

Asset Management by Derivation

From the asset manager's questions to probabilistic inference and decisions

Draft · 465 pages

Ricardo Marquez, Ph.D.

ricardo-marquez.com

Asset Management by Derivation

Volume I

From the asset manager's questions to probabilistic inference and decisions

Ricardo Marquez

November 1, 2026 (work in progress: chapters are added as they are written)

How to read this book

This book is written for the asset manager, the maintenance or reliability planner and the condition-monitoring engineer. Its questions are the ones an asset's life raises: what condition it is in today, how fast it is changing, when it will need attention, what to do about it and when, and what the monitoring itself is worth. A trend chart answers them from raw performance: the approach temperature, the return temperature, the efficiency, plotted month by month with a line through it. Raw performance mixes aging with load and weather, so the chart reads a summer as degradation and a light load as a recovery. It cannot say when the coil will no longer hold the hall, whether it will last to the planned shutdown, or whether restoring it is worth more than replacing it. Each of these is a probability or an expected consequence over the unknowns of one model of the plant, and the book derives and computes each one.

The health index in this book is a parameter of the physics: a coil's conductance, a battery's capacity and internal resistance, read against the value the commissioning test found. Load and weather enter the model as known inputs, so the parameter's trend is aging and nothing else. Time enters as rows that tie one month's parameter to the next, so a trend is more rows in the same fit rather than a second tool, and every date and every decision is read from the posterior that fit produces.

The plant grows with the book. Part I works on one equipment hall, a state of two and then three entries, so that every object of the method can be seen whole. From Part II the hall has two cooling coils and three racks, and it is followed from its commissioning test through three years of monthly records. Two smaller assets join it: the site's UPS battery string and a fleet of fifty coils of the same design.

Every chapter that answers a question follows the same progression, and the progression is the method.

Step 0: the plant graph. The asset and the plant around it, before any question is asked: its nodes, paths and sources; which quantities are pinned and which are free; the physical relations that must hold among them. The free quantities form the state vector $\mathbf{z}$.

Step 1: the question. The asset manager's question in their own words, then as a probability or an expectation of a function of the state, with the unknown on the left of the conditioning bar and everything treated as known on the right.

Step 2: the graph extended for the question. The windows of the record and the trend rows that tie them; candidate explanations when something may have changed; the actions when a decision is asked; and the records attached as rows.

Step 3: the posterior of the state. Derived and computed once, before any question is answered: the MAP state, the information, the covariance, and the exact posterior where one exists.

Step 4: the answers. Each question read from that posterior down to its number: an entry of the state or the gradient of a derived quantity; a trend and the date it crosses a limit; a ratio of evidences when explanations compete; or an expected consequence when actions do.

Step 5: evaluation and checks. The gaps between the reported values and the exact values or draws collected, the approximations tested and, because the record is constructed from a stated truth, the answers tested against that truth.

Step 6: the report. The date, the probability, the band or the action the asset manager acts on, with the exact value beside it where one exists.

Step 7: the next measurement. If the band is too wide, the measurements that could be added ranked by the width each would remove from the answer that matters.

Five kinds of question cover what the book asks, distinguished by the mathematical object that answers them rather than by the word the maintenance office uses: *estimation*, *prediction under a threshold*, *hypothesis comparison*, *diagnosis* and *design*. Beneath them sit five computations: the *posterior* of the state, the *trend* of a parameter across windows, the *evidence* of a drawing of the plant, the *preposterior* after a measurement not yet taken, and the *expected consequence* of an action. Asset management is mostly trend and expected consequence. The notation chapter that follows fixes the symbols, and the appendix collects the asset manager's questions in an atlas that maps each to its mathematical form, its computation and the section that answers it.

The framework the book teaches is physics-constrained probabilistic modeling: a plant written as a graph, its unknowns as a state vector, every reading as a row, and every question as a probability over that state. TERRA is the author's own implementation of this framework; the book names the framework and the computations, not the software, and every number in it is reproducible from the derivations as written. The method and its notation are those of the companion books *Reliability Engineering by Derivation* and *Operations and Control by Derivation*, which ask the reliability engineer's and the operator's questions of the same hall.

Each chapter closes with a statement of what to retain, a table of its terms, and exercises at three levels, formulation, derivation and application, each followed by its answer.

Contents

How to read this book **ii**

Notation and vocabulary **viii**

I Thinking in Probabilities **1**

1 Physical assets as graphs **2**
1.1 One asset and several questions . . . 2
1.2 The posterior of the state (step 3) . . . 15
1.3 The answers (step 4) . . . 21
1.4 The checks (step 5) . . . 27
1.5 The report (step 6) . . . 29
1.6 The next measurement (step 7) . . . 31
1.7 What to retain . . . 35
1.8 Exercises . . . 38

2 The language of asset-management questions **40**
2.1 Asset-management questions and their mathematical form . . . 40
2.2 The grammar of probability statements . . . 43
2.3 The general forms and the five computations . . . 45
2.4 Estimation, prediction and joint events . . . 59
2.5 Model comparison, diagnosis and decisions . . . 70
2.6 Measurements, classification and workflow . . . 74
2.7 What to retain . . . 80
2.8 Exercises . . . 83

3 Uncertainty and conditional probability **86**
3.1 Uncertainty and evidence in asset-management questions . . . 86
3.2 Conditional probability and Bayes' theorem . . . 92
3.3 Bayesian diagnosis of fouling from a condition test . . . 94
3.4 Marginalization and dependence . . . 97
3.5 Conditioning in a physical model . . . 101
3.6 Predictive variance and the posterior covariance . . . 110
3.7 What to retain . . . 117
3.8 Exercises . . . 120

4 The anatomy of probabilistic inference **122**
4.1 From a plant graph to an acceptance probability . . . 122
4.2 The posterior of the state (step 3) . . . 131
4.3 The answers (step 4) . . . 144

4.4 The checks (step 5) . . . 149
4.5 The report (step 6) . . . 155
4.6 The next measurement (step 7) . . . 156
4.7 What to retain . . . 167
4.8 Exercises . . . 170

II Assessing Present Condition 173

5 Health indices from physical parameters 174
5.1 The dashboard's health index and the coil's . . . 174
5.2 The posterior of the state (step 3) . . . 180
5.3 The answers (step 4) . . . 185
5.4 The checks (step 5) . . . 192
5.5 The report (step 6) . . . 195
5.6 The next measurement (step 7) . . . 196
5.7 What to retain . . . 199
5.8 Exercises . . . 200

6 Hidden degradation under control 202
6.1 The held set point and the effort behind it . . . 202
6.2 The posterior of the state (step 3) . . . 208
6.3 The answers (step 4) . . . 210
6.4 The checks (step 5) . . . 216
6.5 The report (step 6) . . . 219
6.6 The next measurement (step 7) . . . 220
6.7 What to retain . . . 223
6.8 Exercises . . . 224

7 Degradation or sensor drift 226
7.1 A rise in the approach with two explanations . . . 226
7.2 The posterior of the state (step 3) . . . 233
7.3 The answers (step 4) . . . 237
7.4 The checks (step 5) . . . 246
7.5 The report (step 6) . . . 250
7.6 The next measurement (step 7) . . . 251
7.7 What to retain . . . 255
7.8 Exercises . . . 256

8 Battery state of health 258
8.1 The state-of-health tile and the string's circuit . . . 258
8.2 The posterior of the state (step 3) . . . 265
8.3 The answers (step 4) . . . 270
8.4 The checks (step 5) . . . 276
8.5 The report (step 6) . . . 280
8.6 The next measurement (step 7) . . . 281
8.7 What to retain . . . 285
8.8 Exercises . . . 286

9 Fleet pooling and hierarchical models 288
9.1 The fleet chart and the coil's place in it . . . 288
9.2 The posterior of the state (step 3) . . . 294
9.3 The answers (step 4) . . . 301
9.4 The checks (step 5) . . . 306
9.5 The report (step 6) . . . 309
9.6 The next measurement (step 7) . . . 310
9.7 What to retain . . . 314
9.8 Exercises . . . 315

III Estimating Degradation Rates 316

10 Degradation rates from condition records 317
10.1 The dashboard's trend line and the coil's rate . . . 317
10.2 The posterior of the state (step 3) . . . 323
10.3 The answers (step 4) . . . 327
10.4 The checks (step 5) . . . 336
10.5 The report (step 6) . . . 342
10.6 The next measurement (step 7) . . . 344
10.7 What to retain . . . 347
10.8 Exercises . . . 348

11 Accelerating degradation 350
11.1 A bend in the trend chart . . . 350
11.2 The posterior of the state (step 3) . . . 356
11.3 The answers (step 4) . . . 359
11.4 The checks (step 5) . . . 367
11.5 The report (step 6) . . . 373
11.6 The next measurement (step 7) . . . 374
11.7 What to retain . . . 378
11.8 Exercises . . . 379

12 Change-point detection 381
12.1 A step on the dashboard and a pump change . . . 381
12.2 The posterior of the state (step 3) . . . 386
12.3 The answers (step 4) . . . 389
12.4 The checks (step 5) . . . 397
12.5 The report (step 6) . . . 403
12.6 The next measurement (step 7) . . . 404
12.7 What to retain . . . 408
12.8 Exercises . . . 409

13 Seasonal effects and aging 411
13.1 The summer drop on the dashboard . . . 411
13.2 The posterior of the state (step 3) . . . 418
13.3 The answers (step 4) . . . 419
13.4 The checks (step 5) . . . 431
13.5 The report (step 6) . . . 434
13.6 The next measurement (step 7) . . . 436
13.7 What to retain . . . 439
13.8 Exercises . . . 440

A Atlas of asset-management questions **442**

Notation and vocabulary

Every symbol below is used with the same meaning in every chapter. The unknown always stands on the left of a conditioning bar and the evidence on the right. Vectors and matrices are bold. Coil subscripts are $u \in \{1, 2\}$, rack subscripts $j \in \{1, 2, 3\}$ and fleet subscripts $i = 0, \ldots, 49$; windows are months, indexed t.

Symbol	Unit	Meaning
The plant graph and its state		
$\mathbf{z}$, z_i	varies	State vector of the plant graph: every free quantity estimated at once; its i-th entry.
$\mathbf{z}_t$, t	varies, –	The state of window t; the window index, $t = 0$ the commissioning test and $t = 1, \ldots, 36$ the months from January of year 1 to December of year 3.
$\mathbf{y}$, y	varies	The evidence: all readings, test results, counts and prior information treated as known; a single reading, subscripted by what it reads.
$r_k(\mathbf{z})$	varies	Row k of the model: a physical relation that must hold ($r_k = 0$), a reading compared with the state, a trend row, or a prior. As a subscript k is this row (and in Chapter 24 a channel's rows); in Chapters 2, 3 and 15, whose bare k is the count of failures, the row is ℓ.
w_k, σ_k	varies	Weight (precision) of row k and the standard deviation it corresponds to, $w_k = 1/\sigma_k^2$.
σ_p	varies	Tolerance of a physics row; an exact constraint is the limit $\sigma_p \to 0$.
The equipment hall of Part I		
T, T_c	°C	Temperature of the hall air; the supply air held by the cooling unit's discharge controller, pinned.
Q, G	kW, kW/K	Heat load of the hall; effective conductance of its air path to the supply air.
g, G_{ref}	–, kW/K	Log-conductance $\ln(G/G_{\text{ref}})$, the entry of the state that carries the conductance; the design conductance on the design sheet, 6 kW/K.
h, h^*, h_F	–	Part I's health index $h = G/G_{\text{ref}} = \mathrm{e}^g$, the conductance against design; the index a requirement needs; the functional limit, at which the asset stops doing its job. From Part II the health index is against commissioning, $h_{u,t}$ below.
μ_g, σ_g	–	Mean and spread of the handbook prior on g, $\mu_g = \ln(7.5/6)$.

Symbol	Unit	Meaning
g_t, $\hat{g}_0$	–	The log-conductance in window t; the commissioned value, the fit of the commissioning test (window 0).
$\sigma_{\hat{g}_0}$, $\hat{G}_0$	–, kW/K	Spread of the commissioning row on g (the commissioning fit carried forward); the commissioned conductance $G_{\text{ref}}\text{e}^{\hat{g}_0}$.
G_{test}, h_{test}	kW/K, –	The conductance and index one condition test reports when its reading is taken as exact, $Q/(y_T - T_c)$.
g^*, g_F	–	A requirement level on g, $\ln h^*$; from Chapter 19, the level on each coil's log-derate at which two equal coils give G^*, $\ln(G^*/(\bar{G}_1 + \bar{G}_2)) = -0.26313$; the functional limit, $\ln h_F$, in Chapter 20 $\ln(5/6) = -0.18232$ on each fleet coil's log-derate against rating.
$T_{\max}$, T^*	°C	A temperature limit: 24°C now and 27.5°C at 60 kW (the owner's request, Chapter 1), 28°C (the commissioning record's acceptance limit, Chapter 4); a threshold on a test temperature.
Q_2, T_2	kW, °C	The requested load and the hall temperature it produces, a scenario node.
ρ, μ_0, σ_0	K/kW	Thermal resistance of the air path, $1/G$; the commissioning record's pre-test estimate of it and its spread (Chapter 4).
b_T	K	Calibration offset of the hall sensor, a nuisance quantity or a bias entry.
Δ_{m}, σ_{m}	kW	A model discrepancy of the physics row and its spread (Chapter 1).
C_{th}, ϑ, ϑ_{th}, ΔT	kJ/K, s or h, s, K	Thermal capacitance of the hall; elapsed time within a record (seconds in Chapter 1's transient, hours into a battery test in Chapter 8, hours within a window in Chapter 24); the hall's time constant C_{th}/G; the departure from steady state.
y_T, σ_T	°C, K	The hall sensor's reading and the standard deviation of its error.
The two-unit hall of Parts II–VI		
$T_s(t)$, $Q_o(t)$	°C, kW	Chilled-water supply temperature and the hall's other heat in window t: known inputs that follow the season.
Q	kW	IT heat load of the hall, an entry of the state read by the hall load meter.
G_u, $\bar{G}_u$	kW/K	Effective conductance of coil u; its rated value, 7 kW/K.
g_u, d_u	–	Log-derate $\ln(G_u/\bar{G}_u)$, the entry that carries coil u's conductance against its rating; the derate factor $d_u = \text{e}^{g_u}$, one at the rating.
$g_{u,t}$, $d_{u,t}$, Q_t	–, –, kW	The log-derate, derate factor and IT load in window t.
$h_{u,t}$	–	Health index of coil u in window t against its commissioned value, $G_{u,t}/G_{u,0} = \text{e}^{g_{u,t}-g_{u,0}}$ (Part II on).
$A_{u,t}$, h_u^{raw}	K, –	Approach of unit u in window t, $T_{d,u} - T_s(t)$; the dashboard's raw-performance index $A_{u,0}/A_{u,t}$, the naive baseline.
$G_{u,\text{test}}$	kW/K	Conductance of coil u with the window's readings taken as exact.

Symbol	Unit	Meaning
σ_W	kW	Standard deviation of a water-side heat meter's reading on a coil's duty.
$g_{u,0}$	–	The commissioned value of g_u, the fit of window 0.
C_u	kW/K	Air capacity rate of unit u, pinned at its fixed fan speed; read by an airflow station only in Chapter 23's candidates.
T_r, T_c, $T_{d,u}$	°C	Return air, cold aisle (the supply air of Part I's hall, now made by two coils), and discharge air of unit u.
T_j, β_j	°C, –	Inlet temperature of rack j; its recirculation fraction.
G^*, f^*	varies	A requirement level: G^* the conductance a duty needs (in Part I, 5.0 kW/K for 24°C at 30 kW and 6.316 kW/K for 27.5°C at 60 kW; from Part II, the total $G_1 + G_2$ of the capacity scenario); f^* a level on any derived quantity. The requirement holds on the safe side of f^*: $f \geq f^*$ for a conductance, $f \leq f^*$ for a temperature.
b_j	varies	Bias of reading j, an entry freed when a sensor is suspected of drifting.
$\check{X}$	as X	The value of X at the scenario node of the commissioned model: today's load and inputs with the commissioned coils ($\check{T}_r$, $\check{D}_u$, $\check{A}_u$; Chapter 6).
$D_{u,t}$	kW	Duty of unit u in window t, the heat its coil passes to the chilled water, $G_{u,t}(T_{r,t} - T_s(t))$.
$\eta_{A,u}$, $\eta_{D,u}$, η_s	K, kW, K	Excess approach, duty shift and supply penalty against the commissioned model at today's conditions (Chapter 6).
$A_{r,t}$, T_c^{sp}	K, °C	Return approach $T_{r,t} - T_s(t)$; a cold-aisle set point.
$G_{\mathrm{tot},t}$, G_t^{app}, $\bar{Q}_o$	kW/K, kW/K, kW	The hall's total conductance $G_{1,t} + G_{2,t}$; its apparent conductance, the metered IT load over the return approach, $y_{Q,t}/(y_{T_r,t} - T_s(t))$, a raw figure; the mean other heat, 8.00 kW (Chapter 13).
ζ_0, s_{season}, s_{load}	K/kW, K, K	The cold aisle's rise per kW at the commissioned coils; the season's and the load's terms of the cold aisle's change since commissioning; from Chapter 13, the season's and the load's terms of a raw figure's change between any two windows.
$b_{d,1}$, σ_b, $\sigma_{\ln h}$	K, K, –	Drift of unit 1's discharge sensor since commissioning (a bias entry); the prior spreads of a freed bias and of a freed change $\ln h_{u,t}$ (Chapter 7). From Chapter 10 the drift is a function of the window, $b_{d,1}(t)$.
$\dot{b}_{d,1}$, t_b, $\sigma_{\dot{b}}$	K/yr, window, K/yr	Drift rate of unit 1's discharge sensor, one entry for the record, with $b_{d,1}(t) = \dot{b}_{d,1}\,(t - t_b + 1)\,\Delta t$ for $t \geq t_b$ and zero before; the drift's first window, $t_b = 20$; the drift rate's prior spread, 2.0 (Chapter 10).
M_{neither}, M_{foul}, M_{drift}, M_{both}	–	The four drawings of Chapter 7: neither term, the derate term, the bias term, both.
K_{kk}, r_k^{N}, n_r	–	Leverage of row k, the diagonal of the hat matrix $\mathbf{K} = \mathbf{J}\boldsymbol{\Sigma}\mathbf{J}^{\mathsf{T}}$, with redundancy number $1 - K_{kk}$; the normalized residual $r_k/\sqrt{1 - K_{kk}}$; the number of rows.

Symbol	Unit	Meaning
ψ_k, $s_{d,1}$, y_b	–, K, K	The ratio in the closed-form Bayes factor of a bias on reading k (Chapter 12's ψ extends it to a step's column); the shift of unit 1's discharge reading the two suspects make; a side-by-side check's reading of a bias.
Time, trend and dates		
Δt	yr	The spacing of two windows, one month.
γ, γ_u, $\dot{\gamma}_u$	yr^{-1}, yr^{-1}, yr^{-2}	Trend (degradation rate) of a parameter; coil u's rate, the trend entry of its log-derate path, read as $100(\mathrm{e}^{\gamma_u} - 1)\,\%/\mathrm{yr}$ (Chapter 10), and in a drawing that frees the curvature the rate at the record's center $\bar{t}_\mathrm{y}$; its curvature (acceleration), the change of the rate per year (Chapter 11). Chapter 9's γ_i indexes the fleet's coils instead: its γ_1 is fleet coil 1, not the hall's coil 1.
t_y, $\bar{t}_\mathrm{y}$, $\gamma_u(t_\mathrm{y})$	yr, yr, yr^{-1}	Time since commissioning, $t_\mathrm{y} = t\,\Delta t$ (Chapter 11); the record's center, 1.5 yr (window 18); coil u's rate at time t_y, $\gamma_u + \dot{\gamma}_u(t_\mathrm{y} - \bar{t}_\mathrm{y})$, so that $\gamma_u(3)$ is the rate today.
x_t, $\dot{g}_{u,t}$, $\sigma_{\dot{\gamma}}$	yr, yr^{-1}, yr^{-2}	The curvature's lever on the increment from window t to $t+1$, $(t+\frac{1}{2})\Delta t - \bar{t}_\mathrm{y}$; the path's monthly rate $(g_{u,t+1} - g_{u,t})/\Delta t$; the curvature's prior spread, 0.05 (Chapter 11).
t_knee	window	The knee of a rate, after which it changes by $\dot{\gamma}_u$ per year: 12 or 24 in Chapter 11's sensitivity forms.
Γ	yr^{-1}	The hall's conductance rate, $\ln(G_{\mathrm{tot},36}/G_{\mathrm{tot},0})/(36\,\Delta t)$ (Chapter 13). With an argument, $\Gamma(\cdot)$ is the gamma function (Chapter 2).
α	–	Level entry of a straight-line trend at the record's center $\bar{t}$.
σ_q	–	Spread of a trend row: one monthly window's departure of a parameter from its straight line, $\sigma_q\sqrt{n_\mathrm{w}}$ across n_w windows without a reading; the straight line is the limit $\sigma_q \to 0$.
$r_{q,u,t}$, $\mathbf{D}_t$, $\mathbf{E}$, $\mathbf{F}$	–, varies	Coil u's trend row from window t to $t+1$; the trend rows' blocks of the stacked Jacobian in window t's columns and in the trend entries (γ_1, γ_2) (Chapter 10), and in the curvatures (Chapter 11).
n_w, n_z, t_as	–, –, window	The number of monthly increments a rate is read over, 36 for the whole record, or of unread windows a trend row spans (Chapters 4 and 10); the number of entries of the state (Chapters 3, 4, 7 and 12); the last window of an as-of fit, the record cut at window t_as.
s_u^raw, $\dot{s}_u^\mathrm{raw}$	yr^{-1}, yr^{-2}	Slope of the dashboard's least-squares line through the 37 tiles $h_{u,t}^\mathrm{raw}$ (Chapter 10); the dashboard's bend, twice the quadratic coefficient of the least-squares quadratic through them (Chapter 11).
$\tilde{\mathbf{z}}^0$, $\tilde{\mathbf{z}}^g$	varies	The stacked state with every window's coils at the commissioned values and no drift; with the fitted coils and no drift: the steps of Chapter 10's sequential split of the dashboard's slope.

Symbol	Unit	Meaning
γ_u°	yr^{-1}	Pseudo-true value of γ_u: the drawing's fit to the truth's noise-free readings (step 5, Chapter 10).
M_0, M_1, M_2, M_{12}	–	The constant-rate drawing (Chapter 10's base fit) and the drawings that free coil 1's, coil 2's and both curvatures (Chapter 11). Chapter 12 writes the first M_{none}.
δ, δ_j, t_0	–; –, K or kW; window	Size of a step change in a parameter (with an argument, $\delta(\cdot)$ is the Dirac delta, Chapter 3); a step of kind j (Chapter 12): δ_1, δ_2 a coil's, δ_{12} both coils', δ_{T_r}, $\delta_{T_{d,1}}$, $\delta_{T_{d,2}}$, δ_{T_1} a temperature sensor's, δ_Q the load meter's; the window at which it occurs.
s, σ_δ, Π_s	window, as δ_j, –	A candidate window of a step, 1 to 36; the step's prior spread, 0.15 on a coil, 1.0 K on a temperature sensor, 10 kW on the meter; the location prior, the prior probability of a step's window s (Chapter 12).
M_{none}, $M_{j,s}$, $\mathbf{a}_{j,s}$	–, –, varies	The drawing with no step (Chapter 10's base fit); the drawing with one step of kind j into window s; the step's column in the normalized rows, $-1/\sigma_q$ on a coil's trend row into s and $1/\sigma_k$ on a sensor's readings from s on (Chapter 12).
$B_{j,s}$, $\bar{B}_s$, $\bar{B}$, $B_{\max}$	–	Bayes factor of $M_{j,s}$ against M_{none}; its mean over the eight kinds at window s; its mean over all 288 candidates; the best candidate's (Chapter 12).
ψ, ψ_a	–	A step's prior variance over its variance from the other rows, $\sigma_\delta^2\,\mathbf{a}^\mathsf{T}(\mathbf{I}-\mathbf{K})\mathbf{a}$; the same after a candidate measurement a, $\sigma_\delta^2/\sigma_{\text{dat},a}^2$ (Chapter 12).
D_h, D_γ, $\mathbf{g}_{D_\gamma}$, $\mathbf{g}_{\text{split}}$	–, yr^{-1}, –, mixed	The difference $\ln h_{2,18} - \ln h_{1,18}$ of the two coils' indices (Chapter 5); the difference $\gamma_1 - \gamma_2$ of the two rates; its gradient; the gradient of coil 2's two parts of loss, $36\,\Delta t\,\gamma_2 + \delta_2$ (Chapter 12).
σ_{dat}^2, $\sigma_{\text{dat},a}^2$, σ_{new}^2	–	A step's data-only variance from the present record, with the prior taken out; the same after a candidate measurement a; that of the candidate's rows alone (Chapter 12, step 7).
$\tilde{\delta}$, $\tilde{\delta}_0$, $\tilde{\delta}_{\text{new}}$, $\tilde{\mu}$, $\tilde{v}$	–	The mean of a step's likelihood, the estimate the rows alone give; the present record's; the candidate rows'; the combined estimate's mean and variance under the posterior predictive (Chapter 12, step 7).
t_1, t_2	window	The two windows of a raw figure's change; for the summer drop, July of year 3 (window 31) and today (window 36) (Chapter 13).
ϕ_t, mo_t	rad, –	The season's phase of window t, $2\pi(\mathrm{mo}_t - 0.5)/12$, with $\cos\phi_t = 1$ between January and February; the month index of window t, 0 for January. Window 0 takes zero seasonal regressors (Chapter 13).

Symbol	Unit	Meaning
Y_t, y_t, $\check{Y}_t$, $\check{Y}_t^{\mathrm{m}}$	as the figure	The model value of a raw figure ($T_{r,t}$ or $\ln G_t^{\mathrm{app}}$) at window t; its reading; its value with the coils at the commissioned ones, at the window's own load and inputs; that value at the season's mean inputs, window 0's (Chapter 13).
s_{coils}, s_{err}	K or –	The coils' term and the reading errors' term of a raw figure's change from window t_1 to t_2, beside s_{season} and s_{load}; in bold, each term's column over the windows (Chapter 13).
$\mathbf{X}$, $\mathbf{X}^+$, $\hat{\mathbf{k}}$, k_c, k_s	varies	The design of a least-squares regression on the windows' times (Chapters 2, 11 and 13); its pseudo-inverse $(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}$; the fitted coefficients; a seasonal regression's cosine and sine coefficients (Chapter 13).
$\mathbf{y}^{\mathrm{raw}}$, $\mathbf{w}$, $\mathbf{w}^{\mathrm{line}}$, $\mathbf{w}^{\mathrm{seas}}$	as the figure, yr^{-1}	The readings of a raw figure in windows 0 to t_{as}; the slope row of a regression, the second row of $\mathbf{X}^+$, without and with an annual harmonic (Chapter 13).
s^{reg}, s^{line}, s^{seas}, $s^{\mathrm{seas}}_{\mathrm{drop}}$	varies	The slope of a regression; of the line and of the seasonal regression through a raw figure, per year; the seasonal regression's reading of the summer drop, its fitted trend over the five months plus the residuals' change (Chapter 13).
$\hat{\nu}_t$, $\hat{\sigma}$, $\mathrm{se}(\cdot)$	as the figure	A regression's residual at window t; the residuals' standard deviation; the practitioner's standard error from the residual scatter (Chapter 13).
$\mathbf{\Sigma}_{31,36}$, $\mathbf{g}_{31,36}$	–	The covariance block of $(g_{1,31}, g_{2,31}, g_{1,36}, g_{2,36})$; the gradient of the coils' term of the summer drop on those entries (Chapter 13).
g_u^c, g_u^s, $\mathbf{g}^{\mathrm{sea}}$, σ_{sea}	–	Cosine and sine amplitudes of coil u's seasonal conductance, on top of its trend; the four amplitudes as one vector; their prior spread, 0.05 (Chapter 13).
M_{base}, M_{sea}	–	The base drawing (Chapter 10's base fit) and the seasonal-conductance drawing (Chapter 13).
t^*, t_c	window	The window at which a parameter's path crosses its requirement level: the remaining useful life is t^* less the present window. From Chapter 14 it is the first passage of $G_{\mathrm{tot},t}$ to G^* on the extended state, read in continuous windows by interpolation between $t_c - 1$ and t_c, with $t_c = \lceil t^* \rceil$ the first window at or below the limit, the date; from Chapter 17, $t_c(\mathbf{z}; \mathcal{P}_{\mathrm{v}})$ is counted from a visit.
t_{sd}	window	A planning date: the October shutdown of year 4 (window 46).
Failures and the fleet		
λ, τ	yr^{-1}, yr	A constant failure rate; the length of a mission window (Chapters 2, 3, 15 and 16).
k, T_{tot}	–, unit-yr	A count of failures or corrective work orders; the exposure behind it.
ω, λ_{ref}	–, yr^{-1}	Log-rate $\ln(\lambda/\lambda_{\mathrm{ref}})$, the entry that carries a rate; its reference, 0.1 per year.

Symbol	Unit	Meaning
σ_ω, T_f, K_τ	–, yr, –	Posterior spread of the log-rate; time to a unit's next corrective failure; count of corrective work orders in a period τ.
F, H, S	–	Events of a diagnosis: the coil fouled; the condition flag fires; an inspection grade, of value s in $\Pr(S = s)$ (Chapter 3).
F_A, F_B, q_S, q_A, q_B	–	Two findings or replacements; the probabilities of a shared cause and of each one's own wear.
a, b, a_λ, b_λ	–, yr	Shape and rate of a gamma prior on a rate (primed after the record); the context tells them from an action a. From Chapter 15, where a is the age, they are written a_λ, b_λ: the prior is a count row of a_λ failures in b_λ unit-years.
$\lambda(a)$, λ_0, θ_a	yr^{-1}, yr^{-1}, yr^{-1}	Failure rate at age a, $\lambda_0 \mathrm{e}^{\theta_a a}$, the rate of occurrence of failures under minimal repair (Chapter 16); its value at age zero; its log-slope in age, a scalar.
n_a, E_a	–, coil-yr	Failures and exposure of the fleet in age bin a, the ages a to $a+1$ (Chapter 16).
μ_γ, τ_γ	yr^{-1}	Mean and spread of the fleet's degradation rates, the shared prior of partial pooling. In Chapter 9 $\tau_\bullet$ stands for either of the fleet's spreads, τ_γ or τ_c, and $n_\mathrm{coil} = 50$ counts the coils.
μ_c, τ_c	–	Mean and spread of the fleet's commissioning values.
$g_{i,0}$, γ_i, y_{ij}, a_{ij}, σ_test, k_i	–, yr^{-1}, –, yr, –, –	Commissioning value and fouling rate of fleet coil i; its j-th annual condition test and the age at it; the test's spread (0.04); its number of tests.
ℓ_c, ℓ_γ, $\varepsilon_{\mathrm{c},i}$, $\varepsilon_{\gamma,i}$	–	$\ln \tau_\mathrm{c}$, $\ln \tau_\gamma$, the entries that carry the spreads; the standardized deviations of the non-centered parametrization.
$\mathbf{\Omega}$, $\mathbf{\Lambda}_i$, $\mathbf{\Lambda}_\mu$, $\mathbf{\Lambda}_\mathrm{pr}$, $\mathbf{\Sigma}_i$	varies	Population covariance $\mathrm{diag}(\tau_\mathrm{c}^2, \tau_\gamma^2)$; coil i's information given the spreads and means; the information on the means; their prior information; coil i's covariance with the means integrated (Chapter 9).
Ψ, ϖ_m	–	The Gaussian rows' sum of squares at the conditional minimum; the weight of node m of a grid on the state: the spreads' grid (Chapters 9 and 20), the log-rate's (Chapter 15), the age slope's (Chapter 16).
π_i, $\mathcal{K}_i$, $\tilde{\gamma}_0$	–	Fleet position: coil i's percentile in the design's population of rates and its rank among the fifty (1 the fastest); coil 0's standardized rate.
γ_new, $\hat{\gamma}_0^\mathrm{chart}$	yr^{-1}	The rate of a new coil of the design (the posterior predictive); the fleet chart's slope of coil 0.
The UPS battery string		
$C_{\mathrm{bat},t}$, $R_{\mathrm{bat},t}$	Ah, mΩ	Usable capacity and ohmic resistance of the string in test t; $\bar{C}_\mathrm{bat}$, $\bar{R}_\mathrm{bat}$ their ratings, 100 Ah and 40 mΩ.
ξ_t, h_C, ΔR	–, –, mΩ	Log-capacity $\ln(C_{\mathrm{bat},t}/\bar{C}_\mathrm{bat})$; capacity against commissioning; resistance rise since commissioning.

Symbol	Unit	Meaning
v_0, Δv, h_{ch}, N	V, V, V, –	A block's loaded curve at the usable-empty point, its rise to full, and the charge curve's offset above it; the number of blocks, 40.
I_d, I_c, SOC	A, A, –	Discharge and recharge currents; state of charge.
$V_{d,t}$, $\dot{V}_{d,t}$, $V_{c,t}$, $\dot{V}_{c,t}$	V, V/h	Level and slope of test t's discharge line and of its recharge line.
$T_{\mathrm{run},t}$, V_{cut}, $\mathrm{SOC}_{\mathrm{cut},t}$, $V_{\mathrm{h},t}$	h, V, –, V	Run time to the cutoff; the cutoff on the string, 420 V; the state of charge there; the loaded head $N(v_0 + \Delta v) - V_{\mathrm{cut}} - I_d R_{\mathrm{bat},t}/1000$.
σ_V, $\sigma_{\mathrm{run},t}$, $\mathbf{s}$, $\mathbf{n}$	V, h, mixed	The voltmeter's spread; the run-time row's spread; the two discharge lines as coordinates; the ridge (null direction) of the discharge rows.
Probability		
$\Pr(A)$, $\Pr(A \mid B)$	–	Probability of event A; conditional on event B.
$p(\cdot)$	varies	A probability density; $p(\mathbf{z} \mid \mathbf{y})$ is the posterior of the state, $p(\mathbf{y} \mid \mathbf{z})$ the likelihood, $p(\mathbf{z})$ the prior.
$\mathbb{E}$, Var, Cov	varies	Expectation, variance, covariance.
$\mathcal{N}(\mu, \sigma^2)$	–	Normal distribution with mean μ and variance σ^2; Φ its standard cumulative distribution function, φ its density.
Inference objects		
$\mathcal{L}(\mathbf{z})$	–	The objective, minus the log posterior up to a constant.
$\hat{\mathbf{z}}$	varies	The maximum a posteriori (MAP) state, the minimizer of $\mathcal{L}$.
$\mathbf{J}$	varies	Jacobian of the normalized rows with respect to the state, evaluated at $\hat{\mathbf{z}}$.
$\mathbf{\Lambda}$	varies	Information matrix (Gauss–Newton Hessian of $\mathcal{L}$), $\mathbf{J}^{\mathsf{T}}\mathbf{J}$ for normalized rows.
$\mathbf{\Sigma}$, Σ_{ij}	varies	Posterior covariance of the state under the Laplace approximation, $\mathbf{\Lambda}^{-1}$; its entries. The posterior spread of entry i is written $\sqrt{\Sigma_{ii}}$.
$f(\mathbf{z})$, $\mathbf{g}$	varies	A derived quantity of the state and its gradient $\partial f/\partial \mathbf{z}$ at $\hat{\mathbf{z}}$.
$\mathbf{N}$, $\mathbf{J}_{\mathrm{red}}$	varies	Null-space basis of the physics rows; the Jacobian of the reduced state (Chapter 4).
f_L, f_U	varies	The ends of the credible interval of a derived quantity.
ϱ, $\boldsymbol{\varrho}$, Φ_2	–	A correlation coefficient (Chapters 2 and 3; in Chapter 24 that of two consecutive logged readings' errors); the posterior's correlation matrix (Chapter 7); the distribution function of two standard normals with correlation ϱ.
υ, $\mathbf{\Upsilon}$, S_{in}	–	Weight of a reading; the gain of the correction form; the innovation variance of one reading (Chapter 4).
x_j, w_j	–	Gauss–Hermite nodes and weights.
ϖ_j, N_{eff}, p_{L}	–	Weight of draw j in the reweighted draws; their effective sample size $1/\sum_j \varpi_j^2$; the Laplace normal used as the proposal (Chapter 5).

Symbol	Unit	Meaning
ℓ_t, ε_q	–	Magnitude of the slope of window t's reading row; the summed departure of unread windows from the line (Chapter 4).
y_{new}, $\tilde{y}_a$	varies	A future reading (the predictive) and the outcome of a candidate measurement a before it is taken (the preposterior).
U_{p}, $U(a)$	varies	A quantity of the posterior after $\tilde{y}_a$; its expectation over the outcomes, which ranks the candidates (Chapter 2's design form).
M, M_j, B_{AB}	–	A drawing (model) of the plant; the members of a family of candidates; the Bayes factor $p(\mathbf{y} \mid M_A)/p(\mathbf{y} \mid M_B)$.
M_{clean}, M_{fouled}, M_{bias}, B_{fc}, B_{fb}	–	Chapter 2's drawings of a warm condition test: the air path clean, the air path fouled, the hall sensor with a bias; the Bayes factors of fouled against clean and of fouled against bias.
M_{one}, M_{two}, $B_{\text{two,one}}$	–	Chapter 4's alternatives for the hall's air path, a single thermal resistance and two distinct heat-removal paths (beside the sensor bias M_{bias}); the Bayes factor between them.
$\mathbf{J}_a$, $\mathbf{\Lambda}_a$, $\mathbf{\Sigma}_a$	varies	The Jacobian rows of a candidate measurement, its information $\mathbf{J}_a^{\mathsf{T}}\mathbf{J}_a$, and the preposterior covariance $(\mathbf{\Sigma}^{-1} + \mathbf{\Lambda}_a)^{-1}$.
Decisions		
a, $L(a, \mathbf{z})$	–, R$	An action (repair, replace, inspect, wait); its loss when the state is $\mathbf{z}$.
$\mathcal{R}(a)$, c_a, C_e	R$	The expected loss of action a, $\mathbb{E}[L(a, \mathbf{z}) \mid \mathbf{y}]$; its fixed cost; the consequence of event e.
c, c^*, c_1^*	varies, –, –	A decision level on a monitored parameter (the work-order trigger), from Chapter 19 a level on the probability $q_{t_{\text{as}}}$; the level of least expected loss with a refit every month, 0.0197; the one-look level $C_{\text{false}}/(C_{\text{false}} + C_{\text{miss}})$, 0.0980 (Chapter 19).
κ	–	Assurance level of an acceptance rule: the posterior probability a requirement must reach (0.95 in Chapter 22's verdict; Chapter 21's break-even p_{be} replaces it with a level the costs set).
EVSI, EVPI, EVPPI	R$	Expected value of sample information: what a measurement is worth before an action. Expected value of perfect information, $\mathcal{R}(a^*) - \mathbb{E}[\min_a L(a, \mathbf{z}) \mid \mathbf{y}]$: what knowing the whole state would be worth, the most any measurement can be (Chapter 17, its general form; Chapter 20 specializes it to the fleet's orders). Expected value of partial perfect information: the same with one open question known instead of the whole state (Chapter 17).

Vocabulary

Plant graph The physical system as nodes (places where a conserved quantity balances), edges (paths along which it flows) and sources, with the relations that must hold among them.

State vector The collection of free quantities the fit estimates at once.

Pinned, free, known input A quantity treated as known (a boundary condition, or an input such as the supply temperature read by a precise instrument) against one carried in the state.

Residual, row A relation written so that it equals zero when satisfied: physics rows constrain the state, reading rows compare it with the records, trend rows tie one window's parameter to the next, prior rows carry handbook or commissioning knowledge.

Window The span of records fitted as one steady condition: one representative hour of a month here.

Health index A physical parameter of the model read against its commissioned value; never a score built from raw performance.

Condition monitoring Following the health index through the windows of a record.

Degradation rate The trend of the health index per year.

Remaining useful life The time until the health index crosses the level at which the asset can no longer do its job, as a date with a band.

P–F interval The time from the first detectable change of the health index to its crossing of the functional limit.

Condition-based maintenance Maintenance triggered by the health index crossing a decision level, rather than by the calendar.

Partial pooling Estimating each unit of a fleet with a shared prior whose mean and spread are themselves estimated, so that the fleet informs the unit and the unit informs the fleet.

Change-point detection Locating a step in a parameter's path by the evidence of a step freed at each candidate window.

State of health The capacity and internal resistance of a battery against their values when new.

Laplace approximation Replacing the posterior by the Gaussian whose covariance is the inverse curvature of the objective at the MAP state.

Delta method Linearizing a derived quantity around the MAP state so that its variance is $\mathbf{g}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{g}$.

Posterior predictive The distribution of a future reading obtained by averaging its model over the posterior.

Evidence, Bayes factor The normalizing constant of one model's posterior; the ratio of two models' evidences.

Preposterior The posterior that would result from a measurement not yet taken; the object that values a measurement before it is made.

The five computations Posterior, trend, evidence, preposterior and expected consequence: the objects every asset-management question is answered from.

Part I

Thinking in Probabilities

Chapter 1

Physical Assets as Graphs

1.1 One asset and several questions

The owner of a data center equipment hall asks its asset manager whether the hall can take twice its present load, from 30 to 60 kW, with the hall air then held below 27.5°C. The hall's air path, from the hall to the supply air that the cooling unit's discharge controller holds at 18°C, has a short history: a design sheet, which specifies its conductance as 6 kW/K, and a commissioning test, at which the one hall sensor, with an error of about half a kelvin, read 23.2°C at 30 kW. The health-index dashboard takes that reading as exact: at 23.2°C under 30 kW the path conducts 5.77 kW/K, an index of 0.962 against the design sheet, and the dashboard reports that the path was commissioned 3.8 % below design. At that conductance the doubled load puts the hall at 28.4°C, over the limit. Read from the dashboard alone, the path was handed over short of design and must be brought up to design before the load doubles.

The handbook tile beside it tells another story. The design handbook gives the conductance of a path of this kind as 7.5 kW/K, right to within a factor of about 1.35, an index of 1.25; heat leaves the hall in proportion to its excess over the supply air, so at that conductance the hall sits at 22°C under 30 kW and at 26°C under 60 kW, a comfortable pass. Neither tile carries a band, and neither is weighed against the other. One test is one window of the path's record, and it carries no degradation rate and no age (Chapter 4 adds a second window). What the decision needs is a probability that the path carries 60 kW within 27.5°C, and, before an upgrade is ordered, the probability that the path was commissioned below design at all. The question of this chapter is therefore

how a physical asset is represented so that one representation serves every question the asset manager asks of it (its condition, its capacity, the next measurement), and what must be fixed before any inference is performed.

Jargon: plant graph, node, path, source, pinned and free quantities, state vector, residual, row, observation, prior, posterior, design value, health index, commissioning test, commissioned value, model discrepancy, scenario node, Jacobian, normal equations, information matrix, covariance, Laplace approximation, factor graph

A *plant graph* represents a physical system as *nodes*, places where a conserved quantity such as energy or mass must balance; *paths* (edges), along which the quantity flows; and *sources*, which supply it. A quantity is *pinned* when it is treated as known, as a boundary condition, and *free* when it is unknown and must be inferred. The free quantities, collected in one vector, form the *state vector* $\mathbf{z}$. A *residual* is a relation written so that it equals zero when it is satisfied; each residual is a *row* of the model. A physical row states a relation that must hold among entries of the state; an *observation* (a reading) is a measured value related to an entry of the state with an error, and its row compares the two. The *prior* expresses what is known about the state before the readings, for example a handbook value of a conductance; the *posterior* $p(\mathbf{z} \mid \mathbf{y})$ expresses what is known after the evidence $\mathbf{y}$. The *design value* of a parameter is the value its design sheet specifies. The *health index* is a parameter of the physics read against a reference value, here the conductance of the air path as a fraction of its design value. A *commissioning test* (acceptance test) is the first test of the asset in service, and its fit, the *commissioned value*, is the baseline against which later condition tests are read. A *model discrepancy* is the difference between the physical system and the model of it that would remain if every entry of the state were known exactly. A *scenario node* is a node added to the graph for an operating condition that has not occurred, so that a question can be asked about it. The *Jacobian* $\mathbf{J}$ is the matrix of derivatives of every row with respect to every entry of the state, one row per row of the model and one column per entry. The *normal equations* $(\mathbf{J}^\mathsf{T}\mathbf{J})\,\Delta\mathbf{z} = -\mathbf{J}^\mathsf{T}\mathbf{r}$ give the step that best reduces all the rows at once when each is replaced by its tangent. The *information matrix* $\mathbf{J}^\mathsf{T}\mathbf{J}$ states how sharply the rows together fix the state, direction by direction, and its inverse is the posterior *covariance* $\boldsymbol{\Sigma}$, whose diagonal holds the variances of the entries and whose off-diagonal entries hold how they vary together. The *Laplace approximation* replaces the posterior by the normal distribution centered on its most probable state with that covariance. A *factor graph* draws the entries of the state and the rows that act on them, one node each.

Key idea

A graph describes the asset and the plant around it; a probabilistic question states what is to be learned about it. The plant graph, its pinned and free quantities and its physical rows are built once, and each question is a probability over the state of that one graph, with the unknown on the left of the conditioning bar and the evidence on the right. The posterior of the state is carried by three objects, the most probable state $\hat{\mathbf{z}}$, the covariance $\boldsymbol{\Sigma}$ and the Jacobian $\mathbf{J}$ of the rows, and every answer is an entry of the state read off them or a function of the state carried through them.

1.1.1 Several questions about one equipment hall

An asset manager responsible for the hall might ask any of the following:

1. What temperature is the hall running at now?
2. How far is the as-commissioned conductance of the air path from its design value?
3. Can the path carry the doubled load within its limit?
4. Is the hall sensor drifting, or is the path fouled?
5. Would another sensor sharpen the health index?
6. What happens if the path loses half its conductance at once?

The questions are different. The first two ask for the value of an unknown quantity; the third asks for the probability that a future temperature stays within a limit; the fourth asks which of two descriptions of the record is true; the fifth asks what a measurement not yet taken would be worth; the sixth asks about an operating condition that has not occurred. All six, however, concern the same physical asset: the same hall, the same air path, the same sensor.

Constructing a separate mathematical model for every question would repeat the same physics six times, and the six models could disagree about it. The method of this book begins instead with one common representation of the equipment and its physical relations, the *plant graph*. It is step 0 of the method stated in *How to read this book*: the system as it is, before any question is asked, built once and kept for every question that follows. Section 1.1.6 writes the questions that this chapter answers over the graph.

1.1.2 The system, the question and the evidence

Three things must be fixed before any probabilistic inference is performed, and the plant graph is where they are fixed. The first is the division of the quantities: which are known and pinned (the supply-air temperature, the heat load), which are unknown and free (the hall temperature, the conductance), and which are observed (the reading of the commissioning test). The field often separates the unknowns into *states*, such as temperatures and flows, and *parameters*, such as conductances and rates; in this book both kinds are entries of one state vector, and the inference treats them alike. The second is the set of physical relations that must hold among the quantities, such as conservation of energy at the hall. The third is the evidence: the readings, the records and the prior knowledge, attached to the graph as rows of their own.

The question is a separate object. It is stated over the graph after the graph is built, and it decides what is computed, not what the system is. The separation makes the asset manager's assessments systematic: when a new question arrives, the graph is already in hand, and the work is to write the question over its state.

> **Principle 1.1.** *The physical model is reusable; the probabilistic question determines what is calculated from it. A plant graph is built once from the equipment and its physics, and every question is a probability or an expectation over the state of that graph.*

1.1.3 Larger plant graphs and factor graphs

The equipment hall of this chapter has one thermal path and two unknown quantities, the smallest graph with every kind of element. The same construction serves a plant of many components, and the graph on which an inference operates is a second object, drawn from the plant graph and the evidence attached to it.

A cooling network. Real plants may include hundreds or thousands of components, each with its own equations and uncertainties. Consider a cooling network with two chillers, two pumps, a common electrical supply and several temperature sensors. Its plant graph may contain temperatures, pressures, flows and electrical states at its nodes; physical components represented by paths or by component relations; heat, mass and power conservation equations; sensor measurements and historical observations; failure and repair states of equipment; and shared parameters and common-cause dependencies.

Figure 1.1 draws a reduced version. The green boxes are free nodes: the two chillers and the two pumps, each carrying its status and capacity or flow, the supply header with its water temperature T_s and flow, and the hall with its temperature T. The gray box on the left is the pinned outdoor air, to which the chillers reject heat; the white box in the middle is the common electrical supply, which feeds all four machines; the white box at the upper right is

the heat load of the hall. Solid green arrows are physical paths of heat and water; thin gray arrows are electrical feeds; dashed arrows are observations, two temperature readings and the maintenance record of a chiller. The common supply is a single node on which four components depend: if it fails, all four stop together, and a graph that omitted it would treat their failures as independent.

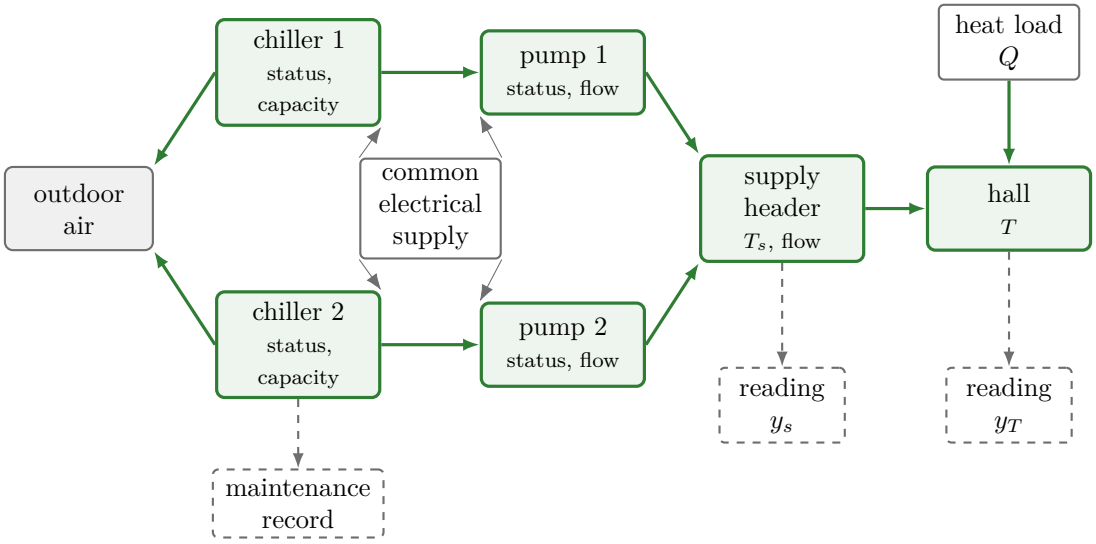


Figure 1.1: Plant graph of a reduced cooling network (illustrative; return piping omitted). Free: the status and capacity of each chiller, the status and flow of each pump, the header water temperature T_s and flow, and the hall temperature T. Pinned: the outdoor air, to which the chillers reject heat, and the heat load Q; the common electrical supply feeds all four machines and is the shared dependency. Observed: T_s and T through two readings, and chiller 2 through its maintenance record.

The mathematical principle is the one the single hall shows below. The network is a collection of physical relations,

$$r_1(\mathbf{z}) = 0, \quad r_2(\mathbf{z}) = 0, \quad \ldots, \quad r_m(\mathbf{z}) = 0,$$

one per balance and per component relation, over a state vector $\mathbf{z}$ that now has many entries. Observations and probability models for the uncertain quantities are attached as rows of their own. What is ultimately sought is the joint posterior of the state,

$$p(\mathbf{z} \mid \mathbf{y}, M),$$

where M represents the assumed system model, including its structure: which components exist, how they are connected and which dependencies are represented. Writing M explicitly matters because the structure itself may be uncertain, and a question about the structure is answered by comparing two such models, not by the posterior inside either.

Physical graphs and factor graphs. A physical graph and a probabilistic factor graph are related, but they are not the same object. The physical graph represents equipment and connections: halls, paths, chillers, pumps, sensors. A *factor graph* represents the variables of the inference and the mathematical relations among them, including the observations, the priors and the dependencies: each entry of the state is one node, each row of the model is one factor node, and each factor is linked to the entries its row involves. It is constructed from the physical graph and the evidence attached to it, and it shows which row acts on which unknown. The distinction becomes important when competing system structures and diagnosis are studied in later chapters. Two physical drawings of the same plant, one with a single air path and

one with two, give two factor graphs with different variables and rows, and comparing them is comparing two models M. A sensor bias adds a variable to the factor graph without changing the equipment at all.

1.1.4 The plant graph (step 0)

The simplest plant graph with every kind of element has one free node, one pinned node, one path, one source and one observation, and the equipment hall is that graph. It is built before any question is asked and before any number is fitted.

Nodes, a path, a source and an observation. Consider a simplified equipment hall with one heat-removal path, the hall's air path to the supply air that the cooling unit delivers. Figure 1.2 draws it as a plant graph. The green box on the left is the hall, a free node whose air temperature T is unknown; the gray box on the right is the supply air, a pinned node whose temperature $T_c = 18°\mathrm{C}$ is held by the cooling unit's discharge controller. The solid green arrow between them is the air path, the asset of this chapter, characterized by its effective thermal conductance G. The box above the hall is the heat load Q supplied by the servers, a source. The dashed box below is the hall sensor, read at the commissioning test, and the dashed arrow from the hall to it is an observation: the sensor reports a value related to T, not T itself. Solid connections are physical relations; the dashed connection is evidence.

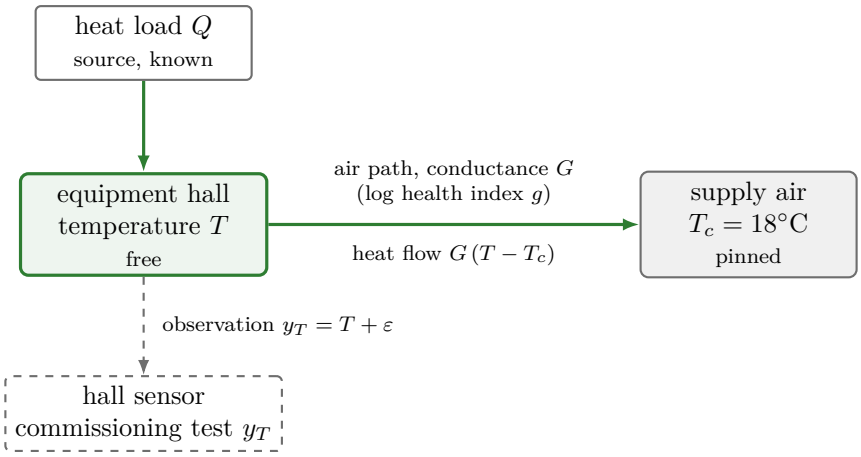


Figure 1.2: Plant graph of the equipment hall in steady state. Free: the hall air temperature T and the conductance G of the air path, carried as the log health index $g = \ln(G/G_{\mathrm{ref}})$ (Section 1.1.4). Pinned: the supply-air temperature $T_c = 18°\mathrm{C}$, held by the cooling unit's discharge controller, and the heat load Q. Observed: T, through the commissioning-test reading y_T with error ε. Solid arrows are physical relations; the dashed arrow is an observation.

Table 1.1 lists the elements of the graph. The first column names each element, the second its role in the graph, the third its symbol and the fourth its unit; the last column states how the element enters this first example.

The modeling choices of the graph. Three choices fix the graph before any question is asked, and the effective conductance is then the one uncertain property of the path. The hall is taken in steady state through the commissioning test, storing no more and no less heat from one moment to the next: this is what makes one reading at one load a statement about the conductance, and it fails in the period after a change of load or of conductance (Section 1.6.5). The heat load is pinned at its 30 kW: this is what leaves the conductance as the one unknown of the path, and it fails if the load is not known. The supply air is pinned at $T_c = 18°\mathrm{C}$ because

Table 1.1: The elements of the plant graph of Figure 1.2.

Element	Role	Symbol	Unit	In the example
Equipment hall	free node	T	°C	air temperature unknown
Supply air	pinned node	T_c	°C	held at 18°C by the discharge controller
Heat load	source	Q	kW	thermal power supplied by the servers, known
Air path	path (edge)	G	kW/K	effective conductance, uncertain; the asset
Hall sensor	observation	y_T	°C	the commissioning-test reading of the hall temperature

the discharge controller holds it at its set point: this is what makes it a boundary rather than an unknown, and it fails if the unit saturates and can no longer hold it (Section 1.6.5). The conductance lumps every mechanism by which the hall's heat reaches the supply air, the airflow through the cooling unit and the mixing of the hall air alike, into one number; how well that lumping represents the hall is a modeling assumption, examined in Section 1.6.5. This is the starting physical model.

Conservation of energy at the hall node. A graph becomes useful for computation when the mathematical relations between its elements are specified. In the steady state of the modeling choices the hall's temperature is constant, and conservation of energy at the hall node requires the heat entering to equal the heat leaving,

$$Q_{\text{in}} = Q_{\text{out}},$$

where Q_{in} and Q_{out} are thermal powers in kilowatts. The heat entering is the load, $Q_{\text{in}} = Q$. The heat leaving is modeled as proportional to the temperature difference across the path,

$$Q_{\text{out}} = G\,(T - T_c),$$

the defining relation of the effective conductance: G is the heat flow per kelvin of difference between the hall and the supply air. Combining the two,

$$Q = G\,(T - T_c), \tag{1.1}$$

where Q is the heat generated by the servers in kW, G the effective thermal conductance in kW/K, T the hall air temperature in °C and T_c the supply-air temperature in °C. A temperature difference has the same value in kelvin and in degrees Celsius, so the units balance: kW/K times K is kW.

Equation (1.1) is a *physical relation*, not a probability statement. It states what must hold if the simplified thermal model is correct, whatever is or is not known about its quantities. Since $G > 0$, dividing by G and adding T_c gives

$$T = T_c + \frac{Q}{G}, \tag{1.2}$$

which gives the temperature when the other three quantities are specified. The hall is warmer than the supply air by Q/G: more load or less conductance raises it.

The forward calculation. The hall is operated at the load $Q = 30\,\mathrm{kW}$, and its air path was designed for a conductance of $G = 6\,\mathrm{kW/K}$, the design value on its design sheet. With $T_c = 18°\mathrm{C}$, equation (1.2) gives

$$T = 18 + \frac{30\,\mathrm{kW}}{6\,\mathrm{kW/K}} = 18 + 5 = 23°\mathrm{C},$$

the steady-state temperature of the hall with a path built exactly to its design. The calculation used no probability, estimated no uncertainty and used no observational evidence. It is a deterministic *forward calculation*: from pinned inputs and a specified conductance to the temperature they imply. Every later chapter needs it, because every inference compares temperatures computed this way with what the sensor reports. The temperature falls as the conductance grows, with the slope $\partial T/\partial G = -Q/G^2$, which at the design point is $-30/36 = -0.833\,\mathrm{K}$ per kW/K: an error of $1\,\mathrm{kW/K}$ in the conductance moves the computed temperature by almost a kelvin. Doubling the load doubles the excess $T - T_c$ at every conductance: at $G = 6\,\mathrm{kW/K}$ the hall at $60\,\mathrm{kW}$ would be at 28°C.

The state vector. The forward calculation required the conductance to be known. In the inference problems of this book it is not, and neither is the hall temperature. The *state vector* is the collection of the unknown quantities that the inference estimates at once. In the equipment hall the unknowns are the hall temperature and the thermal conductance, and one possible representation is

$$\mathbf{z} = \begin{bmatrix} T \\ G \end{bmatrix}.$$

The state holds everything the inference treats as free; the pinned quantities T_c and Q are not in it, and neither is the reading y_T, which is evidence. The concept recurs throughout the book: the graph fixes the state, the evidence constrains it, and each question is a statement about it. Quantities that are entries of the state, such as T here, are answered directly by the inference; quantities that are functions of the state, such as the heat flow through the path, require one further transformation.

The log-conductance. A conductance is positive, and an estimate of it must be positive too. In practice the conductance is therefore often represented by a logarithmic entry,

$$g = \ln\!\left(\frac{G}{G_{\mathrm{ref}}}\right),$$

where $G_{\mathrm{ref}} > 0$ is a reference conductance in kW/K and g is dimensionless. Inverting,

$$G = G_{\mathrm{ref}}\,\mathrm{e}^{g},$$

which is positive for every real g: an inference free to move g anywhere on the real line can never produce a negative or zero conductance. A change in g is a relative change in G, since $\mathrm{d}G = G_{\mathrm{ref}}\,\mathrm{e}^{g}\,\mathrm{d}g = G\,\mathrm{d}g$, so that $\mathrm{d}g = \mathrm{d}G/G$; uncertainty in g reads directly as a fractional uncertainty in G. The reference value only sets the origin of g. In this chapter it is taken as $G_{\mathrm{ref}} = 6\,\mathrm{kW/K}$, the design conductance on the design sheet and in the forward calculation, so that $g = 0$ there, $G = 7.5\,\mathrm{kW/K}$ corresponds to $g = \ln 1.25 = 0.2231$, and a conductance 10 % above the reference to $g = \ln 1.1 = 0.0953$. The design value is chosen as the origin and not attached as a row: it states what the path was specified to deliver, not what it delivers, so it carries no evidence about the conductance, and the fit would be the same with another origin. Read against the same reference, the exponential of the entry is the *health index* of the path (in the framework, a parameter of the physics read against its design value),

$$h = \frac{G}{G_{\mathrm{ref}}} = \mathrm{e}^{g}, \tag{1.3}$$

the fraction of its design conductance the path has: $h = 1$, $g = 0$, is a path built to design, and the handbook's 7.5 kW/K is $h = 1.25$. The state vector becomes

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} T \\ g \end{bmatrix}, \tag{1.4}$$

with entries $z_1 = T$ in °C and $z_2 = g$, dimensionless. It is the one state of this chapter: every question below is a statement about these two entries.

The physics row. The thermal relation can now be written in terms of the state. Substituting $G = G_{\text{ref}}\,\mathrm{e}^{g}$ into equation (1.1),

$$Q = G_{\text{ref}}\,\mathrm{e}^{g}\,(T - T_c),$$

or, moving everything to one side,

$$r_{\text{phys}}(\mathbf{z}) = G_{\text{ref}}\,\mathrm{e}^{g}\,(T - T_c) - Q = 0. \tag{1.5}$$

The function $r_{\text{phys}}(\mathbf{z})$ is a *residual*, in kW. A residual measures how far a candidate state is from satisfying a relation: it is zero when the relation holds exactly and grows as the candidate departs from it. With $G_{\text{ref}} = 6$ kW/K, $T_c = 18$°C and $Q = 30$ kW, the candidate $\mathbf{z} = (23°\text{C},\ 0)$ gives $r_{\text{phys}} = 6 \times 1 \times 5 - 30 = 0$: it is the forward calculation. The candidate $\mathbf{z} = (23.2°\text{C},\ 0)$ gives $6 \times 5.2 - 30 = 1.2$ kW: at that temperature and conductance the path would remove 1.2 kW more than the servers supply, so the hall could not be in steady state. The candidate $(23.2°\text{C},\ \ln(30/31.2)) = (23.2°\text{C},\ -0.0392)$ again gives zero: the temperature 23.2°C is consistent with the physics if the conductance is $30/5.2 = 5.769$ kW/K. The zero set of r_{phys} is the curve $T = T_c + Q/(G_{\text{ref}}\,\mathrm{e}^{g})$ in the plane of (g, T); the physics alone allows every point of that curve and no point off it.

The same row has a second form, in the units of the quantity it fixes. Dividing equation (1.5) by $G_{\text{ref}}\,\mathrm{e}^{g} = G$, which is positive, leaves the zero set unchanged and gives the row in kelvin, the departure of a candidate temperature from the temperature the physics implies. Divided further by a *model tolerance* σ_p, the temperature error in kelvin that the physics is allowed, it becomes a dimensionless row,

$$r_1(\mathbf{z}) = \frac{T - T_c - (Q/G_{\text{ref}})\,\mathrm{e}^{-g}}{\sigma_p}, \tag{1.6}$$

which counts the violation of the physics in multiples of its tolerance. This is the form in which the physics row enters every computation of the book. The formulation extends to systems of any size. A larger physical system is represented by a collection of residuals, one per relation: an electrical network has rows for Kirchhoff's current and voltage laws, a fluid network rows for mass conservation at its junctions and for the pressure losses along its pipes, and a cooling plant rows for the energy balance at each of its nodes. Each row has the form of equation (1.5), and the inference treats them alike.

Exact constraints and model discrepancy. Physical rows and reading rows should not automatically be treated as equivalent kinds of noisy measurement. Conservation of energy is a law: if the hall is in steady state and the model has captured every path, equation (1.5) holds exactly, and it can be imposed as an *exact constraint*, $r_{\text{phys}}(\mathbf{z}) = 0$, which restricts the state to the zero set of the row. In a computation the exact constraint is carried as the physics row (1.6) with a model tolerance σ_p far smaller than every other spread of the problem, and exactness is the limit $\sigma_p \to 0$; the limit is checked by showing that the results stop moving as σ_p shrinks, never by using the row to eliminate an entry of the state. A reading row, by contrast,

is meant to be violated by an amount of the order of the sensor's error: the reading is not the temperature.

Approximate physics is a third case. The single effective conductance of equation (1.1) is an idealization; the real hall has several paths, a conductance that varies with airflow, and stored heat. When the relation is only approximately true, the honest model is

$$G_{\mathrm{ref}}\, \mathrm{e}^{g}\, (T - T_c) - Q = \Delta_{\mathrm{m}},$$

with an explicit *model discrepancy* Δ_{m} in kW, an entry of its own with a stated prior, for example $\Delta_{\mathrm{m}} \sim \mathcal{N}(0, \sigma_{\mathrm{m}}^2)$. Algebraically this is a physics row with the weight $1/\sigma_{\mathrm{m}}^2$, and in the objective it looks like a reading row. It behaves differently. A measurement error is new at each reading and averages out over many readings; a discrepancy is a property of the model and repeats wherever the same operating condition repeats, so that many readings at one condition do not reduce it. Treating a discrepancy as measurement noise therefore makes the inference overconfident, and treating a law as noisy lets the inference trade physics against data. The choice between an exact constraint and an explicit discrepancy is made when the graph is built, and it is stated with the model. In this chapter conservation of energy is exact.

1.1.5 The record and the handbook value

The record of the hall is one reading, taken at the commissioning (acceptance) test of the air path: window 0 of the path's record. The hall sensor, the one temperature sensor installed in the hall, reported at the load of 30 kW

$$y_T = 23.2^{\circ}\mathrm{C}.$$

The sensor is not perfectly accurate. Its measurement model is

$$y_T = T + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma_T^2),$$

where ε is the measurement error in kelvin, assumed normal with zero mean, and σ_T its standard deviation, which the sensor's specification gives as

$$\sigma_T = 0.5\,\mathrm{K}.$$

The reading is the only observation: the supply-air temperature and the load are pinned, and nothing measures the conductance. The *commissioned value* $\hat{g}_0$ is the fit of this test, the most probable value of the log-conductance under the posterior of step 3, written $\hat{g}$ in this chapter of one window; it is the baseline against which later condition tests are read. Chapter 4 adds a later condition test of the same path at window 24, and Chapter 5 reads each coil of a larger hall against its own commissioning fit.

Knowledge of the conductance that does not come from the reading is also in hand. A design handbook, or earlier tests of halls of the same kind, give an estimate of the effective conductance with an uncertainty; for this hall (illustrative) the handbook gives 7.5 kW/K, with the true conductance within a factor of about 1.35 of it with probability of about 0.68. The handbook value is not a measurement of this hall, and it is not the design value either: the design sheet states what the path was specified to deliver, the handbook what paths of its kind are found to deliver. It is information held before the reading, and it enters the model as a prior.

1.1.6 The questions (step 1)

The asset manager's questions can now be written over the graph, each as a probability or an expectation over the state, with the unknown on the left of the bar and the evidence on the

right. Table 1.2 lists the five that this chapter answers: the first column is the question in the asset manager's words; the second the general form that every question of its category takes, written in the book's notation for a quantity $f(\mathbf{z})$ derived from the state, a threshold f^* on it and the posterior density $p(\mathbf{z} \mid \mathbf{y})$; the third its mathematical form for the hall; the fourth its category among the five named in *How to read this book*, which Chapter 2 develops; and the last the subsection where it is answered. The first two are estimation, of an entry of the state and of a quantity derived from one, the health index $h = G/G_{\text{ref}} = \mathrm{e}^{g}$ of equation (1.3), with the probability that it lies below one, below design. The third and fourth are predictions under a threshold, tail probabilities of the posterior: the third at the present load, the fourth at a load of 60 kW at which the hall has never run, whose temperature T_2 is a quantity the graph acquires in step 2. The fifth is a design question about a measurement not yet taken; it asks nothing new of the present evidence, and it is answered as the next measurement for the first four.

Table 1.2: The questions asked of the equipment hall at its commissioning test, the general form of their category, their mathematical form for the hall, their category and where each is answered. h is the health index of equation (1.3), T_2 the hall temperature at the load $Q_2 = 60\,\text{kW}$ (Section 1.1.7); $\mathbf{\Sigma}_a$ the covariance expected after a candidate row with Jacobian row $\mathbf{J}_a$.

Asset manager's question	General form	Mathematical form	Category	Answered in
What temperature is the hall running at now?	$\mathbb{E}[f(\mathbf{z}) \mid \mathbf{y}]$, $\mathrm{Var}(f(\mathbf{z}) \mid \mathbf{y})$	$p(T \mid \mathbf{y})$, $T = z_1$	estimation	1.3.1
How far is the as-commissioned conductance from its design value?	$\mathbb{E}[f(\mathbf{z}) \mid \mathbf{y}]$, $\mathrm{Var}(f(\mathbf{z}) \mid \mathbf{y})$, $\Pr(f(\mathbf{z}) \le f^* \mid \mathbf{y})$	$p(h \mid \mathbf{y})$, $h = G/G_{\text{ref}} = \mathrm{e}^{z_2}$; $\Pr(h < 1 \mid \mathbf{y})$	estimation of a derived quantity	1.3.1
Is the hall within its 24°C limit now?	$\Pr(f(\mathbf{z}) < f^* \mid \mathbf{y})$	$\Pr(T < 24^\circ\text{C} \mid \mathbf{y})$	prediction under a threshold	1.3.2
Can the path carry 60 kW within 27.5°C?	$\Pr(f(\mathbf{z}) < f^* \mid \mathbf{y})$	$\Pr(T_2 < 27.5^\circ\text{C} \mid \mathbf{y})$, $T_2 = T_c + (Q_2/G_{\text{ref}})\,\mathrm{e}^{-z_2}$	prediction under a threshold	1.3.2
Which added sensor would narrow the health index?	preposterior covariance $(\mathbf{\Sigma}^{-1} + \mathbf{J}_a^{\mathsf{T}}\mathbf{J}_a)^{-1}$	$\mathbf{\Sigma}_a$ for each candidate row	design	1.6.1

The questions differ in what they compute and agree in what they are computed over: the same two entries of the state, under the same posterior. Two of the asset manager's six questions are not in the table. Whether the hall sensor is drifting or the path is fouled compares two complete models of the same reading, the plant with a fouled path read by a true sensor and the plant with a bias entry on the reading row, and its answer is a ratio of their evidences rather than a property of either posterior; when every sensor of a larger plant is a candidate, the same comparison run over the candidates is a diagnosis. Chapter 7 answers it, and the residual check of Section 1.4.3 takes a first, coarse look at it. What happens if the path loses half its conductance at once is a question about a transient that the steady-state graph cannot represent, and Section 1.6.5 states what the graph would need for it.

1.1.7 The evidence as rows (step 2)

Step 2 attaches the reading and the handbook value to the graph as rows on the state $\mathbf{z} = (T, g)^{\mathsf{T}}$, and extends the graph with the operating condition the fourth question names. The three rows,

END OF THE PREVIEW

Asset Management by Derivation

These are the opening pages. The complete book runs 465 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