity $T^{\text{obs}}_{17:00}$ does not exist. The operationally realizable model is

$$P_{17:00} = f(\hat{T}_{17:00|14:00}), \tag{1.12}$$

and equation (1.10) says by how much its error exceeds the reported figure. A model with a 3 MW in-sample error and the sensitivity of the previous subsection has an operational error of $\sqrt{3^2 + 6^2} = 6.7$ MW, more than double the number that was reported.

Leakage of this kind is easy to introduce accidentally, because historical archives usually store the observed weather and not the forecast that was current at each issue time. Three habits prevent it. Reconstruct the information set explicitly for each training row, by asking what existed at that row's issue time, rather than selecting columns from a table. Store weather forecasts with both their valid time and their issue time, so that $\hat{T}_{t+h|t}$ can be retrieved for the correct t. And evaluate with the same inputs that will be used operationally, a discipline taken up in Chapter 9.

Leakage through the target. The same error occurs on the target side when the series being predicted has been revised after the fact. A verified flexibility number computed weeks

later with a settled baseline model is not the number an operator could have had, and a model trained to predict the settled figure while evaluated against it will overstate what the operator can expect.

1.4.4 Latency makes the information set smaller than the clock

In a deployed system the information set at 14:00 does not extend to 14:00. Meter reads are aggregated and transmitted on a cycle; a reading timestamped 13:55 may not be available to the model until 14:07. The effective information set therefore ends at some $t - \delta$, with δ a delivery latency of minutes to an hour, and this is the red sliver drawn inside $\mathcal{I}_{14:00}$ in Figure 1.5.

Latency changes the problem rather than merely degrading it. A forecast nominally at $h = 5$ minutes built on data that is 12 minutes old is in truth a 17-minute-ahead forecast, and the persistence term of equation (1.8) is correspondingly weaker. The specification of a forecasting problem should therefore state the latency alongside the horizon, and a model should be trained on data lagged by the same δ it will face in production.

1.5 The forecast as a conditional statement

A forecast is a statement about a distribution, and a single number is one summary of it. This section contrasts two forecasts that share a mean, identifies the summary the decision actually needs, and adds the control action to the conditioning set.

1.5.1 A point forecast is a summary of a distribution

The statement $\hat{F} = 45\,\text{MW}$ is a *point forecast*: one number for the target. The statement $p(F \mid \mathcal{I}_t)$ is a *probabilistic forecast*: a distribution over the target. The first is recoverable from the second by choosing a summary — the mean, the median, or some other quantile — and the choice of summary is not innocuous, as Chapter 10 shows. The second is not recoverable from the first at all.

Consider two probabilistic forecasts of the same target,

$$\text{A: } F \sim \mathcal{N}(45,\, 2^2), \qquad \text{B: } F \sim \mathcal{N}(45,\, 15^2), \tag{1.13}$$

in MW. Both have $\mathbb{E}[F] = 45\,\text{MW}$, so both collapse to the same point forecast, and they describe very different situations. Forecast A says the fleet will deliver between about 41 and 49 MW; forecast B admits anything from 15 to 75 MW. An aggregator holding forecast A can make a firm commitment near 45 MW; one holding forecast B cannot.

1.5.2 The decision context selects the summary

Suppose the utility requires at least 40 MW of reduction and pays nothing, or imposes a penalty, if the fleet delivers less. The operationally relevant quantity is then not the expectation but the exceedance probability

$$\Pr(F \geq 40\,\text{MW} \mid \mathcal{I}_t), \tag{1.14}$$

which the two forecasts answer very differently. Under A the shortfall threshold is 2.5 standard deviations below the mean and $\Pr(F \geq 40) = 0.994$; under B it is one third of a standard deviation below and $\Pr(F \geq 40) = 0.631$. The same point forecast corresponds to a near-certainty and to a situation that fails on more than one occasion in three.

Figure 1.6 plots the two densities in separate panels, because their vertical scales differ by a factor of seven. In each panel the horizontal axis is delivered flexibility, the curve is the forecast density, the vertical line is the 40 MW requirement and the shaded area to its left is the shortfall probability. Those two areas are the complements of the two numbers above, 0.006

and 0.369; the first is too small to see at this scale, and the second occupies most of the left half of the lower panel. The figure is the first appearance of the theme that connects forecasting to operations: a decision needs a statement about a region of the distribution, usually a tail, and the mean is rarely that statement.

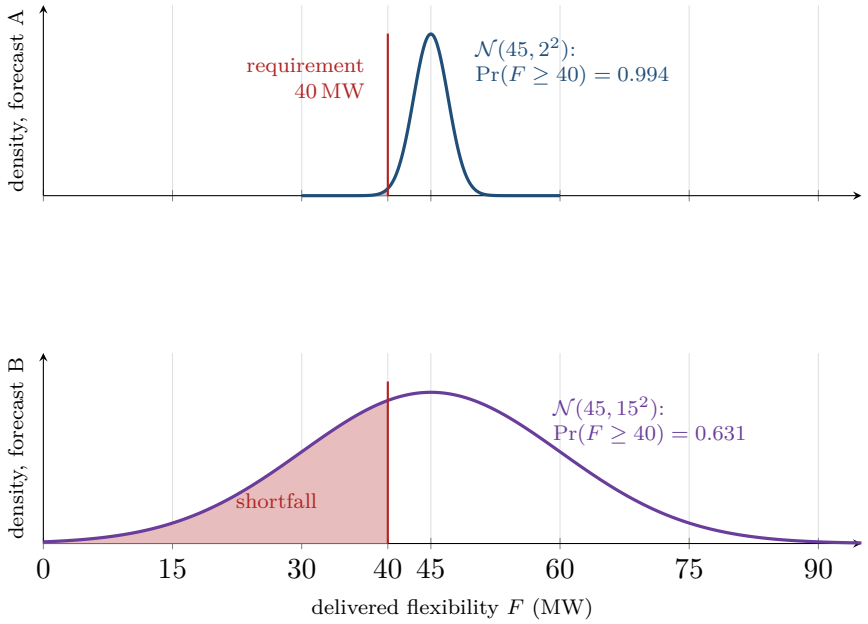


Figure 1.6: Two probabilistic forecasts with the same point forecast. Both densities are normal with mean 45 MW, as in equation (1.13); forecast A has $\sigma = 2$ MW and forecast B has $\sigma = 15$ MW. The vertical line is the 40 MW requirement and the shaded area is the shortfall probability, the complement of equation (1.14). The panels have separate vertical scales.

1.5.3 Conditioning on the planned control action

For the aggregator the conditioning set has one more element. Because two of the three targets depend on the intervention, the object to produce is

$$p(y_{t+h} \mid \mathcal{I}_t,\ u), \tag{1.15}$$

the distribution of the target given both the information and the planned action of equation (1.5). Reporting a single unconditional distribution for P^{ctrl} or F at a future hour is incomplete in the same way as reporting a point forecast for a wide distribution: it suppresses a dependence the user needs.

The addition of u is small in notation and large in consequence. A weather forecaster does not change tomorrow's weather, so the forecast is a statement about an exogenous process. An aggregator's flexibility forecast is a statement about a system its own organization is driving, and the forecast and the dispatch decision are parts of one problem. This is the point at which ordinary forecasting becomes forecasting for control, and Chapter 11 and the chapters beyond it return to it.

Principle 1.2. *For a controllable fleet the forecast is conditional on the planned action as well as on the information: the object is $p(y_{t+h} \mid \mathcal{I}_t, u)$. A forecast reported without the assumed action is not reproducible.*

1.6 The forecasting loop and a worked example

The pieces assembled above form an operating loop rather than a single calculation. This section draws the loop, works the chapter's numbers through it, and extracts the four questions that should be answered before any model is built.

1.6.1 From data to decision

A common mental model of forecasting is "historical load produces a forecast". Figure 1.7 replaces it with the structure this chapter has built. The boxes along the top row are the stages: the information set assembled at the issue time; the forecast model; the two target distributions $p(P^{\text{base}})$ and $p(P^{\text{ctrl}} \mid u)$; the flexibility distribution formed from them; and the optimization that consumes it. The inputs feeding the first box are listed beneath it, the planned action u enters the model in red because it is the aggregator's own choice, and the arrow returning from the decision to the fleet closes the loop: the dispatch changes the devices, which changes the next information set.

Two features of the diagram are worth naming. The forecast model produces distributions, not numbers, so the flexibility box is formed from a joint distribution rather than by subtracting two point values; Section 1.6.2 shows why that distinction has consequences. And the loop has a direction: information flows forward to a decision, and the decision flows back into the physical state. A forecasting system with no return arrow is a forecasting system nobody is acting on.

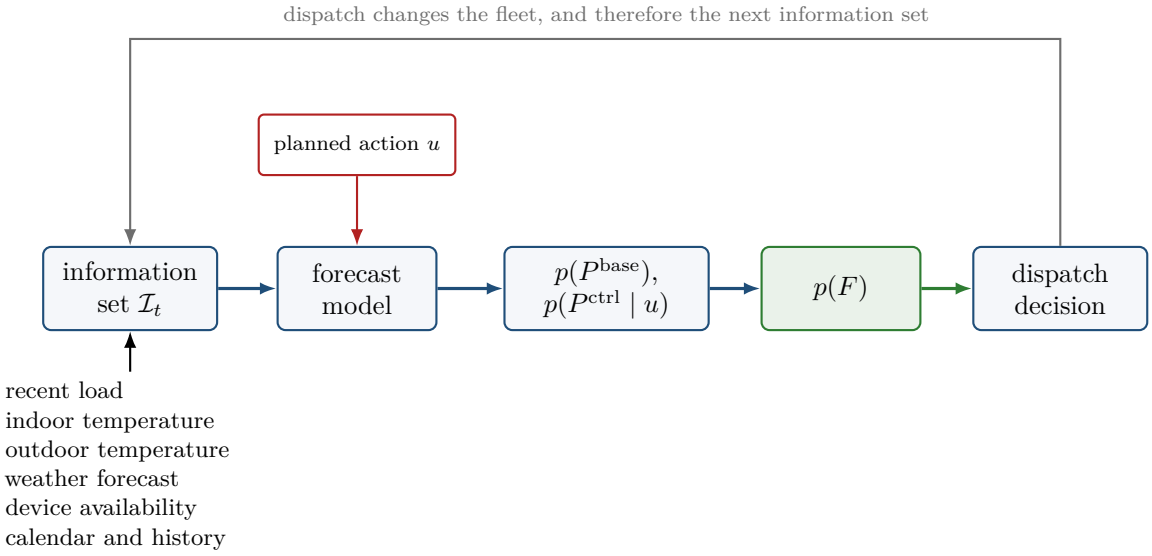


Figure 1.7: The forecasting loop for a controllable fleet. The stages run left to right from the information set to the dispatch decision; the inputs to the information set are listed below it, the aggregator's planned action enters the model from above, and the gray return path closes the loop through the physical fleet.

1.6.2 The chapter's numbers, worked through

Example 1.1 (The 2:00 PM question)**.** At 14:00 the aggregator holds the following information. The fleet is currently drawing $P_{\text{fleet}}(14{:}00) = 90\,\text{MW}$. The forecast outdoor temperature for 17:00 is $\hat{T}_{17:00|14:00} = 36°\text{C}$. Under comparable conditions the fleet's 5:00 PM load has historically been about 120 MW. The intended action is a $+2°\text{C}$ setpoint change, and the historical response of similar devices to such a change has ranged from 35 to 50 MW. A model built on this

information returns

$$P^{\text{base}}(17{:}00) \sim \mathcal{N}(120,\, 5^2), \qquad P^{\text{ctrl}}(17{:}00 \mid u) \sim \mathcal{N}(77,\, 7^2), \tag{1.16}$$

in MW, so that the expected flexibility is $\mathbb{E}[F] = 120 - 77 = 43\,\text{MW}$.

The mean of the difference is unambiguous. Its variance is not, and the naive calculation is wrong. Adding the variances, $5^2 + 7^2 = 74\ \text{MW}^2$, assumes that the two forecast errors are independent, and they are not: both P^{base} and P^{ctrl} are driven by the same forecast temperature, so a weather forecast that is two degrees too warm inflates both. In general

$$\text{Var}(F) = \sigma^2_{\text{base}} + \sigma^2_{\text{ctrl}} - 2\rho\,\sigma_{\text{base}}\sigma_{\text{ctrl}} = 74 - 70\rho \quad [\text{MW}^2], \tag{1.17}$$

with ρ the correlation between the two forecast errors. Three cases are instructive. Independence, $\rho = 0$, gives $\sigma_F = 8.6\,\text{MW}$. Strong common weather error, $\rho = 0.8$, gives $\sigma_F = 4.2\,\text{MW}$, less than half as wide. Perfect correlation, $\rho = 1$, gives $\sigma_F = |7 - 5| = 2.0\,\text{MW}$: the shared error cancels exactly in the difference and only the discrepancy between the two sensitivities survives.

Figure 1.8 shows the computation in two panels. The upper panel plots the two power densities of equation (1.16) against fleet power, with their means marked; the lower panel plots the resulting flexibility density under $\rho = 0$ and under $\rho = 0.8$ against flexibility, with the 40 MW requirement marked and the shortfall region of the wider case shaded. The two lower curves have the same mean, 43 MW, and the exceedance probability $\Pr(F \geq 40)$ is 0.64 in the independent case and 0.76 in the correlated case. The correlation assumption is therefore not a refinement; it moves the number the decision depends on. It reappears as a central issue in Chapter 11, where the errors being correlated belong to thousands of devices sharing one weather forecast.

1.6.3 Four questions before building a model

Everything in this chapter reduces to four questions, to be answered before a model class is considered.

1. What exactly is the target, including population, resolution, units and sign?

2. At what horizon, and from what issue time?

3. What exactly is known at that issue time, including latency and including which weather variables are forecasts rather than observations?

4. Which decision will consume the forecast, and therefore which feature of the distribution must be reported?

The questions are cheap and they prevent a specific class of failure: a system that is accurate against its own test set and useless in operation. The two most common instances have both appeared above. Question 3 catches the leakage of Section 1.4.3, in which a model was fitted to a variable that will not exist at issue time. Question 1 catches the mixture of Section 1.2.4, in which a target was assembled from events that are not comparable. Question 4 is the one most often skipped, and it is the reason a well-calibrated distribution is worth more than a slightly better mean.

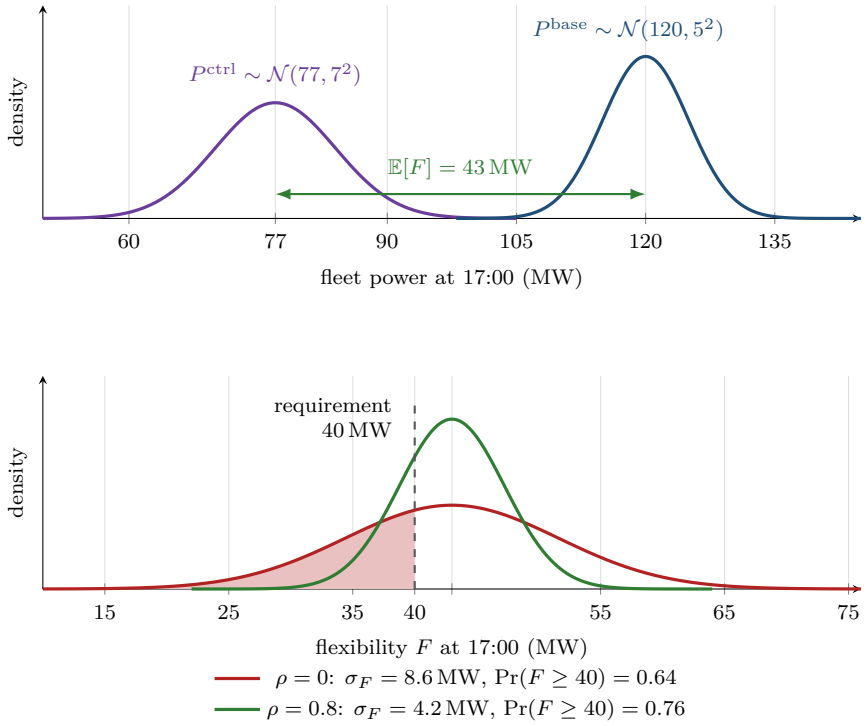


Figure 1.8: The worked example of Section 1.6.2. Upper panel: the two power densities of equation (1.16) against fleet power at 17:00, with the gap between their means labeled. Lower panel: the density of their difference under two correlation assumptions, from equation (1.17), with the 40 MW requirement dashed and the shortfall region of the independent case shaded. All densities are normal with the stated parameters.

Conceptual questions

"We want to forecast thermostat flexibility three hours ahead. How would you start?"

A strong answer does not begin with a model. It begins by defining the target: whether the quantity wanted is the baseline load, the controlled load under a specified dispatch action, or the difference between them, and it notes that the third is a counterfactual that will have to be estimated rather than measured. It then fixes the issue time and verifies that every intended input genuinely exists at that moment, separating observed weather from the forecast that will actually be available and accounting for telemetry latency. Next it establishes simple reference forecasts — persistence, a similar-day average, a temperature regression — so that any later model has something to beat. Only then does it compare more sophisticated models, and it reports the comparison at the horizon asked for rather than at a convenient one. A good answer also asks which decision the number serves, because a capacity commitment needs a lower quantile rather than a mean.

1.7 What to retain

Definition 1.1 (The forecasting problem)**.** The fundamental object of forecasting is not a value y_t but the conditional distribution $p(y_{t+h} \mid \mathcal{I}_t)$, in which t is the issue time, h the horizon, y_{t+h} the named target and $\mathcal{I}_t$ the information genuinely available at t. For a controllable fleet the target may depend on the operator's own action, so the object is $p(y_{t+h} \mid \mathcal{I}_t, u)$; that addition of u is what turns ordinary forecasting into forecasting for control.

Table 1.3 collects the chapter's objects with the subsection that defines each.

Table 1.3: The objects of a forecasting problem specification.

Object	Definition	Subsection
t, h	issue time and horizon; target time $t + h$	1.3.1
y_{t+h}	the target: a named quantity, population, resolution and sign	1.2.4
$\mathcal{I}_t$	information set: measured history, device state, exogenous forecasts, calendar	1.4.1
$\hat{y}_{t+h} = f(\mathcal{I}_t)$	point forecast: one summary of the distribution	1.5.1
$p(y_{t+h} \mid \mathcal{I}_t)$	probabilistic forecast: the complete object	1.1.3
$P^{\text{base}}(t)$	baseline load: consumption without intervention	1.2.1
$P^{\text{ctrl}}(t \mid u)$	controlled load under the action u of equation (1.5)	1.2.2
$F(t)$	flexibility $P^{\text{base}} - P^{\text{ctrl}}$; counterfactual, never measured	1.2.3
$\hat{T}_{t+h\mid t}$	weather forecast for $t + h$ issued at t; not the observed value	1.4.2
leakage	use of anything outside $\mathcal{I}_t$ in training or evaluation	1.4.3
δ	telemetry latency; the information set ends at $t - \delta$	1.4.4
$\Pr(F \geq F_0)$	the decision-relevant functional when a threshold must be met	1.5.2
$p(y_{t+h} \mid \mathcal{I}_t, u)$	the forecast for a controllable fleet	1.5.3

1.8 Exercises

Exercise 1.1. An operator asks three questions about the same thermostat fleet at 18:00 tomorrow: (a) "how much will these customers consume?"; (b) "if we raise setpoints by three degrees for ninety minutes, what will they consume?"; (c) "how much can we offer the utility?". Name which of P^{base}, $P^{\text{ctrl}}(\cdot \mid u)$ and F each question asks for, and state for each whether the quantity can be verified against a meter after the fact. For the one that cannot, say what must be supplied in its place.

Exercise 1.2. A load model fitted on observed temperatures reports a root-mean-square error of 2.5 MW. The fleet's sensitivity is $\partial P/\partial T = 4\,\text{MW}$ per degree Celsius and the three-hour-ahead temperature forecast has $\sigma_T = 1.5°\text{C}$. Using equation (1.10), compute the operational error at $h = 3$ hours. Repeat for a day-ahead forecast with $\sigma_T = 2.5°\text{C}$. In which of the two cases is the load model itself the limiting factor?

Exercise 1.3. For the two forecasts of equation (1.13), compute the largest reduction the aggregator could commit to while keeping $\Pr(F \geq \text{commitment}) \geq 0.95$. State both answers, compare them with the shared point forecast of 45 MW, and say what the difference is worth if committed capacity is paid \$40 per MW.

Exercise 1.4. Using equation (1.17), find the correlation ρ for which $\sigma_F = 3\,\text{MW}$, and compute $\Pr(F \geq 40\,\text{MW})$ at that value. Then explain, in terms of the shared weather forecast, why a *positive* correlation between the two forecast errors makes the flexibility forecast narrower rather than wider, and what would have to be true for ρ to be negative.

Chapter 2

Structure of energy time series: level, seasonality, weather dependence and persistence

2.1 The decomposition of a load series

Chapter 1 fixed the forecasting problem: at issue time t, using only the information set $\mathcal{I}_t$, produce $p(y_{t+h} \mid \mathcal{I}_t)$ for a target h hours ahead. Nothing in that statement promises the target is predictable. Whether it is, and by how much, depends on the structure of the series, and this chapter's question is

What is an hourly energy series made of, and which of its parts can be predicted?

A fleet load series is not an arbitrary sequence of numbers. It is the measured output of a large physical system whose inputs are weather, household schedules, device states and, for an aggregator, its own control commands. Each of those inputs leaves a recognizable signature in the series: a level, repeating daily and weekly cycles, a kinked dependence on outdoor temperature, and a strong dependence of each hour on the hours before it. Those signatures are what every forecasting method in this book exploits, and naming them precisely is a prerequisite for choosing a method at all.

Jargon: level, trend, seasonality, autocorrelation, stationarity, balance temperature

The *level* of a series is its local average, the number the series fluctuates about. A *trend* is a slow change in that level over months or years. *Seasonality* is any cycle of fixed period: electricity load has a 24-hour cycle, a 168-hour (weekly) cycle and an annual cycle, so "seasonal" in this book means "periodic", not "to do with summer". *Autocorrelation* is the correlation of a series with its own past, $\mathrm{Corr}(P_t, P_{t-k})$ at lag k. A series is *stationary* when its statistical properties do not depend on the time of observation; load is not stationary, and much of forecasting practice consists of removing the parts that make it nonstationary. An *exogenous variable* is a predictor that comes from outside the series, principally weather and calendar. The *balance temperature* (also called the balance point or base temperature) is the outdoor temperature at which a building needs neither heating nor cooling; below and above it the load responds to weather, between them it does not. *Common-mode uncertainty* is error shared by every device in a fleet, such as a single wrong temperature forecast for the whole territory; averaging over devices does not reduce it.

Key idea

An energy series is the output of a physical system driven by weather, by calendar-locked human behavior and by control. Its level, cycles, weather response and persistence are signatures of those mechanisms rather than statistical curiosities, and that physical origin is why the series is forecastable at all.

This section writes the series as a sum of parts, restates the same object in state-space form so that the persistence of Section 2.4 has somewhere to live, and decomposes one synthetic week of fleet load into its four components.

2.1.1 From causes to components

Write the fleet load at hour t as a function of its causes,

$$P_t = f(x_t,\ w_t,\ b_t,\ c_t,\ u_t) + \varepsilon_t, \tag{2.1}$$

with x_t the physical state of the fleet (indoor temperatures, stored heat in the building fabric, which compressors are running, how many devices are connected), w_t the weather, b_t occupant behavior, c_t the calendar, u_t the aggregator's control action, and ε_t everything left over. Equation (2.1) is not a model that can be fitted; it is a statement about where the numbers come from, and it already contains the chapter's agenda, because each argument produces a different kind of structure.

For working purposes the function is usually replaced by a sum of components,

$$P_t = \underbrace{L_t}_{\text{level and trend}} + \underbrace{S_t}_{\text{calendar}} + \underbrace{W_t}_{\text{weather response}} + \underbrace{C_t}_{\text{control}} + \underbrace{\varepsilon_t}_{\text{residual}}, \tag{2.2}$$

which is the classical additive decomposition with one term added for the aggregator's own intervention. Three properties of equation (2.2) should be stated at once, because all three matter later.

The decomposition is a modeling choice, not a property of the data. Load could equally be written multiplicatively, $P_t = L_t S_t W_t \varepsilon_t$, which is often a better description of long-horizon growth, since a fleet twice the size has roughly twice the daily swing as well as twice the level. The additive form is used here because the components can then be plotted on the same axis and read in megawatts.

The components are not orthogonal. Nothing forces S_t and W_t to be uncorrelated, and in practice they are strongly correlated: afternoons are both the hot part of the day and the part of the day when the hour-of-day profile is high. The consequence is that the variances of the components do not add up to the variance of the series, and that "how much of the variance is weather" has no unique answer. A number for it is only meaningful once the decomposition that produced it is stated.

The residual is defined by what the other terms contain. ε_t is not a property of the physical world. It is whatever the other four terms failed to represent, so it shrinks as the model improves and, importantly, it is usually not white noise. The residual of a calendar-and-weather model retains autocorrelation, which is exactly the structure Chapter 6 models.

2.1.2 The state-space viewpoint

Equation (2.2) has no memory in it: every term is a function of the current hour. That cannot be right for a fleet of buildings, because a building that was hot an hour ago is still warm now. The compact way to carry memory is the state-space pair

$$x_{t+1} = g(x_t,\ u_t,\ w_t), \qquad P_t = h(x_t) + \varepsilon_t, \tag{2.3}$$

in which g is the dynamics, h is the measurement, and the state x_t is the only channel through which the past reaches the present. For a thermostat fleet the state includes quantities a forecaster never measures directly: the distribution of indoor temperatures across the fleet, the heat stored in walls and floors, the on/off phase of each compressor, and the number of devices currently reachable. Figure 2.1 draws this structure.

The figure is read left to right. The four boxes on the left are the exogenous drivers of equation (2.1): weather, calendar, occupant behavior, and the control action, drawn in red because it is the only input the aggregator chooses. They feed the large central box, the latent state x_t, whose contents are listed inside it. The curved arrow above that box is the dynamics g: the state at the next hour depends on the state at this one, which is where the series' memory physically resides. The arrow on the right is the measurement h, which maps state to metered power, with the noise ε_t entering at the meter. The dashed path along the bottom is the part of household load that reaches the meter without passing through the thermal state at all: lighting, cooking, consumer electronics. That path has calendar structure but almost no thermal memory, which is why a fleet series shows both fast and slow dependence on its own past.

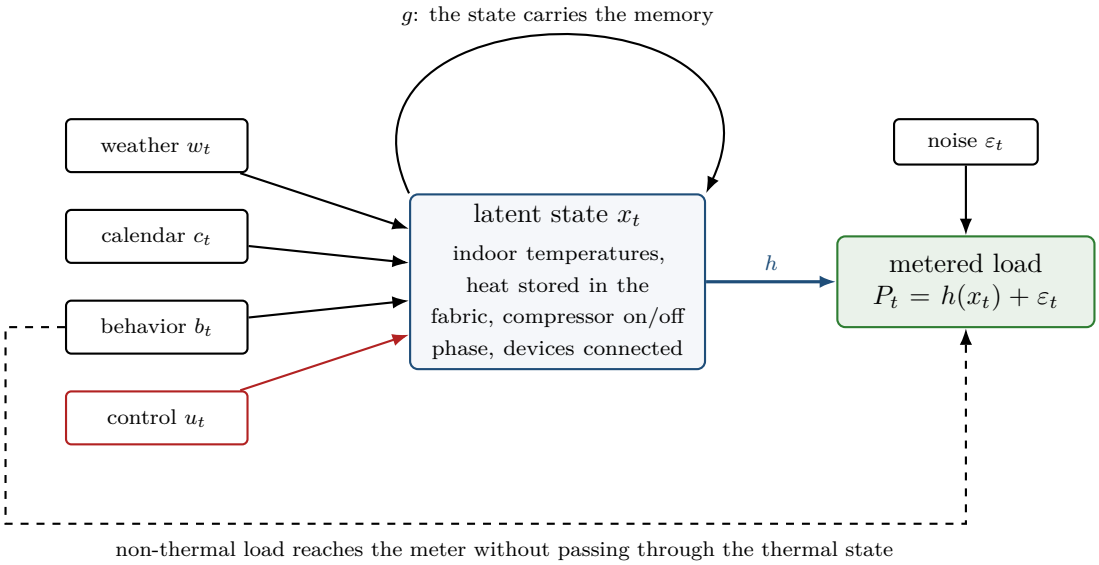


Figure 2.1: The generative structure of a fleet load series. Exogenous drivers (left) act on a latent physical state (center) whose own dynamics g carry memory from hour to hour; the measurement h turns state into metered power. The red input is the only one the aggregator chooses. The dashed path is non-thermal household load, which has calendar structure but no thermal memory.

The state-space form also explains a fact that would otherwise look like a coincidence: recent observations of load are useful out of proportion to their apparent information content. A single number, P_t, is useful because it is a measurement of $h(x_t)$, and therefore evidence about the hidden state that will generate the next several hours. Persistence, in other words, is a crude state estimator. That reading is developed in Section 2.4 and is the reason Chapter 7 can replace hand-built lag features with a learned hidden state.

END OF THE PREVIEW

Time-Series Forecasting for Energy

These are the opening pages. The complete book runs 363 pages.

The full PDF opens with a token. Write to **rmarquez123@gmail.com**, name the book, and I will send you one.

One token opens every book on the site.

Ricardo Marquez, Ph.D.

ricardo-marquez.com