Reliability of a cooling service from trip tests, outage logs and disturbance windows, by probabilistic inference on a two-hall data center

Ricardo Marquez
rmarquez123@gmail.com

November 10, 2026

Abstract

A data center sells its tenants an availability, and the cooling redundancy behind that promise is claimed on the strength of a reliability model whose inputs, the success structure of the equipment, the unavailabilities of the components and their independence, are asserted rather than measured. A wrong input is silent: the plant runs at a risk it does not know it carries until the day a unit fails and a hall overheats. The plant already keeps the records that could test every one of those inputs, trip tests, outage logs and disturbance windows, and today those records are filed, not used. This report shows how to use them. It is an inference study on a two-hall system with one wall opening between the halls and a shared electrical feeder: readings with declared sensor noise are weighed against what the plant believed beforehand, the handbook rates and the drawing's values with their stated spreads (the priors), to give what it should believe afterward (the posteriors); the N+1 claim is read as the posterior probability that hall B stays within its allowable rise with its unit out and the full load applied, and the uncertainty of the service availability is read from the posteriors, so that each asserted input is replaced by a tested one with a stated confidence and a stated regime in which the record cannot tell. Five questions a reliability engineer asks of the plant are each answered from its own record. A trip test that takes one unit out establishes the claim at 0.97 when the opening's capacity is well above the load, and leaves it at 0.53 from a load-bank test when the capacity is within a tenth of the load; the design the forward study called redundant is not acceptable at the usual confidence from a load-bank test, and the margin an acceptable design needs grows as the test load falls. Three years of outage record put a band on the service availability that spans a factor of about nine in unavailability around the asserted value, and an unsettled structure widens it far more than the record narrows it. The measurement that tightens the band most is a matched sensor pair while the structure is unsettled and a year of the feeder's record once it is settled, with every trip test then worth nothing. Three years of the site's disturbance log name the logged disturbance as raising one cooling unit's failure rate, put the other at the decision threshold, where the Laplace form names it and the grid reference does not, and leave its effect on the feeder open, at a factor near three on the units' rates with an interval the log leaves wide; the same posterior, carried through a second query, puts the cost to the service at 0.020 of availability in a full disturbance month, from 0.983 to 0.962, with its interval and a posterior probability of 0.90 that it is a loss. What the owner gains is a redundancy claim that states what it rests on and names the measurement that would strengthen it, which a count of units cannot give. The system is small enough that every structure and every band is checked against a closed form; the readings and records are synthetic and the report says where that limits it.

1 Introduction

A data center sells an availability, and the availability rests on a reliability model of its cooling: a block diagram or a fault tree in which the equipment is drawn, each component carries an hour-level unavailability from a handbook or a vendor, and the components are taken to fail independently, or a common-cause fraction is asserted beside them. The model's arithmetic is exact once its inputs are given (Rausand and Høyland 2004; Billinton and Allan 1992; Vesely et al. 1981; the series and parallel formulas of IEEE Std 3006.5, IEEE 2014), and the inputs are what the redundancy claim actually rests on: the success structure, which combinations of failed components lose the service; the unavailabilities; and the independence that the products assume, or the common-cause fraction asserted in its place. In practice all three are asserted when the model is built and rarely tested against what the plant records afterwards. The plant does record things. A commissioning or periodic trip test takes a unit out of service and logs what the halls do; the maintenance system logs every outage and its repair; the site's disturbance log says which months ran under stress. Each of these is evidence about one of the model's inputs, and none of it enters the claim. The quantity the claim states is the service availability, A_{service}, the probability that the cooling service is delivered at a random hour: an availability in the sense of IEEE Std 3006.7 (IEEE 2013), not the survival probability over a mission time that the word reliability denotes in that standard. The forward study named the same quantity the service availability; this report names it for what it computes.

What a record can and cannot say about a reliability model is an old question with standard tools. Failure data are analyzed with rates and their confidence bounds (IEEE 2007), the comparison of structural hypotheses on readings is the Bayes factor (Kass and Raftery 1995), the worth of a measurement before it is taken is the expected information it adds (Lindley 1956) or the decision it improves (Raiffa and Schlaifer 1961), the sensitivity of a system's reliability to one component's is its Birnbaum importance (Birnbaum 1969), and the treatment of components that share a cause of failure is the common-cause-failure literature (Mosleh et al. 1988). The thermal side has its own standard: the allowable temperature envelope of IT equipment (ASHRAE 2021) is what decides whether heat crossing between halls is acceptable. What is less standard is doing all of this on one model of one plant, so that a trip test, an outage log and a disturbance log each update the same service availability.

TERRA is a general-purpose engine for probabilistic inference on physical networks. It treats a network as a graph on which the physics is written locally: a conservation law at every node, a constitutive law on every edge, and, attached to the same graph, whatever the sensors report and whatever is known in advance. Each of these is a relation that holds only to within a stated tolerance, and the engine weighs them together to find the most probable state of the system, the uncertainty of every quantity in it, and the relative plausibility of competing explanations of the same readings. Each physical domain brings its own laws, units and file formats; the inference and the bookkeeping are shared. This study writes the two-hall thermal network as such a graph, attaches the trip test's readings to it, and uses the engine's posteriors and evidences to answer five questions; the outage record and the disturbance log enter through the same component model the forward reliability calculation uses. Section 3 states the framework and each of these constructs precisely.

This report takes the two-hall reliability model of a forward study, in which every input was asserted, and tests its inputs against records. Testing the inputs instead of re-asserting them matters because a wrong input costs nothing until it costs everything: a claim of one spare unit on the wrong structure reads the same on a dashboard as one on the right structure, and the first evidence against it is an outage. The records that could test it are cheap and already exist. What

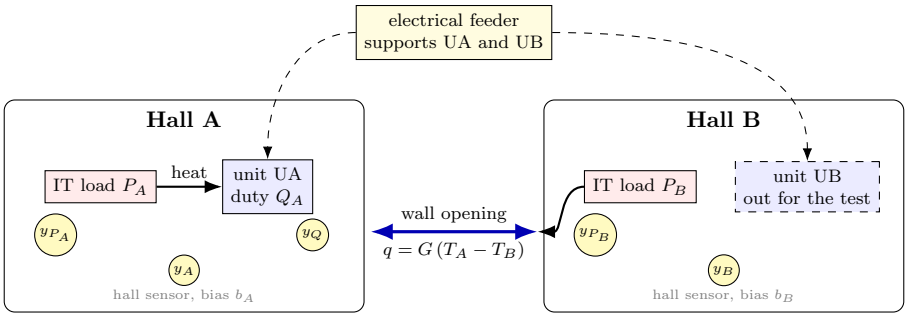


Figure 1: The system and its sensors. Two data halls, each with an IT load of 60 kW in service and one cooling unit, connected by a wall opening through which heat crosses in proportion to the temperature difference. One electrical feeder supports both units (dashed arrows). The trip test takes unit B out, so that hall B's heat must cross the opening to unit A; the circles are the readings taken: the two hall temperatures, each from a sensor with its own offset bias, the surviving unit's duty, and the two PDU powers.

the process adds to each record is a confidence and the regime in which the record cannot tell, which is what an owner needs to decide whether to spend on a test, a sensor or a longer log. The question is what a reliability engineer who owns the plant can learn about the redundancy claim from what the plant logs: whether the N+1 claim is true, how uncertain the service availability is given the plant's own outages, which measurement would tighten that most, whether a shared disturbance raises the components' failure rates, and what that does to the service availability. The construction is a factor graph of the two halls on which the trip test's readings are reconciled, a Poisson posterior on each component's rates from its record, the delta method on the service availability, and a fit of each component's sensitivity to the logged severity of the disturbance. What the report adds is the demonstration that each asserted input of a cooling-redundancy claim can be tested by a record the plant already keeps, with the cases in which the record cannot tell named as such. Section 2 describes the system, the readings and the records, Section 3 the model and the five arms, Section 4 the results, Section 5 their limits, and Section 6 the answers; the appendices carry the reproduction recipe, the data statement, one arm worked by hand and the notation.

2 System and data

The system. Figure 1 shows it. Two data halls, A and B, each carry an IT load of 60 kW in service and each has one cooling unit that could carry both; a wall opening connects them, through which heat crosses in proportion to the temperature difference with conductance G; one electrical feeder supports both units. The forward reliability model of this system has one step: with one unit out, the other hall's 60 kW must cross the opening, and the units are a parallel pair when the opening can carry it and a series pair when it cannot. The forward model states that limit as a heat-flow capacity C. In a steady test the heat an opening carries is the conductance times the temperature difference, and what is physically limited is the difference the halls may run at: hall B may not rise more than the allowable difference of 8 K above hall A, the span from the 22°C set point to the top of the allowable envelope (ASHRAE 2021). The N+1 claim is therefore a relation on the state: with unit B out and the full load applied, hall B's rise over hall A is the load over

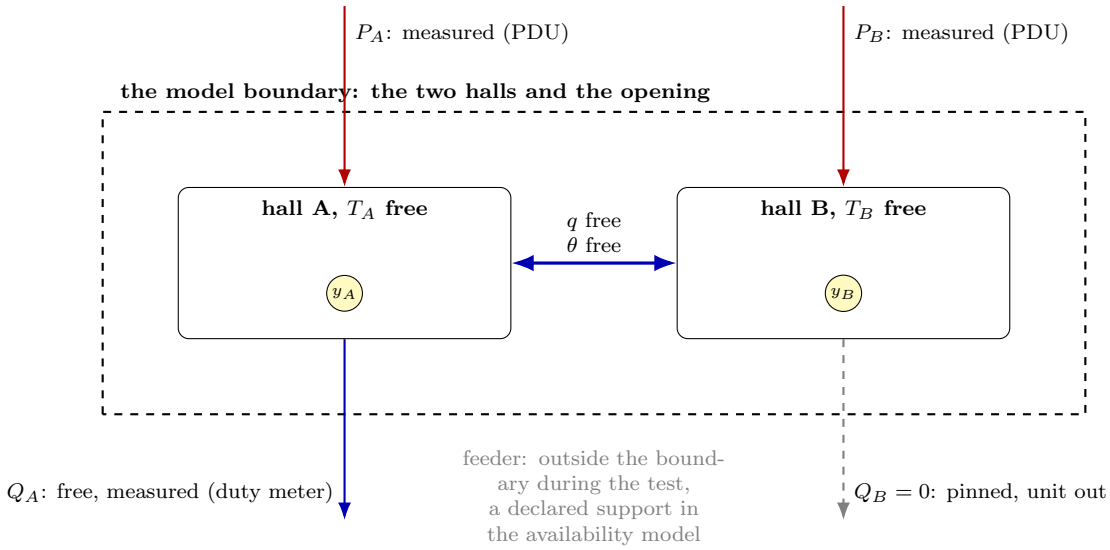


Figure 2: The partition. The model's boundary encloses the two halls and the opening; every crossing is labeled. The IT powers enter as measured flows, the surviving unit's duty is a free flow with a meter on it, unit B's duty is pinned at zero, the opening's heat and its conductance (the entry θ) are free, and both hall temperatures are free: the level of the temperatures is fixed by hall A's reading under the set-point prior, never by a boundary. The feeder lies outside the boundary during the test.

the conductance, and the claim holds when that rise is within the allowable,

$$\Delta T_{\text{full}} = \frac{Q_{\text{hall}}}{G}, \tag{1}$$

$$\Delta T_{\text{full}} \le \Delta T_{\text{allow}}, \tag{2}$$

$$C = G\,\Delta T_{\text{allow}}. \tag{3}$$

Equation (1) is the rise under the contingency, (2) the claim, and (3) the forward model's capacity read as the heat passed at the allowable difference, so that (2) is $C \ge Q_{\text{hall}}$. The success structure follows from the relation and is not chosen when the model is built: the units are a parallel pair when (2) holds and a series pair when it fails. A hard cap on the flow of the kind a damper or a duct fan would impose is not observable in a test below it and is not estimated; Section 5 returns to this. Two cases of the opening are studied: a far case whose true capacity is 80 kW, the forward study's own and well above the hall's load, and a near case at 66 kW, within a tenth of it. Appendix C pairs each with a drawing below the load, 30 and 54 kW, for the registered comparison of drawings.

The partition. The model's boundary, with every crossing labeled, is drawn in Figure 2. The trip test takes unit B out, so its duty is pinned at zero; the IT powers enter as measured flows, the surviving unit's duty as a free flow with a meter on it, the opening's heat and its conductance as free quantities, and both hall temperatures as free quantities. The balances enter only through temperature differences, so one level is a gauge: hall A's reading fixes it, under a prior that the unit holds its set point within s_T. Pinning hall A's temperature instead would identify its sensor's bias from its own reading and shrink every width in the report by the square root of two, which is why the choice is registered. The feeder lies outside the boundary during the test and enters the reliability model as a declared support, as in the forward study.

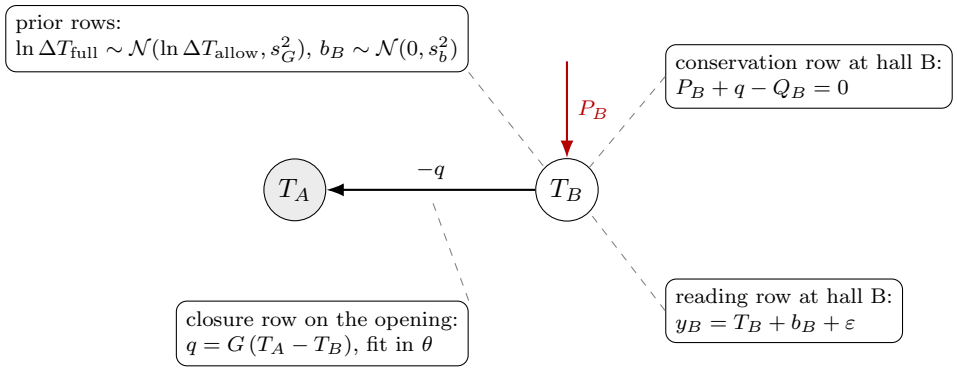


Figure 3: The model around hall B. The unknowns are the two hall temperatures, the heat on the opening, its conductance as the entry θ, the IT loads, the surviving unit's duty and the two sensor biases. The rows are the heat balance at each hall, the closure law on the opening with the conductance in the log so that it stays positive, one reading row per sensor with its bias, and a prior row per approximately known quantity. Every row carries a weight, and the fit minimizes the weighted squared misfits of all rows at once.

The model. The rows around hall B are shown in Figure 3. The unknowns are the two hall temperatures, the heat on the opening, its conductance, the two IT loads, the surviving unit's duty and the two sensor biases. The rows are a heat balance at each hall, the closure law on the opening, one reading row per sensor carrying its bias, and a prior row for each approximately known quantity: the conductance, the set point and the biases. Section 3.3 states them.

The data. Figure 4 shows the raw record of the trip test and Figure 5 the outage records; every input is listed in Table 1. All readings and records are synthetic, generated from the stated models with the stated noise, and the report is a study of what such records can establish, not of a particular plant. The trip test logs each hall's temperature once a minute for 30 minutes after settling (sensor noise 0.3 K per sample), the surviving unit's duty (2 %) and the two PDU powers (1 %); each temperature sensor carries an offset bias whose prior spread is 0.5 K, and the bias is what averaging cannot remove: the floor on the hall difference is $\sigma_{\Delta T} = 0.71$ K, of which the bias is almost all. The outage records are 1, 3, 10 unit-years per component at the expected counts of a failure rate of 2.5 per year and a 180 h repair for each unit (the unavailability 0.05 of the forward model) and 3.69 per year and 24 h for the feeder (its 0.01), each failure with its repair in the log. The disturbance log is 36 months of 720 h, each with its severity index, a normalized intensity of the stress the site logged, the months of July and August of each year at one, June and September at one half and the rest at zero, and the failures and downtime each component logged in it; under a full disturbance month a unit's failure rate is four times its baseline and the feeder's twice. The assumptions are these. The thermal physics is steady and linear, read after settling. A sensor's bias is constant over the test. Failures are Poisson in operating time and repairs Poisson in downtime, each arriving at a steady rate with no memory of the last, so that the unavailability is $\lambda r/(1+\lambda r)$. The feeder's effect on the units is declared. The disturbance acts on a component's failure rate as a factor exponential in the month's severity, with the sensitivity unknown and the repair rate unchanged. Every threshold, grid, noise, prior, count and seed was registered before any result was computed, with dated additive amendments, and the results file records the registration it ran under.

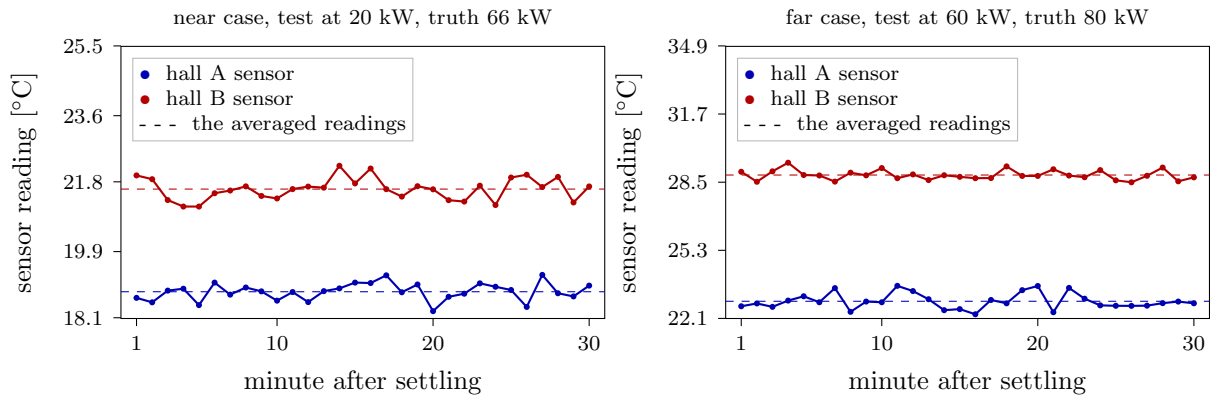


Figure 4: The raw record of the trip test for the registered seed, two of the six tests: the two hall sensors logged once a minute over the settled half hour, with the averaged readings the fit consumes as dashed lines. Left, the near case at the 20 kW load-bank test: the halls differ by a few kelvin and the sensor noise is a visible fraction of it. Right, the far case at full load: the difference is several kelvin and the noise is not. What neither panel shows is the sensor bias, a constant offset that averaging cannot remove; it is carried by the prior.

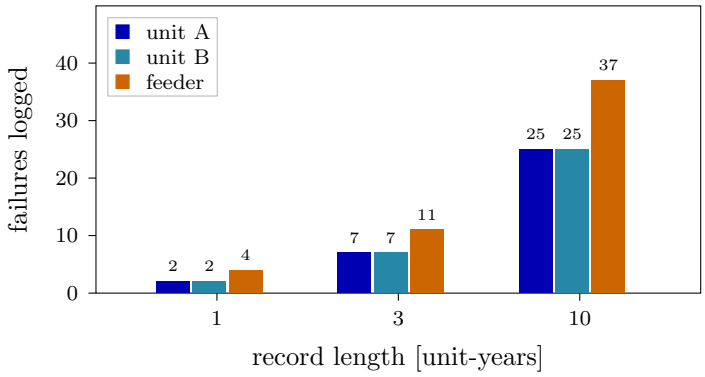


Figure 5: The outage records the second question is asked of: the failures logged per component over records of one, three and ten unit-years, with the number above each bar. Each failure carries its repair in the repair log (the mean repair times of Table 1). The records are synthetic, at the expected counts of the stated rates; a real record would have fewer or more, which is the point of the posterior.

3 Methods

3.1 Overview

The five arms are laid out in Figure 6. Each starts from a record the plant keeps, passes through one inference, and ends in a quantity the engineer acts on. The first arm (Section 3.4) reconciles a trip test's readings with the factor graph of Section 3.3 and reads the claim's probability off the posterior of the full-load rise, with its acceptance rule, its margin, and its power over the tests that could happen, with no test simulated; the registered comparison of the forward study's two drawings, a question apart, is Appendix C. The second (Section 3.5) turns an outage record into a posterior on each component's rates and propagates it through the enumeration into a band on the service availability, by draws and in closed form. The third (Section 3.6) ranks candidate measurements by the variance they take off that band. The fourth (Section 3.7) fits each component's sensitivity to a shared disturbance from the site's disturbance log and names the disturbance where it raises

Table 1: The registered inputs, each with the symbol the text uses for it (Table 20 says what each symbol means). Every number in the report is computed from these; the decision gaps and grids were committed before any result, and one dated amendment (the balance-row tolerance) is recorded in the registration file.

quantity	symbol	unit	value
IT heat load per hall in service	Q_{hall}	kW	60
trip-test loads per hall	Q	kW	20, 40, 60
allowable hall difference	ΔT_{allow}	K	8
design pairs, far / near	C	kW	30 / 80 and 54 / 66
temperature sensor noise per sample, samples per test	σ, N	K, –	0.3, 30
sensor bias prior, set-point prior	s_b, s_T	K	0.5, 2.0
duty meter, PDU meter, relative noise	σ_Q, σ_P	–	2 %, 1 %
prior on the full-load rise (unrestricted): center, spread	ΔT_{allow}, s_G	K, –	8, $\ln 3$ (a factor of 3)
decision gap, drawing tolerance	Δ^*, τ	nats, –	1, 2 %
unit failure rate, mean repair time	λ, r	1/yr, h	2.5, 180
feeder failure rate, mean repair time	λ, r	1/yr, h	3.69, 24
rate priors: center at the stated rate, spread	s_λ, s_μ	–	$\ln 3$ (a factor of 3)
disturbance log: months, hours per month	M, T	–, h	36, 720
severity schedule: months at 0 / 1/2 / 1	d_t	–	24 / 6 / 6
sensitivity prior: center, spread	s_β	–	0, $\ln 3$ (a factor of 3)
true sensitivities, unit / feeder	β	–	$\ln 4$ / $\ln 2$
confidence that names the disturbance	–	–	0.95

a rate. The fifth (Section 3.8) carries that posterior, with no new inference, into the change the disturbance makes in the service availability. The five arms come in the order the owner's decisions come: first whether the claim's structure is right, since nothing else matters if it is not; then how much the plant's own history says about the number; then which spend would say more; then whether the independence the whole product rests on has been violated; then what the violation costs the service. Section 3.2 states the probabilistic framework every arm rests on, Section 3.3 its instance for the trip test, and Section 3.9 the measures.

3.2 The probabilistic model

Every arm of this study rests on one construction, stated here once and written out for the trip test in Section 3.3. Every relation the state must satisfy, a physics balance, a reading, a prior, a count, is a factor $\psi_k(z)$, and the state is inferred from one objective, the negative logarithm of their product,

$$C(z) = -\sum_k \log \psi_k(z), \tag{4}$$

whose mode, and whose curvature at the mode inverted, are the two operations every arm invokes; the class of a factor changes only its term in the sum. The Gaussian factor is the principal case and is written out next; the Poisson factor of a count, which the record arms attach, adds its term to the same sum. A physical network is a graph, and the engine writes its physics on that graph. The state is a vector z of unknowns: one potential per node, one flow per edge, and whatever else is uncertain, such as demands or parameters. Every relation the state must satisfy is a factor, written as a residual $r_k(z)$ that vanishes when the relation holds and carrying a weight w_k, the inverse variance of how far it may be off. A conservation law at a node, a constitutive law on an edge, a sensor reading and a prior on an approximately known quantity are all factors of this one

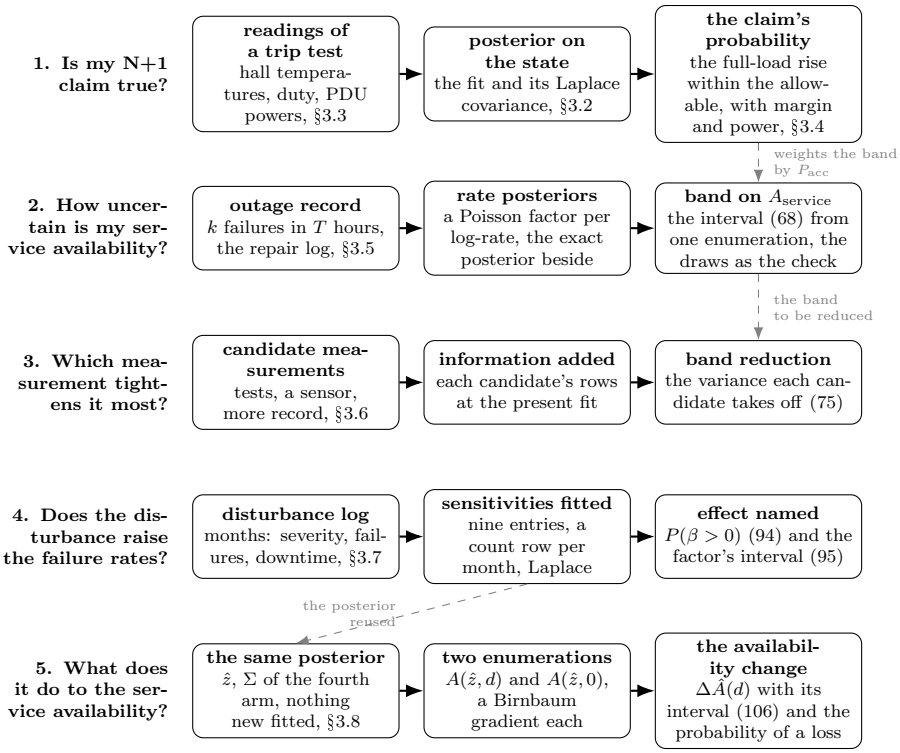


Figure 6: The method: five arms, one per question, each a chain of three steps from a record to a verdict. The first arm's probability that the claim holds weights the second arm's band, the second arm's band is what the third arm reduces, and the fourth arm's posterior is what the fifth arm queries. Every arm runs on the same two-hall network and the same component model.

kind; they differ only in their residual and their weight, a physics relation being trusted to a small tolerance, a reading to its sensor's noise, a prior to its stated spread. Treating each factor as an independent Gaussian belief about its residual gives the posterior density of the state given the readings y,

$$p(z \mid y) \propto \prod_{k} \exp\!\big(-\tfrac{1}{2}\, w_k\, r_k(z)^2\big). \tag{5}$$

The state that maximizes it, the *maximum-a-posteriori* (MAP) state z^*, minimizes the weighted sum of squared residuals $C(z)$ and is found by Gauss–Newton iteration on the sparse Jacobian J of the residuals, $J_{ki} = \partial r_k / \partial z_i$, with W the diagonal matrix of weights, from a flat start until the gradient norm falls below a tolerance ϵ:

$$C(z) = \tfrac{1}{2} \sum_{k} w_k\, r_k(z)^2, \tag{6}$$

$$z^* = \arg\min_{z} C(z), \tag{7}$$

$$z \leftarrow z - (J^\top W J)^{-1} J^\top W\, r(z) \quad \text{until } \|J^\top W\, r(z)\| < \epsilon. \tag{8}$$

Near the minimum the cost (6) is approximately quadratic, so the posterior is approximately Gaussian about z^* with information matrix Λ. This is the *Laplace approximation*, and the *posterior standard deviation* σ_i of any unknown is read from the diagonal of Λ^{-1}, from the same sparse

factorization the last Gauss–Newton step used:

$$p(z \mid y) \approx \mathcal{N}\big(z^*, \Lambda^{-1}\big), \tag{9}$$

$$\Lambda = J^\top W J, \tag{10}$$

$$\sigma_i = \sqrt{(\Lambda^{-1})_{ii}}. \tag{11}$$

When several factor graphs compete to explain the same readings — a leak here, a leak there, no leak — each is compared by its *evidence* $Z = \int \prod_k \exp(-\frac{1}{2} w_k r_k(z)^2)\, dz$, the probability of the readings under the model once the extra freedom of its unknowns has been paid for. In the Laplace approximation the evidence has a closed form, and normalizing every factor by its own weight lets graphs with different numbers of unknowns d and factors compare as densities,

$$\log Z = -C(z^*) + \tfrac{d}{2} \log 2\pi - \tfrac{1}{2} \log\det \Lambda + \sum_k \tfrac{1}{2} \log \frac{w_k}{2\pi}. \tag{12}$$

Differences of $\log Z$ between hypotheses, from (12), are in *nats*, natural-log units of odds; five nats is odds of about 150 to 1. The posterior probability of a hypothesis is its evidence normalized over all hypotheses under equal prior odds. Everything that follows is a choice of variables and factors on this graph, followed by these three operations.

For a Gaussian factor (4) is the weighted square (6) and the information (10) is its Gauss–Newton curvature; for a count row it is the Poisson term of Section 3.5.2, whose exact curvature at the mode enters the information matrix where a weight stands for a Gaussian row. Appendix F derives the chain from (5) to (12) in three steps of calculus, so that nothing here has to be taken on trust.

Four of the five questions below end in the same form. Each asks the probability of a condition on the state, and a condition defines an indicator, one where it holds and zero where it does not, so that the probability is the posterior expectation of the indicator,

$$P(h(z) \le b \mid E) = \int \mathbf{1}\big[h(z) \le b\big]\, p(z \mid E)\, dz, \tag{13}$$

with h the engineering quantity the question is about and b the bound it is held to. Under the Laplace posterior (9) and a linearization of h at the mode the expectation has a closed form on the two operations above and the gradient of h,

$$P(h(z) \le b \mid E) \approx \Phi\!\left(\frac{b - h(\hat{z})}{\sigma_h}\right), \tag{14}$$

$$\sigma_h^2 = g_h^\top \Sigma\, g_h, \tag{15}$$

$$g_h = \nabla h(\hat{z}), \tag{16}$$

with Φ the standard normal distribution function. A two-sided condition $a < h(z) < b$ is a difference of two such terms, and an equality $h(z) = b$ has no probability under a continuous posterior until a tolerance makes it two-sided. The interval that holds 90 % of the posterior, $h(\hat{z}) \pm 1.645\, \sigma_h$, is a reading of (14), the two values of b at which it is 0.05 and 0.95, and not a second method. Where the condition involves the equipment's configuration S as well as the state, the indicator is summed over the configurations inside the query, $A(z) = \sum_S P(S \mid z)\, I_{\text{succ}}(S)$, and (14) is then taken on A: two layers, the configurations averaged inside the query and the state averaged over the posterior. The first arm has h the full-load rise and b the allowable; the second h the service availability; the fourth h a sensitivity and b zero; the fifth h the availability change and b a threshold or zero.

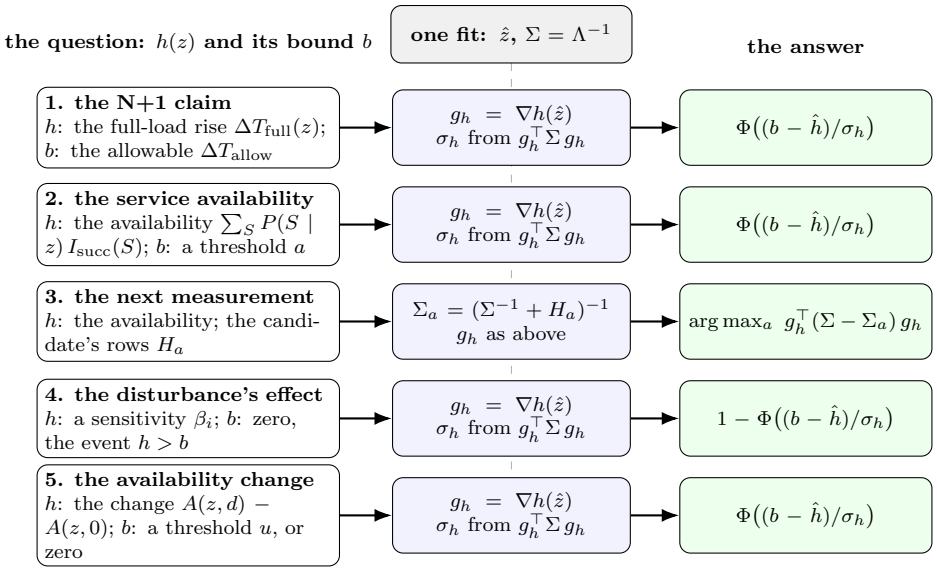


Figure 7: The one pattern. Four of the five questions ask the probability of a condition $h(z) \le b$ on an engineering quantity h of the state, and each is answered on the same two numbers from one fit, the mode $\hat{z}$ and the covariance Σ, through the gradient of h and the delta method, as Φ of a standardized margin (14); the 90 % intervals the results report are readings of the same Φ at 0.05 and 0.95. The third question is the exception: it asks which candidate takes most off σ_h^2, a maximization on the same blocks. A condition defines an indicator, a probability is the indicator's expectation, and the posterior decides how the expectation is evaluated: by Φ here, by the sum over configurations inside the availability query, by the grid or the draws in the references.

The third arm is the one exception: it asks not for a probability but for the candidate that takes most off σ_h^2, a maximization of $g_h^\top (\Sigma - \Sigma_a)\, g_h$ over the candidates on the same blocks. Figure 7 lays the five out. Figure 8 sets the operations beside the questions they answer and the plant and evidence they act on: the fit and its covariance, the query and its gradient, the delta method, a measurement's information update and the value it buys, and a probability or an evidence read from the posterior. Every arm below is a choice of query and evidence followed by some of these, and nothing else.

3.3 The factor graph of the trip test

The construction above is now written out for the trip test. The test takes unit B out, so its duty is zero and all of hall B's heat must cross the opening to unit A. With P_A and P_B the IT powers of the halls, Q_A the duty of the surviving unit and q the heat on the opening counted from A to B, all in watts, the heat balances at the two halls and the closure on the opening are

$$P_A - q - Q_A = 0, \tag{17}$$

$$P_B + q = 0, \tag{18}$$

$$q = G\,(T_A - T_B), \tag{19}$$

with G the opening's conductance in W/K and T_A, T_B the hall temperatures. With hall B hotter, q is negative and (18) says that hall B's whole load crosses the opening. The unknown the fit carries for the opening is $\log G$ rather than G, so that the conductance stays positive and its prior is a factor either way; the closure law (19) is written in G itself. These three rows are the physical law

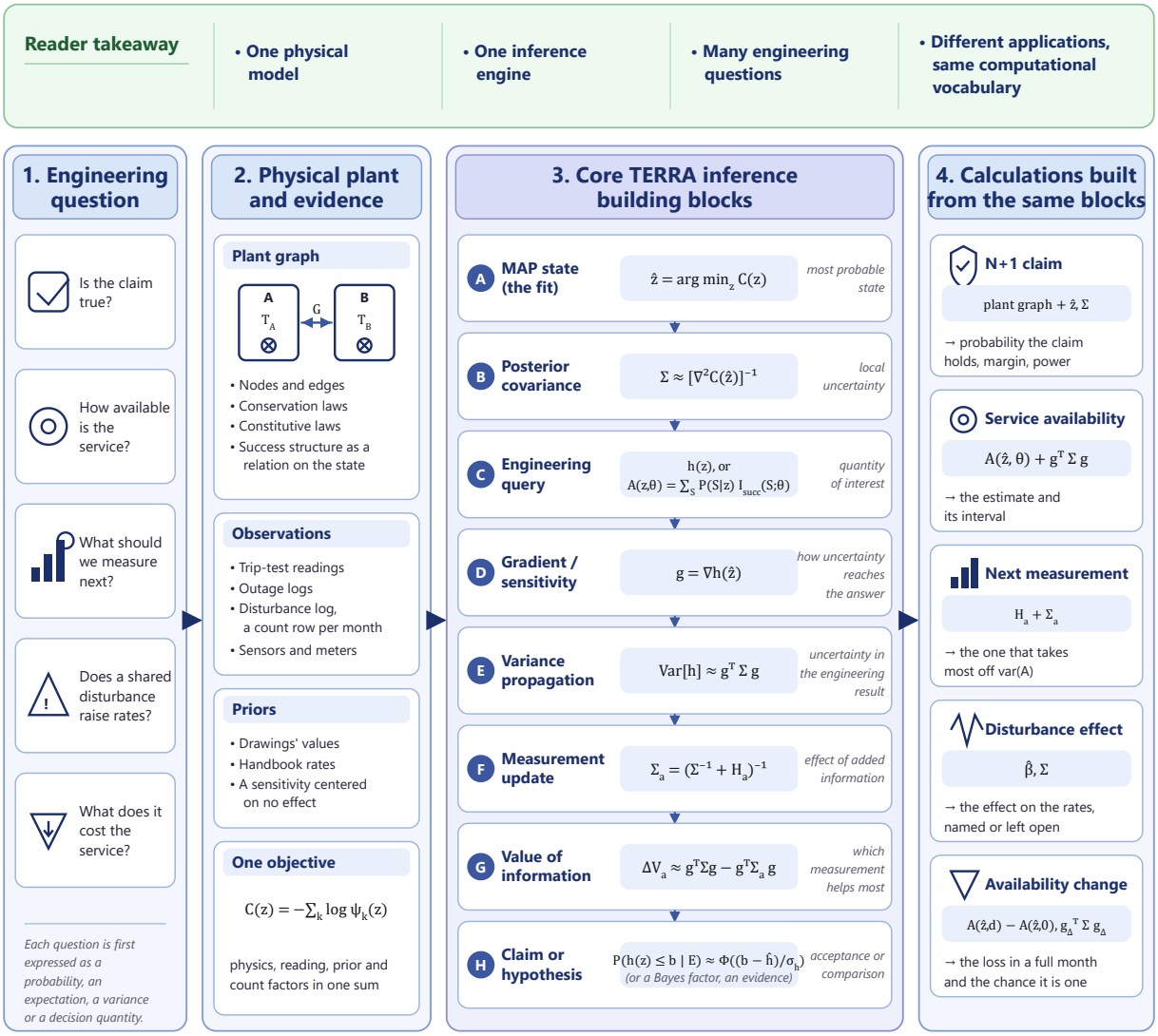


Figure 8: The building blocks. Across the top, the takeaway: one physical model, one inference, many engineering questions, the same computational vocabulary. Below it, left, the five questions, each first written as a probability, an expectation, a variance or a decision quantity; next, the plant graph with the observations and priors attached to it and the one objective they form; then the eight operations every arm invokes, the fit, its covariance, the query, its gradient, the delta method, a measurement's information update, the value of a measurement, and a probability or an evidence read from the posterior; and right, the calculation each arm builds from them. TERRA's contribution is not a formula per problem but the reduction of every question to these blocks.

of the test. The measurement model is one row per sensor: each hall sensor reads its temperature plus a constant offset bias plus noise, and the duty meter reads the surviving unit's duty plus noise,

$$y_A = T_A + b_A + \varepsilon_A, \tag{20}$$

$$y_B = T_B + b_B + \varepsilon_B, \tag{21}$$

$$y_Q = Q_A + \varepsilon_Q, \tag{22}$$

where b_A, b_B are the sensor biases in kelvin and the ε are zero-mean noise. Averaging N readings divides the noise's standard deviation by $\sqrt{N}$, so the averaged temperature readings carry the noise $\sigma/\sqrt{N}$ of 30 samples at 0.3 K; the duty and PDU readings carry their proportional noise. Each prior is a normal distribution, written $x \sim \mathcal{N}(m, s^2)$ and read "x is distributed as a normal with mean m and standard deviation s", under which x lies within s of m about two times in three. The priors are a lognormal law on the full-load rise centered at the allowable, $\ln \Delta T_{\text{full}} \sim \mathcal{N}(\ln \Delta T_{\text{allow}}, (\log 3)^2)$, a factor of 3 either way, which takes no side on the claim (2) and is the prior row on the opening's entry through (1); $T_A \sim \mathcal{N}(22, s_T^2)$ with $s_T = 2$ K; and $b_A, b_B \sim \mathcal{N}(0, s_b^2)$ with $s_b = 0.5$ K. The balance rows are trusted to 30 W and the reading rows to 0.0001 K, which are tolerances on the relations, not on the sensors. The energy tolerance is also a numerical choice: a row trusted a thousand times more tightly than the meters beside it makes the iteration of Section 3.2 crawl, which is why it is registered at 30 W rather than the 1 W first written, within two decades of the meter noise. What the readings determine about the conductance is the difference $y_B - y_A$ against the difference the opening's law (19) produces at the test load, and that difference has a floor set by the two sensors' noise and biases. Subtracting (20) from (21) cancels the set point, which enters both readings alike, and leaves the true difference plus the two biases and the two averaged noises,

$$y_B - y_A = (T_B - T_A) + (b_B - b_A) + (\varepsilon_B - \varepsilon_A), \tag{23}$$

and the spread of that difference about its true value is the floor. It is nothing the graph does not already hold: the delta method on the derived quantity $y_B - y_A$ under the graph before the readings are attached and with the conductance pinned, with Σ_0 that graph's covariance and g the gradient of $y_B - y_A$ in its state,

$$\sigma_{\Delta T}^2 = g^\top \Sigma_0 \, g. \tag{24}$$

By (23) the gradient is one on the reading of sensor B and minus one on that of sensor A, so (24) collects the two biases at s_b^2 each and the two averaged noises at σ^2/N each, plus the part the meters feed in through the balance rows: 0.712 K on the near case's 20 kW test, of which the bias and noise terms alone, $\sqrt{2\,(\sigma^2/N + s_b^2)}$, are 0.711 K. The registered floor is that hand value, and the closed forms of this report use it. The relative width of the conductance is this floor over the difference the test runs at, so it grows as the test load falls and as the opening widens; the floor is dominated by the bias term, which is why a longer test does little and a matched sensor pair does much. Collected, the unknowns the fit carries are one vector,

$$z = (T_A,\ T_B,\ q,\ \theta,\ Q_A,\ P_A,\ P_B,\ b_A,\ b_B), \tag{25}$$

$$\theta = \log G, \tag{26}$$

nine entries, and every relation above is one row, a residual $r_k(z)$ that is zero when the relation holds exactly, with a spread σ_k that says how far from zero it is allowed to sit. Table 2 lists the twelve rows one per line, with the residual, the spread and what the relation says. A reading row and a prior row have the same form, a quantity held near a value within a width; they differ only

Table 2: The twelve rows of the trip-test graph, one per line: the residual that is zero when the relation holds, the spread within which it may sit, and what the relation says. The physics rows are trusted to a tolerance, the reading rows to the sensor's noise, the prior rows to the stated width of what the plant believed. The prior on the opening is the unrestricted one of the fit, which takes no side on the claim; the two drawings' priors are stated in Appendix C.

kind	residual $r_k(z)$	spread σ_k	the relation
physics	$P_A - q - Q_A$	30 W	the balance at hall A
physics	$P_B + q$	30 W	the balance at hall B; unit B's duty is zero
physics	$q - e^{\theta}\,(T_A - T_B)$	30 W	the opening's law
reading	$y_A - (T_A + b_A)$	$\sigma/\sqrt{N} = 0.055$ K	thermometer A, through its bias
reading	$y_B - (T_B + b_B)$	$\sigma/\sqrt{N} = 0.055$ K	thermometer B, through its bias
reading	$y_Q - Q_A$	2 % of the reading	the duty meter of unit A
reading	$y_{P,A} - P_A$	1 % of the reading	the PDU meter of hall A
reading	$y_{P,B} - P_B$	1 % of the reading	the PDU meter of hall B
prior	$\ln \Delta T_{\text{full}} - \ln \Delta T_{\text{allow}}$	$\ln 3$, a factor of 3	the full-load rise $Q_{\text{hall}} e^{-\theta}$, unrestricted
prior	$T_A - 22$	2 K	the set point of hall A
prior	b_A	0.5 K	the bias of sensor A, near zero
prior	b_B	0.5 K	the bias of sensor B, near zero

in whether the value came from a sensor or from what the plant believed beforehand, which is why a drawing's value can be replaced by a measurement of it without changing the model. Figure 9 draws the graph these rows make. In the figure a square touches every circle its residual contains, so the opening's law reaches four unknowns and a meter reaches one; twelve rows constrain nine unknowns, and without the four prior rows the eight remaining would leave one direction free, the split of each reading into a temperature and a bias, which is the floor (24) in another form.

3.4 The probability of acceptance and the acceptance margin

The N+1 claim is signed off once, on one test, and then relied on. The question this arm answers is what rule turns one trip test into an acceptance of the claim, that the opening can carry one hall's load within the allowable rise, and how often that rule accepts a true claim and rejects a false one across the scatter of tests the plant could actually run. It matters because the test is run light, at a fraction of the hall's load, and the claim is about full load: a rule that works on the design case and fails on real readings accepts plants that will overheat, and one that is too strict refuses plants that are fine. In general terms, the claim is a region Θ_{claim} of the parameter θ, and the readings y are accepted as supporting it when the posterior puts enough probability there,

$$\text{accept}(y) \iff P(\theta \in \Theta_{\text{claim}} \mid y) \geq 1 - \alpha, \tag{27}$$

where the bar reads "given": the probability is computed from the readings of the test, not before them, and α is the risk of a false acceptance the owner will carry. The power of the test at a truth θ is the probability, over the readings that truth generates, that the rule accepts,

$$\text{power}(\theta) = P_{y|\theta}\big[\text{accept}(y)\big], \tag{28}$$

the quantity a test's design should be judged by: a rule can be right on the readings it was given and still be a poor test. Here the claim is the relation (2) on the full-load rise, a derived quantity of the state under the contingency: with unit B out and the full load applied, (18) and (19) give $\Delta T_{\text{full}} = Q_{\text{hall}}\, e^{-\theta}$ at a state, and the acceptance probability is the posterior mass of that rise

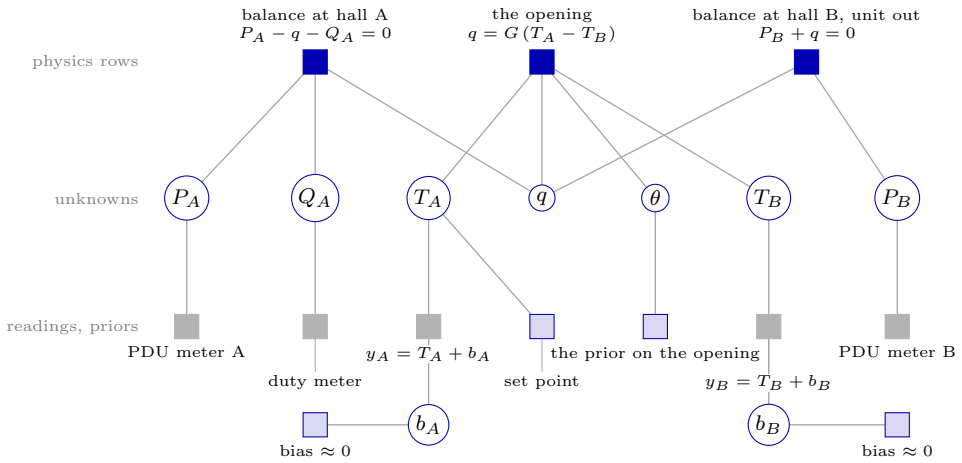


Figure 9: The factor graph of the trip test. Circles are the unknowns, the entries of (25); squares are the rows, each tied by a line to the unknowns its residual contains: the physics rows of (17) to (19) in blue, the reading rows of (20) to (22) and the two PDU rows in gray, the prior rows in light blue. Unit B's duty is pinned at zero by the trip and is not a circle. Every square carries the spread Table 2 gives it, and the fit weighs them all at once.

within the allowable, which is also the probability that H_{AND} is the plant's tree given the readings, independently of any drawing,

$$P_{\text{acc}} = P(\Delta T_{\text{full}} \le \Delta T_{\text{allow}} \mid y), \tag{29}$$

read against $1-\alpha$ of 0.95 and 0.99, with the posterior that of the unrestricted fit of Section 3.3. The probability is evaluated through TERRA; a quadrature outside TERRA, a normal approximation and a closed form are its references, and a count over simulated tests the check on what follows from it.

3.4.1 The evaluation through TERRA

Two numbers come out of this arm. For the test the plant ran, the probability (29) that the opening passes one hall's load. For a test not yet run, the *power* of the test: the probability that a test at load Q accepts the claim at a true capacity C, which is (28) with the truth written as the capacity,

$$\text{power}(C) = P\big(P_{\text{acc}} \ge 0.95 \,\big|\, C\big), \tag{30}$$

with P_{acc} the probability of acceptance a test's readings give and 0.95 the confidence a test accepts at. Figure 10 draws the two activities the power connects: one test and its rule, and the power with no test.

The closed form rests on two facts. The first: the probability of acceptance falls as the observed hall difference ΔT_{obs} rises, so the rule accepts exactly the tests whose difference lies below a critical value, one number per test load, defined by

$$P_{\text{acc}}(\Delta T_{\text{crit}}) = 0.95, \tag{31}$$

and found by bisection on the design-case readings with the difference set, fourteen evaluations per load. The second: over the tests that could happen, the observed difference is normal about the design-case value $\Delta T_{\text{test}}(C) = Q\,\Delta T_{\text{allow}}/C$, the test's difference at the true capacity by (1) and (3)

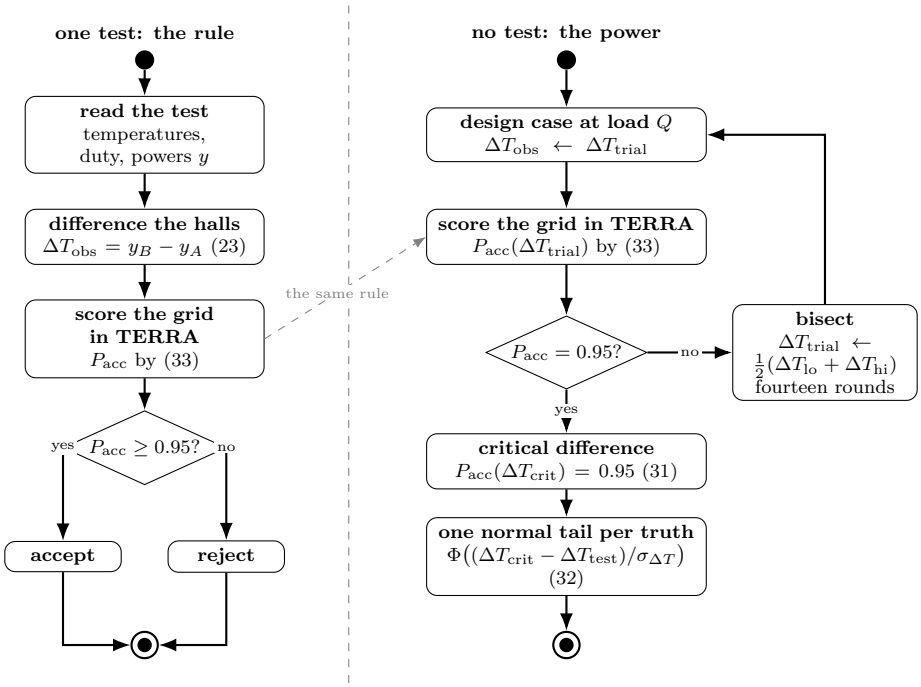


Figure 10: The steps from a test's readings to its acceptance, left, and from a test load to its power, right. Each lane runs from the filled start node to the ringed end node. Left lane, one test: the readings are differenced, the grid is scored in TERRA to the probability of acceptance (33), and the decision accepts at 0.95. Right lane, no test: the design-case readings at load Q with the difference set to a trial value are scored by the same rule, the dashed arrow, and the loop moves the trial difference until the probability of acceptance equals 0.95, fourteen rounds per load; the difference that stops the loop is the critical difference (31), and one normal tail with the floor as its spread turns it into the power (32) at every truth. No test is simulated in either lane; the simulated tests and the coverage are the checks of Section 3.4.2.

scaled by the test load, with the floor (24) as its spread, since by (23) the set-point offset cancels and the biases and the averaged noise add. The power is therefore one normal tail per truth,

$$\text{power}(C) = \Phi\left(\frac{\Delta T_{\text{crit}} - \Delta T_{\text{test}}(C)}{\sigma_{\Delta T}}\right), \tag{32}$$

with Φ the standard normal distribution function, the probability that a normal variable of mean zero and standard deviation one falls below its argument. No test is simulated.

What remains is P_{acc} itself, the posterior mass of the full-load rise within the allowable. On the differenced channel, with the nuisances folded into the floor, that posterior is a one-dimensional sum, and TERRA evaluates it with the discrete-hypothesis query, the one Appendix C also uses on the two drawings: 8,001 fixed full-load rises ΔT_i on a grid over eight prior widths, each a hypothesis that predicts the test's hall difference $(Q/Q_{\text{hall}})\,\Delta T_i$, scored on the observed difference with the

floor as the noise and the prior's mass on its cell as the weight,

$$P_{\mathrm{acc}} = \frac{\sum_{i:\,\Delta T_i \le \Delta T_{\mathrm{allow}}} \pi_i \, L_i}{\sum_i \pi_i \, L_i}, \tag{33}$$

$$L_i = \exp\!\left(-\frac{\left(\Delta T_{\mathrm{obs}} - (Q/Q_{\mathrm{hall}})\,\Delta T_i\right)^2}{2\,\sigma_{\Delta T}^2} \right), \tag{34}$$

$$\pi_i = \exp\!\left(-\frac{(\ln \Delta T_i - \ln \Delta T_{\mathrm{allow}})^2}{2\,s_G^2} \right), \tag{35}$$

with L_i the likelihood of the observed difference under hypothesis i and π_i the unrestricted prior of Table 2 at that rise, the lognormal law on the full-load rise centered at the allowable with spread s_G. The sum is the exact posterior to the resolution of the grid; it costs one evaluation of the query and no fit. On the design case, the readings the model predicts at the truth, (33) is a curve against the true capacity, one per test load, and the *acceptance margin* is the first registered grid cell at which the curve reaches a stated confidence. So (33) gives P_{acc} at any difference, (31) the critical difference and (32) the power at every truth: three equations, fourteen evaluations per load.

3.4.2 The references and checks

Three references stand beside (33). The first is the same quadrature computed outside TERRA, the likelihood of the observed difference at each of 8 001 full-load rises times the unrestricted prior, summed where the full-load rise is within the allowable; it agrees with (33) to the last digit, as it must. The second is the normal approximation. TERRA fits the full unrestricted graph of Section 3.3, the maximum-a-posteriori state (7) and the Laplace curvature (9) to (11) of all nine unknowns at once, so the posterior of the opening's entry θ, and with it of the logarithm of the full-load rise, is normal with mean m and standard deviation s, and the exceedance query returns the mass of the rise within the allowable, with $\widehat{\Delta T}_{\mathrm{full}} = Q_{\mathrm{hall}}\, e^{-m}$ the rise at the fit,

$$P_{\mathrm{acc}} = \Phi\!\left(\frac{\ln \Delta T_{\mathrm{allow}} - \ln \widehat{\Delta T}_{\mathrm{full}}}{s} \right), \tag{36}$$

whose argument is the number of standard deviations by which the fitted rise clears the allowable. This *Laplace rule* is the registered one, and the fit behind it supplies the posterior figures the report quotes; it undercounts the skewed posterior at low test load, which is why the exact sum decides where the two part. Its own critical difference and power follow by (31) and (32) with the tail as the rule, fourteen fits per load. The third reference is a closed form. On the design case the fit returns the truth, whose full-load rise is $\Delta T_{\mathrm{true}} = Q_{\mathrm{hall}}\,\Delta T_{\mathrm{allow}}/C$ by (1) and (3), and the clearance of the allowable and the spread s take simple forms: the test's own difference is that rise scaled by the load, and the reading determines the rise's logarithm with the spread of the difference relative to the difference itself,

$$\ln \Delta T_{\mathrm{allow}} - \ln \widehat{\Delta T}_{\mathrm{full}} = \ln(\Delta T_{\mathrm{allow}}/\Delta T_{\mathrm{true}}), \tag{37}$$

$$s = \sigma_{\Delta T}/\Delta T_{\mathrm{test}}, \tag{38}$$

$$\Delta T_{\mathrm{test}} = \frac{Q}{Q_{\mathrm{hall}}}\,\Delta T_{\mathrm{true}}. \tag{39}$$

Substituting both into the tail (36) gives the Gauss–Newton closed form of the probability of acceptance,

$$P_{\mathrm{GN}} = \Phi\!\left(\frac{\Delta T_{\mathrm{test}}\, \ln(\Delta T_{\mathrm{allow}}/\Delta T_{\mathrm{true}})}{\sigma_{\Delta T}} \right), \tag{40}$$

the test's own difference times the log clearance of the allowable, over the floor. It is reported beside the exact curve as a check on the approximation: its argument goes as $x \ln(1/x)$ in $x = \Delta T_{\text{true}}/\Delta T_{\text{allow}}$ and peaks at $x = 1/e$, so a probability of acceptance that falls as the true rise shrinks further is the approximation's symptom, not a property of the test.

Two checks close the arm. The power (30) was registered as a count: 200 tests per truth and load, each a fresh draw of the readings scored by the exact rule, and the fraction that accept. That count checks the closed form (32), which must reproduce it within the binomial resolution of 200 draws; the Laplace rule's closed form is checked the same way against its own count at one truth per load, 200 realized tests each fitted through TERRA and scored by its tail (36). The posterior behind the Laplace rule is checked for calibration by its coverage, the fraction of 200 draws per test load, each with the truth drawn from the prior and fitted afresh, whose 90 % interval on the full-load rise, taken on its logarithm, contains the truth,

$$\text{coverage} = \frac{\#\{\text{draws whose } 90\,\% \text{ interval contains the truth}\}}{n}, \tag{41}$$

read against the registered band with the binomial interval of the count, the uncertainty a fraction estimated from 200 yes-or-no outcomes carries.

3.4.3 The ranking of the next test

When a test cannot tell, the next test is the question: of the probes the plant could run, the one that adds most information on the conductance (Lindley 1956), ranked by the expected information gain, the uncertainty a probe is expected to take off the posterior. This ranking is by information on the parameter; the ranking of Section 3.6 is by the variance taken off the answer, and the two are compared there. The entropy of a normal posterior with covariance Σ, its uncertainty volume in nats, is $\frac{1}{2} \log\det(2\pi e\, \Sigma)$; information adds, so after a probe that adds the information H_a the covariance is $(\Sigma^{-1} + H_a)^{-1}$, and the expected information gain is the entropy taken off,

$$\text{EIG}_a = \tfrac{1}{2} \log\det \Sigma - \tfrac{1}{2} \log\det \left(\Sigma^{-1} + H_a\right)^{-1}, \tag{42}$$

$$= \tfrac{1}{2} \log\det \left(I + \Sigma\, H_a\right), \tag{43}$$

where I is the identity matrix and det the determinant, so that the gain is, in nats, the expected shrinkage of the uncertainty volume of the parameters. Two matrices enter. The posterior covariance Σ of the parameters, the matrix of their variances and of how they vary together, is the inverse of the information matrix Λ of (10) from the unrestricted fit; the Fisher information H_a a probe a adds is the same construction on the probe's own readings alone,

$$H_a = J_a^{\top} W_a\, J_a, \tag{44}$$

with J_a the matrix of the derivatives of the probe's readings with respect to the parameters, one row per reading, and W_a the diagonal matrix of the inverse variances of those readings' noise: the curvature the probe's rows would add to the fit, computed without adding them. Every probe here reads one quantity, the hall difference a trip test at load Q_a would show, which the closure (19) with unit B out predicts as the fitted full-load rise scaled by the load; so J_a is the derivative of that prediction in the opening's entry θ at the fit, and W_a the inverse of the floor (24) under the probe's own graph, the pre-reading covariance built with the probe's sample count N_a and bias prior $s_{b,a}$. TERRA evaluates (43) on the posterior of the unrestricted fit, with each probe described to it by those two numbers, so that H_a is the information matrix of rows that were never added to the graph. The probes are a longer test, a hotter test, and a matched pair of sensors whose common

bias cancels in the difference; each changes one number in (44): the longer test raises N_a, the hotter tests raise Q_a, the matched pair lowers $s_{b,a}$. Each carries the hall-B rise it causes and the margin that leaves to the allowable, because a probe that breaches the allowable is not a test the plant can run.

3.5 The outage record

The service availability of Section 2 is a function of the components' unavailabilities, and those come from a handbook as one number each, a point with no width. The question this arm answers is how wide the service availability is once the plant's own outage record E, k failures in T operating hours per component and the repair hours they cost, tempers the handbook. It matters because a point value cannot say whether the plant is at 0.99 or 0.9999, and the two call for different decisions; the recommended practice itself puts a 12 % chance on the field rate exceeding twice the handbook value (IEEE 2025, §6.13.3). The answer is the posterior of the service availability given the record,

$$P(A_{\text{service}} \le a \mid E) = \int \mathbf{1}\big[A(\lambda,\mu) \le a\big]\, p(\lambda,\mu \mid E)\, d\lambda\, d\mu, \tag{45}$$

the posterior probability of every set of rates at which the service availability falls at or below a, summarized as the interval that holds 90 % of it, the band the owner acts on; A stands for R_{service} through this section.

3.5.1 The service availability on the plant graph

The service availability is a query on the plant graph of Section 3.3, put at the design condition: the full IT load Q_{hall} in each hall, each hall's temperature counted as a rise above the set point an available unit holds in its own hall, and the duties as free controls. An outage configuration S is the triple of availability indicators s_A, s_B and s_F, each one when that unit or the feeder is available; under S the physical state, the two hall temperatures, the heat on the opening and the two duties, must satisfy the balances (17) and (18) with both duties free, the closure (19), the allowable envelope on both halls, a rise of at most ΔT_{allow} above the set point, and the duty limits each unit's availability and the feeder's impose on nonnegative duties,

$$Q_A \le s_A\, s_F\, Q_{\max}, \tag{46}$$

$$Q_B \le s_B\, s_F\, Q_{\max}. \tag{47}$$

A configuration is feasible when a physical state satisfies the balances, the closure, the envelope and the limits (46) and (47) at once, and the success indicator is existential,

$$I_{\text{succ}}(S;\theta) = \begin{cases} 1 & \text{a feasible physical state exists under } S, \\ 0 & \text{otherwise,} \end{cases} \tag{48}$$

with θ the physical parameters, here the conductance. For this plant the balances solve by hand and the indicator has a closed form,

$$I_{\text{succ}}(S;\theta) = s_F \Big[s_A s_B + (s_A + s_B - 2 s_A s_B)\, \mathbf{1}[\Delta T_{\text{full}} \le \Delta T_{\text{allow}}] \Big], \tag{49}$$

the feeder's indicator multiplying the whole bracket: both units available deliver, one available delivers when the relation (2) holds, and a feeder outage loses everything, with either unit or both; Figure 11 draws its two values as fault trees. In this implementation the query is answered by

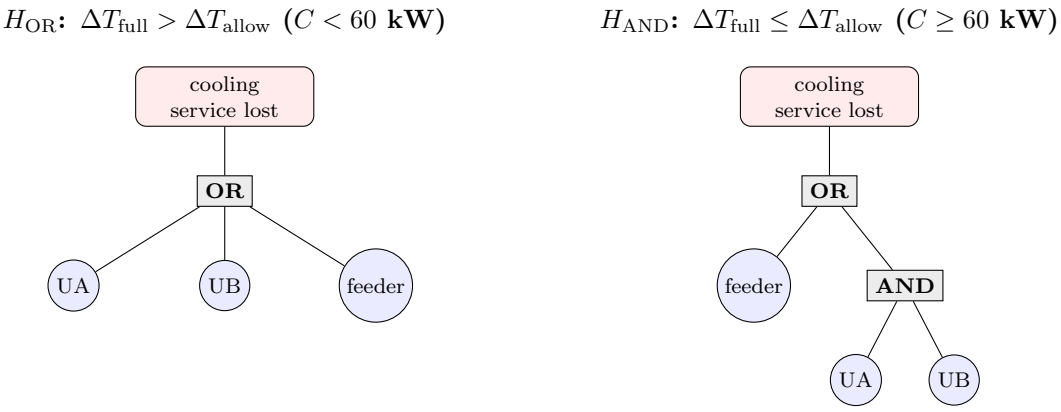


Figure 11: The two values of the success structure as fault trees. Left, the opening cannot pass one hall's load: the top event is the OR of three basic events, either unit or the feeder, and the units are a series pair. Right, the opening passes the load: the units sit under an AND gate beside the feeder, a parallel pair. Each tree carries its validity condition, the full-load rise (2) against the allowable, and the trip test is a test of that condition.

(49), the hand solution of the balances, and the graph is not re-solved per configuration; a plant whose balances have no closed form would re-solve it, with (48) unchanged. The indicator does not depend on the rates. The rates enter through the probability of each configuration: a component is down for the fraction of time its mean down period takes of its mean cycle, $(1/\mu)/(1/\lambda + 1/\mu)$, which rearranges to the unavailability

$$q = \frac{\lambda/\mu}{1 + \lambda/\mu}, \tag{50}$$

and the components fail independently of one another, so that the service availability at a vector of unavailabilities q is the enumeration of every configuration, its probability from the q's and its indicator,

$$A(q) = \sum_{S} P(S \mid q)\, I_{\mathrm{succ}}(S;\theta), \tag{51}$$

$$P(S \mid q) = \prod_{i \text{ out}} q_i \prod_{i \text{ in}} (1 - q_i). \tag{52}$$

The unknowns the service availability depends on are therefore the six rates, two per component, gathered in the vector x of their logarithms, a log failure rate $\theta_i = \log \lambda_i$ and a log repair rate $\phi_i = \log \mu_i$ per component, so that each rate stays positive and its prior is a factor either way. They are the inputs the industry's recommended practice asks for (IEEE 2025, §6.3), under its four stated assumptions, a constant failure rate, an exponentially distributed repair time, each failure event independent of every other, and up times much larger than down times; the third is what the common-cause arm tests. The physics enters the arm through the feasibility query alone: the rates are learned component by component, from counts that know nothing of the plant, and the graph is what couples them, through the configurations (48) declares feasible.

3.5.2 The observation model

A component's record is k failures over T operating hours and m repairs over the D hours of downtime those failures cost. Failures that arrive at a steady rate λ, each independent of the last,

give a count over T hours that is Poisson with mean λT: the probability of exactly k is

$$p(k \mid \lambda) = \frac{(\lambda T)^k \, e^{-\lambda T}}{k!}, \tag{53}$$

whose negative logarithm is $\lambda T - k \log \lambda + \log k!$. Written in the log-rate θ, this negative logarithm, the quantity the fit minimizes, is

$$-\log p(k \mid \theta) = T e^{\theta} - k\,\theta + \text{const}, \tag{54}$$

where const is $\log k!$, which does not depend on θ. The repair log is the same observation on the repair rate, m completed repairs over D hours of downtime, Poisson with mean μD, and gives the same form in ϕ with D and m in place of T and k. Each count row is local to its component: it constrains that component's rate and no other. Before the record, each log-rate has a normal prior centered on the handbook's value, the failure rate λ_0 and the repair rate μ_0, one over its mean time to repair, rates of the kind IEEE Std 3006.8 tabulates (IEEE 2018), with spread $s = \ln 3$, a factor of three either way. Each component's record is therefore a graph of its own with two unknowns and four rows, a prior row and a count row on each log-rate, and the misfit the fit minimizes, the negative log of the four factors' product, is

$$C(\theta, \phi) = \tfrac{1}{2}\frac{(\theta - \log \lambda_0)^2}{s^2} + \tfrac{1}{2}\frac{(\phi - \log \mu_0)^2}{s^2} + \big(T e^{\theta} - k\,\theta\big) + \big(D e^{\phi} - m\,\phi\big). \tag{55}$$

No row holds both unknowns, so the two separate.

3.5.3 The inference through TERRA

TERRA's Newton iteration (7) finds the mode where the gradient of (55) vanishes,

$$T e^{\hat{\theta}} - k + \frac{\hat{\theta} - \log \lambda_0}{s^2} = 0, \tag{56}$$

$$D e^{\hat{\phi}} - m + \frac{\hat{\phi} - \log \mu_0}{s^2} = 0, \tag{57}$$

each a balance: the failures the current rate predicts against the failures the log holds, with the prior pulling toward the handbook at a strength $1/s^2$, less than one failure's worth at a factor of three. The fitted rates are $\hat{\lambda} = e^{\hat{\theta}}$ and $\hat{\mu} = e^{\hat{\phi}}$; with a weak prior, $\hat{\lambda} \approx k/T$, the observed rate. The information matrix (10) is the curvature of (55) at the mode: the prior rows contribute their weight $1/s^2$, the count rows their expected count,

$$\Lambda = \begin{pmatrix} 1/s^2 + T e^{\hat{\theta}} & 0 \\ 0 & 1/s^2 + D e^{\hat{\phi}} \end{pmatrix}, \tag{58}$$

diagonal because no row couples the two. The Laplace posterior (11) is its inverse, and the spread of the log failure rate is

$$\sigma_{\theta} = \frac{1}{\sqrt{T\hat{\lambda} + 1/s^2}}, \tag{59}$$

which is about one over the root of k, since at the mode the predicted count $T\hat{\lambda}$ is close to k, and the spread of the log repair rate is likewise about one over the root of m: a record of k failures

pins the rate to a relative precision of about one over the square root of k, four failures to 50 %, twenty-five to 20 %. A record with no failure is not empty: its count row holds Te^{θ} alone, which pushes the rate down, so the posterior's upper tail is set by the hours, a rate of more than a few per T being ruled out, while its lower tail is the prior's. The three components' graphs share no unknown, so the covariance of all six log-rates is the diagonal of their spreads,

$$\Sigma = \operatorname{diag}\big(\sigma^2_{\theta,\mathrm{UA}},\ \sigma^2_{\phi,\mathrm{UA}},\ \sigma^2_{\theta,\mathrm{UB}},\ \sigma^2_{\phi,\mathrm{UB}},\ \sigma^2_{\theta,\mathrm{F}},\ \sigma^2_{\phi,\mathrm{F}}\big). \tag{60}$$

3.5.4 The return to the service availability

The service availability at the fitted rates is $\hat{A} = A(\hat{q})$ by (51) at $\hat{q}_i = q(\hat{\lambda}_i, \hat{\mu}_i)$, one enumeration. Its spread follows from the covariance by expanding A to first order about $\hat{x}$, $A(x) \approx \hat{A} + g^{\top}(x - \hat{x})$ with g the gradient of A with respect to the six log-rates, which makes A a linear function of a normal vector, and the variance of a linear function of a normal vector with covariance Σ is

$$\sigma^2_A = g^{\top}\,\Sigma\,g, \tag{61}$$

the delta method of error propagation, the operation that gave the floor (24) and the spread (38) of the acceptance arm. The gradient comes from the enumeration already made. A is linear in each q_i on its own, because every configuration has component i either out or in,

$$A = q_i\,A(i \text{ out}) + (1 - q_i)\,A(i \text{ in}), \tag{62}$$

so its derivative with respect to q_i is minus the difference of the two conditional values, the Birnbaum importance (Birnbaum 1969),

$$I_B(i) = A(i \text{ in}) - A(i \text{ out}), \tag{63}$$

the change in the service availability when component i goes from surely out to surely available; both conditional values are sums over the configurations already enumerated, split on component i, so (63) costs nothing more. The chain rule through (50) needs the derivatives of the unavailability with respect to the two log-rates,

$$\frac{dq}{d\log\lambda} = q\,(1 - q), \tag{64}$$

$$\frac{dq}{d\log\mu} = -q\,(1 - q), \tag{65}$$

and gives the two entries of the gradient per component, the derivatives of A with respect to its log failure rate and its log repair rate at the fitted unavailability,

$$g_{\lambda,i} = -I_B(i)\,\hat{q}_i\,(1 - \hat{q}_i), \tag{66}$$

$$g_{\mu,i} = +I_B(i)\,\hat{q}_i\,(1 - \hat{q}_i), \tag{67}$$

six numbers in all. With $\hat{A}$ and σ_A the posterior of the service availability is normal to first order, and (45) is the pattern (14) with $h = A$ and $b = a$,

$$P(A \le a \mid E) \approx \Phi\left(\frac{a - \hat{A}}{\sigma_A}\right), \tag{68}$$

and the interval that holds 90 % of it, $\hat{A} \pm z\,\sigma_A$ with $z = 1.645$ the 95 % quantile of the standard normal, is its reading at 0.05 and 0.95, not a second method. One enumeration per structure and record, its gradient, the fits' spreads, and no draw. Two approximations sit in it beyond the Laplace form of the rates: the linearization (61), and the symmetric normal (68), which does not know that A is at most one and whose upper end passes one when σ_A is of the order of $1 - \hat{A}$, as it is for the parallel pair on a record of a few failures. Both fail gently as the record grows, and the exact reference of Section 3.5.6 checks both.

3.5.5 The structure left open

The indicator (49) has two values, the AND tree and the OR tree of Figure 11, and which is in force is the claim (2), whose probability the trip test gave. A structure is *settled* when that probability (29) is effectively one or zero, as the far case's tests leave it, and *unsettled* when it is between, as the near case's 20 kW test does; the branch the two cases make is the decision of Figure 12. When the structure is unsettled, the service availability is one of two quantities, and which one is itself uncertain: with probability p the relation holds and the plant's tree is the AND tree, in which the service is lost only when both units are out, and with probability $1 - p$ it fails and the tree is the OR tree, in which either unit's outage loses the service,

$$A = \begin{cases} A_{\mathrm{AND}} & \text{with probability } p, \\ A_{\mathrm{OR}} & \text{with probability } 1 - p, \end{cases} \tag{69}$$

where, with the feeder outside the boundary, the enumeration (51) at any q gives

$$A_{\mathrm{AND}} = 1 - q_{\mathrm{A}}\, q_{\mathrm{B}}, \tag{70}$$

$$A_{\mathrm{OR}} = (1 - q_{\mathrm{A}})(1 - q_{\mathrm{B}}), \tag{71}$$

and with the feeder inside it both are multiplied by $1 - q_{\mathrm{F}}$. The weight p is the claim's probability (29), the posterior mass of the full-load rise within the allowable on the test's readings,

$$p = P_{\mathrm{acc}}, \tag{72}$$

which is 0.534 on the near case's 20 kW test (Section 4.1). The third arm takes the same weight in its Laplace form (36), since its preposterior term is built on the Gaussian posterior (Section 3.6.1). The registered comparison of the forward study's two drawings, Appendix C, gives a weight of its own, the posterior probability of the wide drawing over the narrow one; it answers which drawing the plant is, not whether the relation holds, and the band does not use it. The posterior of A is then the mixture of the two normal forms (68), one per structure, each with its own enumeration and spread,

$$P(A \le a \mid E, y) \approx p\,\Phi\left(\frac{a - \hat{A}_{\mathrm{AND}}}{\sigma_{\mathrm{AND}}}\right) + (1 - p)\,\Phi\left(\frac{a - \hat{A}_{\mathrm{OR}}}{\sigma_{\mathrm{OR}}}\right), \tag{73}$$

and the ends of its 90 % interval are the values of a at which (73) reaches 0.05 and 0.95.

3.5.6 The exact reference

The exact posterior of each log-rate is the same prior and likelihood multiplied on a grid of values of θ and normalized to sum to one, with its quantiles, the values below which a stated fraction of the probability lies, read from the running sum. It is computed outside the Laplace approximation

because a count of a few failures gives a skewed posterior, with a longer tail toward high rates than toward low. Draws of the six rates from it, turned into unavailabilities by (50) and pushed through the enumeration (51) one draw at a time, give the posterior of the service availability with no linearization and no normal form, and the 5–95 % of those draws is the exact band. Draws from the Laplace posterior pushed through the same enumeration isolate the first approximation, the normal form of the rates, which the registration bounds by the ratio of the two bands' endpoints in unavailability. A draw that puts an unavailability at or above one half lies outside the enumerator's domain and is counted, never clipped. While the structure is unsettled each draw is assigned to the AND tree with probability p and to the OR tree otherwise and takes the value (70) or (71) accordingly, the exact form of (73). The interval (68) is the arm's derivation and the exact band is its check; where the two part, at the parallel pair's upper end on a short record, the check governs, and Section 4.2 reports both.

3.6 The value of a measurement

A test, a sensor or another year of record each costs money and each narrows something; the question this arm answers is which of them narrows the number the owner acts on, the service availability, and by how much, before it is bought. It matters because a measurement can be informative about a parameter the service availability barely depends on, so a ranking by information alone can send the budget to the wrong instrument: the Fisher information H_a of (43) measures what a candidate adds to the parameter vector θ of the conductance and the six rates, not what it takes off the answer. In general terms, the value of a candidate measurement a is the expected reduction in the variance of the answer A once its outcome y_a is in hand, taken before the outcome is known,

$$V_a = \operatorname{var}(A) - \mathbb{E}_{y_a}\big[\operatorname{var}(A \mid y_a)\big], \tag{74}$$

the preposterior analysis of Raiffa and Schlaifer (1961), with $\mathbb{E}_{y_a}[\cdot]$ the average over the readings the candidate might return. Here A is the service availability at the fitted rates, and under a normal posterior the variance after the candidate does not depend on which outcome it returns, so the average drops and the value is a difference of two variances, one quadratic form each when the structure is settled and a quadrature over the structure probability when it is not. This arm has no closed-form reference; its comparison is with the ranking by information of Section 3.6.2.

3.6.1 The evaluation through TERRA

For each candidate a, a test, a sensor or a year of record, its value to the owner is the variance it takes off the service availability, written A in this section, (74) with its average dropped,

$$V_a = \operatorname{var}(A) - \operatorname{var}_a(A), \tag{75}$$

with $\operatorname{var}(A)$ the variance of the band as it stands and $\operatorname{var}_a(A)$ the variance after the candidate's information is added, before the candidate's outcome is known. One calculation serves whether the structure is settled or still open, and Figure 12 lays it out with the loop over the candidates. Under one structure S, the AND tree or the OR tree, A depends on the parameter vector θ of the conductance and the six log-rates; expanding it about the posterior mean $\bar{\theta}$ to first order, $A(\theta) \approx A(\bar{\theta}) + g_S^{\top}(\theta - \bar{\theta})$ with g_S the gradient of A under that structure, makes A a linear function of a normal vector with covariance Σ, and the variance of a linear function of a normal vector is

$$\operatorname{var}_S(A) = g_S^{\top}\, \Sigma\, g_S, \tag{76}$$

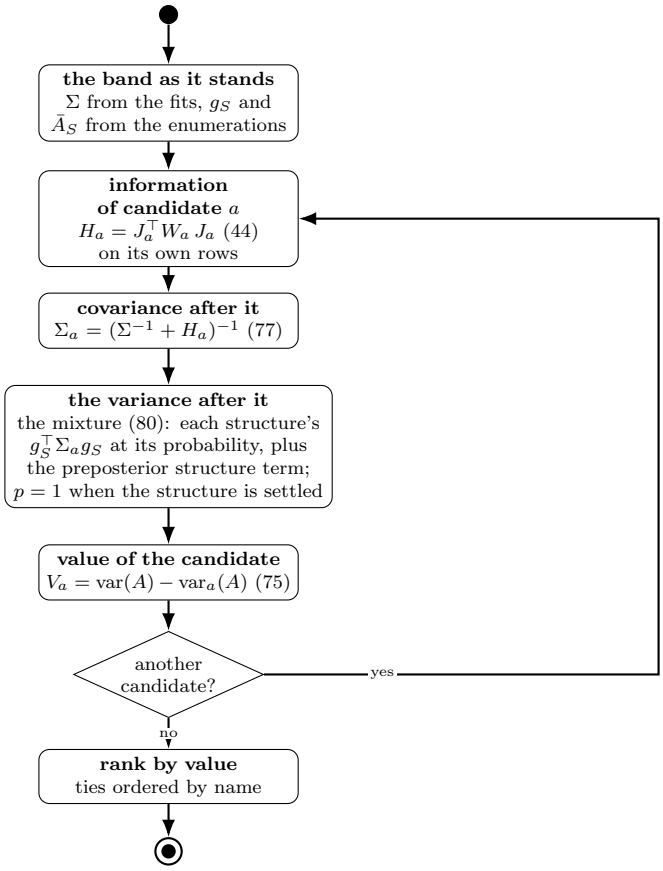


Figure 12: The steps from the band as it stands to the ranking of the candidate measurements. The band is assembled once from the three engine outputs; for each candidate its information (44) is added to the posterior's by (77), and the variance after it is the mixture (80) with the preposterior structure term, one calculation for both regimes: at $p = 1$, the structure settled, it is the AND tree's variance alone. The loop closes over the candidates and the ranking is by (75).

where $g_S^\top$ is the row form of g_S, so that the product is a single number; this is the delta method of error propagation, the one (61) used for the record's band. Information adds: the candidate's rows add their information matrix H_a to the posterior's, so the covariance after it is

$$\Sigma_a = \left(\Sigma^{-1} + H_a\right)^{-1}, \tag{77}$$

and the variance under S after the candidate is (76) with that covariance,

$$\mathrm{var}_{S,a}(A) = g_S^\top \, \Sigma_a \, g_S; \tag{78}$$

the gradient does not move, because the candidate's outcome is unknown and the expansion stays about the present mean. Three TERRA outputs enter. The covariance Σ is diagonal, with the trip-test fit's Laplace spread on the opening's entry by (11) and the record fits' spreads on the six log-rates by (60). A candidate's information H_a is (44) on its own rows, evaluated at the present fitted state: a trip test's reading rows on the opening's entry of the seven-parameter matrix, a year of record's count rows on a component's log failure rate and log repair rate; the rows each class adds are the ones Section 3.4.3 and Section 3.5.2 built, and no further form of H_a is needed here. The gradient g_S is the Birnbaum form (66) under structure S, read off that structure's enumeration; the service availability is a step in the conductance where the full-load rise crosses

the allowable and has no gradient there, so the conductance enters only through the probability that the relation (2) holds, the AND tree, the last piece. This arm writes that probability p and takes it as the Laplace form (36) of the claim's probability on the test's readings, since the structure term after a candidate, below, is built from the fitted mean and spread that (36) rests on; the band of Section 3.5.3 uses the exact form (72), and the two part only by the Laplace tail where the test is weak. The covariance update below is a local, expected-information approximation: it takes the candidate's Jacobian and the availability gradient as fixed at the present mean.

The band is a mixture over the structure, the AND tree with probability p and the OR tree with $1-p$, and the law of total variance, that the variance of a mixture is the average of the variances plus the variance of the averages, gives

$$\operatorname{var}(A) = p\operatorname{var}_{\mathrm{AND}}(A) + (1-p)\operatorname{var}_{\mathrm{OR}}(A) + p(1-p)\,(\bar{A}_{\mathrm{AND}} - \bar{A}_{\mathrm{OR}})^2, \tag{79}$$

whose first two terms are (76) under each structure, each with its own gradient on its own enumeration, and whose last term is the structure term, with $\bar{A}$ the mean service availability under each structure. After a candidate the two structures' variances are (78), and the structure term needs the structure probability the candidate will leave, which depends on readings not yet taken; its expectation over those readings, the preposterior expectation $\mathbb{E}[p'(1-p')]$ (Raiffa and Schlaifer 1961), is a one-dimensional quadrature over the future posterior mean of the opening's entry, written out in Appendix D. A candidate that adds nothing on the conductance, a year of record, leaves the term at $p(1-p)$. The variance after the candidate is then

$$\operatorname{var}_a(A) = p\operatorname{var}_{\mathrm{AND},a}(A) + (1-p)\operatorname{var}_{\mathrm{OR},a}(A) + \mathbb{E}[p'(1-p')]\,(\bar{A}_{\mathrm{AND}} - \bar{A}_{\mathrm{OR}})^2, \tag{80}$$

(79) with (78) in each structure's term and the preposterior expectation in the structure term, and the value (75) is the difference of (79) and (80). No case distinction is needed: when the structure is settled, $p = 1$ as the far case's full-load test leaves it, the structure term vanishes and the two reduce to the AND tree's variance alone,

$$\operatorname{var}(A) = \operatorname{var}_{\mathrm{AND}}(A), \tag{81}$$

$$\operatorname{var}_a(A) = \operatorname{var}_{\mathrm{AND},a}(A), \tag{82}$$

the fixed-structure form, in which no trip test can move the band since the conductance carries no gradient; when the structure is open, $0 < p < 1$, the structure term is what a trip test takes off, since it alone moves p. One formula, and the two regimes of Section 4.3 are its two values of p.

The fits give Σ, the enumerations g_S and the two $\bar{A}$, the candidate's rows H_a; (76) and (78) give the variance under a structure before and after, (79) to (80) the band's, with (81) its settled case, and (75) the value. One approximation sits in it, the linearization of (76), the same one the record's closed-form band carries and checks. The candidates are ranked by (75), with reductions that are round-off stated as bounds and ties ordered by name; the arm has no exact reference, so its comparison is with the ranking by information alone, next.

3.6.2 The ranking by information gain

The alternative the report compares with is the ranking by the information a candidate adds on the parameters, (43), the quantity a design-of-experiments criterion would maximize. It is reported beside the variance ranking in Section 4.3, and the two differ because a parameter can be narrowed without moving the number the owner acts on.

3.7 The shared disturbance

Every block diagram multiplies the components' availabilities, which assumes that one failure says nothing about the next; the recommended practice states the assumption outright, each failure event independent of any other (IEEE 2025, §6.3), and reports what its failure costs, an actual double-circuit supply failing more than 200 times as often as two independent sources would (IEEE 2025, §6.13.2). The question this arm answers is whether a disturbance the site shares, the months in which its utility supply ran outside its normal band, raises each component's failure rate, and by how much. It matters because a shared disturbance is the one failure a parallel pair cannot survive by redundancy, since it raises both units' rates at once, so it is the input a redundancy claim is most sensitive to; a common-cause fraction, the beta factor of the practice, is asserted where one is used, not inferred from the plant's own record. The quantity is, for each component, the posterior probability that the disturbance raises its rate; what the disturbance does to the service availability is the fifth arm's question, Section 3.8, on the posterior this arm fits.

3.7.1 The rate model under the disturbance

A component's failure rate under the disturbance is its baseline rate times a factor that grows with the severity,

$$\lambda_i(d) = \lambda_{i,0}\, e^{\beta_i d}, \tag{83}$$

where $d \in [0,1]$ is the month's disturbance index, a normalized measure of the stress the site logged in the month, zero for the ordinary state and one for the declared full disturbance, $\lambda_{i,0}$ the rate at $d = 0$ and β_i the component's sensitivity: e^{β_i} is the factor by which a full disturbance month multiplies its rate, one when the disturbance does not act on it. The index is an intensity, not a fraction of hours: a month split between two conditions would carry the hours-weighted mean of two rates, which (83) is not, and the record is drawn and read under the one model. The exponential keeps the rate positive at any sensitivity. The repair rate μ_i is taken constant, a repair taking as long under the disturbance as outside it. The unavailability at the severity follows by (50),

$$q_i(d) = \frac{\lambda_i(d)}{\lambda_i(d) + \mu_i}, \tag{84}$$

and the configuration probabilities by (52) at these $q_i(d)$, the components independent given the severity. The unknowns are therefore nine, three per component,

$$z = \big(\alpha_{\mathrm{A}}, \beta_{\mathrm{A}}, \gamma_{\mathrm{A}},\ \alpha_{\mathrm{B}}, \beta_{\mathrm{B}}, \gamma_{\mathrm{B}},\ \alpha_{\mathrm{F}}, \beta_{\mathrm{F}}, \gamma_{\mathrm{F}}\big)^{\top}, \tag{85}$$

with $\alpha_i = \log \lambda_{i,0}$ and $\gamma_i = \log \mu_i$, so that each rate stays positive and its prior is a factor either way. Conditional on the severity the components fail independently, as the recommended practice assumes; the dependence the practice's beta factor asserts arises here from the severity they share. Averaged over the months of the schedule, at the frequency w_d of each severity, the probability that both units are out at once is

$$\overline{q_{\mathrm{A}} q_{\mathrm{B}}} = \sum_d w_d\, q_{\mathrm{A}}(d)\, q_{\mathrm{B}}(d), \tag{86}$$

which exceeds the product of the two averaged unavailabilities whenever both sensitivities are positive: the common cause of the practice, written as a shared explanatory variable with an inferred effect rather than as an asserted fraction.

3.7.2 The disturbance log

The record E is M calendar months, each with its severity d_t from the disturbance log and, per component, its failures k_{it} and the downtime they cost; the operating hours T_{it} are the month less that downtime. Failures at the month's rate give a Poisson count,

$$k_{it} \sim \text{Poisson}\big(T_{it}\,\lambda_{i,0}\,e^{\beta_i d_t}\big), \tag{87}$$

the term (54) of Section 3.5.2 with the month's log-rate $\theta_{it} = \alpha_i + \beta_i d_t$ in place of a single one; a month without a failure still bounds the rate from above by its hours. The repairs are m_i completed over the summed downtime D_i, Poisson at the repair rate as in Section 3.5.2, one row per component since the repair rate does not depend on the severity. Each count row is local to its component and its month; what ties the months of one component together is the pair (α_i, β_i) they share, and what ties the components together is the severity they share. The priors are the records' on α_i and γ_i, centered on the handbook's rate $\bar\alpha_i$ and repair time $\bar\gamma_i$ at a spread $s = \ln 3$, and on each sensitivity a normal prior centered on no effect with spread $s_\beta = \ln 3$, a factor of three per unit of severity either way. The misfit the fit minimizes, the negative log of every factor's product, is

$$C(z) = \sum_i \Big[\frac{(\alpha_i - \bar\alpha_i)^2}{2s^2} + \frac{\beta_i^2}{2s_\beta^2} + \frac{(\gamma_i - \bar\gamma_i)^2}{2s^2} + \sum_t \big(T_{it}\,e^{\alpha_i+\beta_i d_t} - k_{it}\,(\alpha_i + \beta_i d_t)\big) + D_i\,e^{\gamma_i} - m_i\,\gamma_i\Big]. \tag{88}$$

No row holds two components, so the nine entries separate into three blocks of three; within a block the months couple α_i and β_i through θ_{it}.

3.7.3 The inference through TERRA

Each month's log-rate θ_{it} is a free variable of the graph, tied to the baseline and the sensitivity by a linear closure row, $\theta_{it} - \alpha_i - \beta_i d_t = 0$, so that the count row (87) is the Poisson term on one log-rate as in the second arm; the rest is prior rows. TERRA's Newton iteration (7) finds the mode of (88), where its gradient vanishes: for component i,

$$\sum_t \big(T_{it}\,e^{\hat\alpha_i+\hat\beta_i d_t} - k_{it}\big) + \frac{\hat\alpha_i - \bar\alpha_i}{s^2} = 0, \tag{89}$$

$$\sum_t d_t\big(T_{it}\,e^{\hat\alpha_i+\hat\beta_i d_t} - k_{it}\big) + \frac{\hat\beta_i}{s_\beta^2} = 0, \tag{90}$$

the first the balance of fitted against logged failures over all months, the second the same balance weighted by the severity, which is what separates the sensitivity from the baseline: a sensitivity moves the fitted counts of the disturbed months and leaves the calm ones. The repair entry's condition is (57) unchanged. The information matrix (10) is the curvature at the mode; with the fitted count of a month written

$$a_{it} = T_{it}\,e^{\hat\alpha_i+\hat\beta_i d_t}, \tag{91}$$

component i's block is

$$\Lambda_i = \begin{pmatrix} 1/s^2 + \sum_t a_{it} & \sum_t d_t\,a_{it} & 0 \\ \sum_t d_t\,a_{it} & 1/s_\beta^2 + \sum_t d_t^2\,a_{it} & 0 \\ 0 & 0 & 1/s^2 + D_i\,e^{\hat\gamma_i} \end{pmatrix}. \tag{92}$$

The off-diagonal term is the information the months carry on the baseline and the sensitivity together: with every month at the same severity the count rows alone would carry the two entries as one unknown, their part of the block singular, and only the prior rows would keep the block invertible, regularizing the pair without informing it; the spread of the schedule's severities is what makes the two entries two. The Laplace covariance (11) is the inverse, block by block,

$$\Sigma = \operatorname{diag}\big(\Lambda_{\mathrm{A}}^{-1},\ \Lambda_{\mathrm{B}}^{-1},\ \Lambda_{\mathrm{F}}^{-1}\big), \tag{93}$$

and its (α_i, β_i) entry is negative: a higher baseline explains the disturbed months' counts with a smaller sensitivity, and the record constrains the pair better than either alone.

3.7.4 The return to the disturbance's effect

The disturbance's effect is read from the sensitivity's marginal, normal with mean $\hat{\beta}_i$ and spread σ_{β_i} from (93): the posterior probability that the disturbance raises component i's failure rate is

$$P(\beta_i > 0 \mid E) = \Phi\left(\frac{\hat{\beta}_i}{\sigma_{\beta_i}}\right), \tag{94}$$

and the disturbance is named to act on the component when this reaches the registered confidence of 0.95, the confidence the first arm reads the claim against; the factor a full disturbance month puts on the rate has the 90 % interval

$$\big[e^{\hat{\beta}_i - z\,\sigma_{\beta_i}},\ e^{\hat{\beta}_i + z\,\sigma_{\beta_i}}\big] \ni e^{\beta_i}, \tag{95}$$

with $z = 1.645$. One fit and nothing else: the effect is an entry of the state, and its posterior is read from the covariance directly. The approximations are the Laplace form of the posterior and, above it, (83) as a model of the disturbance's effect, which the record tests only at the severities the schedule holds.

3.7.5 The brute-force reference

The reference starts from the same posterior and takes nothing from the Laplace form but the box it evaluates in, so it is an independent integration inside a Laplace-informed domain; a wider and finer grid is run beside it, below, and the change it makes is its resolution. Per component the posterior kernel is e^{-C_i}, with C_i that component's terms of (88); on a grid of $(\alpha_i, \beta_i, \gamma_i)$ centered on the fit and six Laplace widths wide each way, the weight of a grid point z_j is

$$w_j = \frac{e^{-C_i(z_j)}}{\sum_k e^{-C_i(z_k)}}, \tag{96}$$

no linearization and no normal form; the blocks are independent, so three grids of 81^3 points are the whole posterior. The sensitivity's mean, spread and $P(\beta_i > 0)$ are read from the weights directly. The registration states what the reference is expected to find: the Laplace sensitivity's mean within 0.1 of the grid's and its $P(\beta_i > 0)$ within 0.03. The reference's own resolution is read on a grid of 121^3 points 8 widths wide beside the registered one, and the change it makes in the sensitivity's moments is reported with the result. The fit is the arm's derivation and the grid is its check; the same grid, carried through the enumeration, is the fifth arm's reference.

3.8 The change of the service availability under the disturbance

The fourth arm names the disturbance; it does not say what the disturbance costs the service. The question this arm answers is by how much the disturbance changes the service availability at a stated severity, and over the plant's year. It matters because the owner sells the availability, not the failure rate: a named effect on a unit's rate is a finding, and its consequence for the service is the decision. The quantity is the posterior of the change, read as an interval and as the posterior probability that the change is a loss; nothing new is inferred, the posterior of Section 3.7 is carried through a second query.

3.8.1 The availability change on the plant graph

At a severity d of the disturbance the service availability is the query (51) of Section 3.5.1, with the configuration probabilities depending on the severity through the rates,

$$A(z, d) = \sum_{S} P(S \mid z, d)\, I_{\mathrm{succ}}(S; \theta), \tag{97}$$

with the same indicator (49), the relation holding: nothing in the physics changes with the disturbance, only the probability of each configuration, by (52) at the $q_i(d)$ of (84). The change the disturbance makes is the difference between a month at the severity and an ordinary one,

$$\Delta A(z, d) = A(z, d) - A(z, 0), \tag{98}$$

a derived quantity of the nine entries (85), and the question is its posterior,

$$P(\Delta A(d) \le u \mid E) = \int \mathbf{1}\big[\Delta A(z, d) \le u\big]\, p(z \mid E)\, dz, \tag{99}$$

the posterior probability of every state at which the change falls at or below a threshold u, summarized as the interval that holds 90 % of it and read at $u = 0$ as the probability that the disturbance lowers the service availability at all.

3.8.2 The return to the change

At the fitted state the two availabilities are one enumeration each,

$$\hat{A}(d^*) = A(\hat{z}, d^*), \tag{100}$$

at the fitted unavailabilities $\hat{q}_i(d^*)$, (84) at $\hat{z}$, and the change is their difference,

$$\Delta\hat{A}(d^*) = \hat{A}(d^*) - \hat{A}(0). \tag{101}$$

The spread of each availability is the delta method (61) on the nine entries,

$$\sigma_A^2(d^*) = g^{\top}\, \Sigma\, g, \tag{102}$$

with g the gradient of (97) in z at $\hat{z}$, read off the same enumeration through the Birnbaum importances (63) and the chain rule of (84), its nine entries written out in Appendix D; the sensitivity's entry is the baseline's scaled by the severity, zero at $d^* = 0$ where the disturbance does not act. The change is linear in the two availabilities, so its gradient is the difference of theirs,

$$g_{\Delta} = g(d^*) - g(0), \tag{103}$$

and its spread is the same quadratic form on the same covariance,

$$\sigma_{\Delta}^{2}(d^{*}) = g_{\Delta}^{\top}\,\Sigma\,g_{\Delta}, \tag{104}$$

which keeps the coupling: the two availabilities rise and fall together under the posterior, since both move with the baselines, and the difference's spread is smaller than either's. The posterior of each is the pattern (14): for the availability at the severity,

$$P(A(d^{*}) \le a \mid E) \approx \Phi\left(\frac{a - \hat{A}(d^{*})}{\sigma_{A}(d^{*})}\right), \tag{105}$$

and for the change,

$$P(\Delta A(d^{*}) \le u \mid E) \approx \Phi\left(\frac{u - \Delta\hat{A}(d^{*})}{\sigma_{\Delta}(d^{*})}\right), \tag{106}$$

the Laplace form of (99). Each is reported two ways: the interval that holds 90 % of it, the estimate plus and minus 1.645 times the spread, and, for the change, its value at $u = 0$, the posterior probability that the disturbance lowers the service availability at the severity, read against the registered confidence of 0.95 as the fourth arm reads (94). The change is reported at $d^{*} = 1$, a full disturbance month, with the availability at $d^{*} = 0$ and 1 beside it. The plant's year is neither: the service availability averaged over the schedule's severities at their frequencies,

$$\bar{A} = \sum_{d} w_{d}\,A(\hat{z}, d), \tag{107}$$

with the gradient averaged the same way, is the long-run service availability under the disturbance as it occurs, and the year's change is its difference from the ordinary month,

$$\Delta\bar{A} = \bar{A} - \hat{A}(0), \tag{108}$$

with the gradient $\bar{g} - g(0)$ in (104); (86) is the dependence the year carries. One posterior, two enumerations per severity, no draw. The approximations are those of Section 3.5.4, the Laplace form of the posterior, the linearization and the symmetric normal, and the rate model's, which the record tests only at the schedule's severities.

3.8.3 The brute-force reference

The reference is the grid posterior of Section 3.7.5 carried through the same two enumerations: each draw takes one grid point per component at its weight, sets the three $q_{i}(d^{*})$ and the three $q_{i}(0)$, passes each through (97), and the 5 % and 95 % of 4000 draws of the difference are the reference interval of the change, both terms on one draw so that the posterior's coupling of the entries is kept; the availability's own reference is the same draws through one enumeration. A draw with an unavailability at or above one half at either severity lies outside the enumerator's domain and is counted, never clipped. The registration states what the reference is expected to find: the interval's ends within 0.02 of the reference's in the availability and within 0.02 in the change. A second, independent draw set of the same size is run beside it, and the change it makes in the ends is reported with the result.

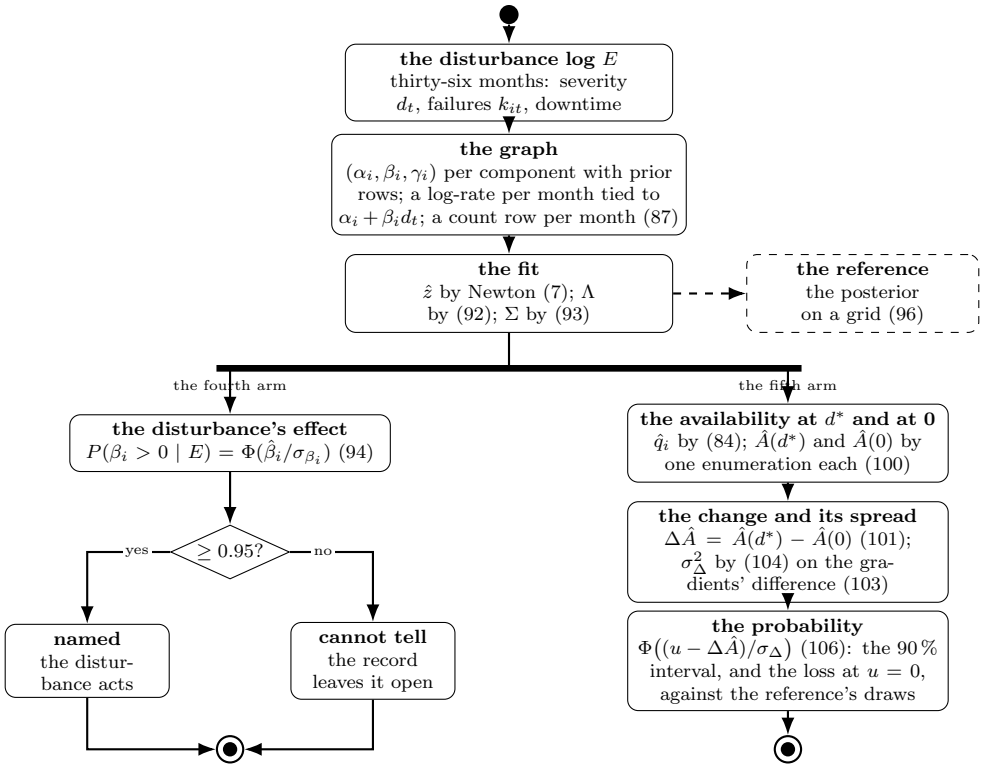


Figure 13: The steps from the disturbance log to the disturbance's effect, the fourth arm, and to the change of the service availability under it, the fifth. The log's months enter the graph as count rows on log-rates tied to each component's baseline and sensitivity, the fit gives the mode and the covariance, and the flow forks at the bar into the two arms. Left, the effect: the posterior probability that each component's sensitivity is positive (94), read against the registered confidence. Right, the change: the fitted unavailabilities at the stated severity and at zero, one enumeration each, the difference, the difference of the two Birnbaum gradients, the delta method, and the probability (106) read as the interval and at zero as the loss. The grid reference (dashed) starts from the same posterior and checks both arms.

3.9 Metrics

Of the relations above, four are references computed outside TERRA, the exact evaluation of (29), the closed form (40), the count's closed form, (32) on the exact evaluation, and the exact posterior of (53); every other quantity is TERRA's, and the comparison of drawings in Appendix C carries a fifth. Each question is read by one measure, every one a numbered relation above. The first question is read by the probability of acceptance (29) against the confidences 0.95 and 0.99, with the power (30) against 0.9 and the coverage (41) against the registered band of [0.85, 0.95] at the 0.90 level, inside which a calibrated posterior falls with high probability at 200 draws. The second is read by the band (51), its 5–95 % against the asserted point value, and by the Laplace band's endpoints relative to the exact ones in unavailability against the registered ceiling of 1.10 at the longest record. The third is read by the variance taken off (76), a candidate's class rather than its name because symmetric candidates tie. The fourth is read by the posterior probability (94) that each component's sensitivity is positive against the registered confidence of 0.95, with the grid's sensitivity moments inside the registered expectations. The fifth is read by the interval (106) read as the interval against the grid reference's, inside the registered expectation, and read at zero as the probability of a loss against the same confidence.

Table 3: The registered seed's single realization of each test, the truth the case's own opening. Columns: the observed difference of (20) and (21); the fitted conductance by (7) and the posterior spread s of its entry by (11); the full-load rise at the fit by (1); and the probability of acceptance (29) as the normal tail (36), as the discrete form (33) and as the exact one-dimensional quadrature.

case	load [kW]	$y_B - y_A$ [K]	$\hat{G}$ [W/K]	s [–]	$\widehat{\Delta T}_{\text{full}}$ [K]	Laplace P_{acc} [–]	discrete P_{acc} [–]	exact P_{acc} [–]
far	20	1.50	12614	0.406	4.76	0.900	0.969	0.969
far	40	3.10	12395	0.215	4.84	0.990	0.999	0.999
far	60	5.93	9900	0.116	6.06	0.991	0.999	0.999
near	20	2.79	7590	0.261	7.91	0.518	0.534	0.534
near	40	4.78	8428	0.146	7.12	0.787	0.820	0.820
near	60	6.39	9206	0.106	6.52	0.973	0.991	0.991

Table 4: The acceptance margin. Columns: the first grid cell [kW] at which the exact quadrature of (29) reaches 0.95 and 0.99 on the design case; the first at which the closed form (40) reaches 0.95; the first at which the power (30) reaches 0.9; the cell the registration placed the exact 0.95 crossing in; and (29) at the forward design's 80 kW opening.

test [kW]	exact 0.95 [kW]	exact 0.99 [kW]	closed form 0.95 [kW]	power 0.9 [kW]	registered [kW]	P at 80 kW [–]
20	100.0	above the grid	never (max 0.91)	above the grid	[95, 105]	0.883
40	80.0	85.0	85.0	95.0	[75, 85]	0.978
60	70.0	75.0	75.0	80.0	[65, 75]	0.998

4 Results

4.1 The N+1 claim and the acceptance margin

Each of the six registered tests is one realization of the trip test of Section 3.3, its readings generated by the measurement model (20) to (22) from the case's true opening with the registered noise, biases and seed; Table 3 lists them. The near case's 20 kW test is the one to read: a fitted conductance of 7590 W/K, a full-load rise by (1) of 7.91 K against the allowable 8 K with a posterior width of 26 %, and a probability of acceptance of 0.53: that test sat on the fence. Its full-load realization accepted the design at 0.99; the far case accepts at 0.97 from its 20 kW test and 0.999 at full load.

Figure 14 plots the probability of acceptance (29) on the design case against the true capacity for the three test loads: solid, the exact quadrature; dotted, the closed form (40); crosses, the power (30); shaded, the cells the registration placed each crossing in. Table 4 reads off the first grid cell at which each curve reaches 0.95 and 0.99, the first at which the power reaches 0.9, the registered cell, and both forms at the forward design's 80 kW opening.

From a full-load test the exact curve reaches 0.95 in the 70 kW cell and 0.99 in the 75 kW cell; from a 40 kW test in the 80 and 85 kW cells; from a 20 kW test it reaches 0.95 only at 100 kW and 0.99 nowhere on the grid. Each crossing fell in its registered cell. The closed form flattens with the margin as (40) says it must, its argument peaking where the true rise is the allowable over e, a capacity of e times one hall's load, so from a 20 kW test it tops out at 0.91 while the exact curve rises to 0.99; the two differ because the normal tail (36) behind (40) undercounts the skew of the posterior in the rise at low load. The forward study's wide design accepts at 0.88 exact and 0.79 by the closed form from a 20 kW test, under 0.95 either way, and at 0.978 from a 40 kW test.

The power (30) lags the design case: at the 70 kW cell, where the full-load design case crosses

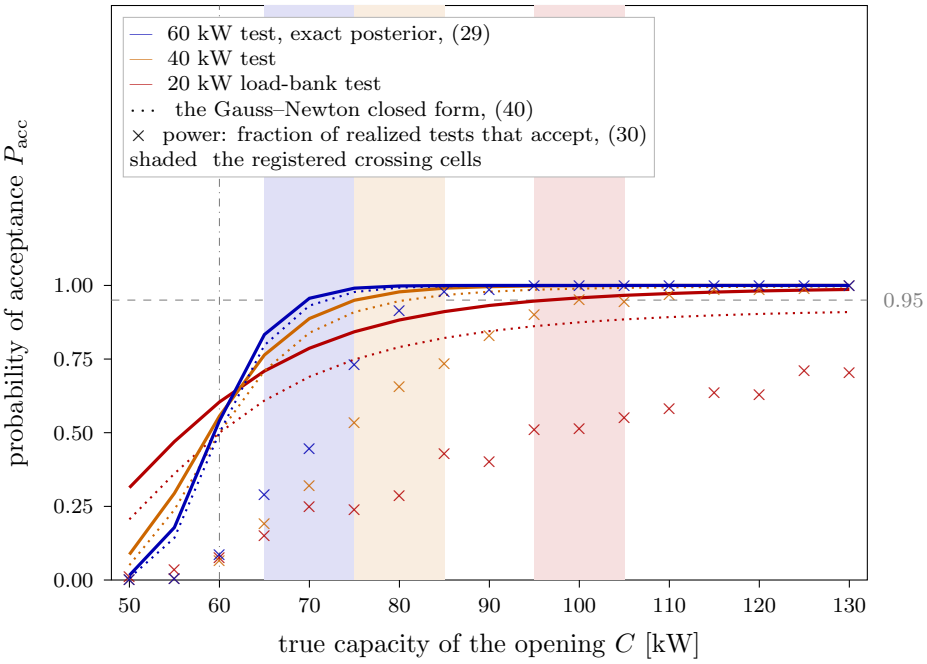


Figure 14: The probability of acceptance against the true capacity of the opening, for the three test loads, on the design case: solid, the exact posterior (29); dotted, the Gauss–Newton closed form (40). The crosses are the power (30), the fraction of two hundred realized tests at that truth whose probability of acceptance reached 0.95. The dashed line is 0.95, the vertical line one hall's load, and the shaded cells are where the registration placed each curve's crossing. The closed form flattens with the margin and never reaches 0.95 from a 20 kW test; the exact curve does, at 100 kW, and the power lags both.

0.95, 89 of 200 realized tests accept; the power reaches 0.9 in the 80 kW cell from a full-load test, the 95 kW cell from a 40 kW test, and nowhere on the grid from a 20 kW test, where even at 130 kW only 142 of 200 accept. A design that is expected to accept is not one that accepts on most days. The discrete form (33) is the exact evaluation to the grid: on the design case it is within 0.000 of the quadrature at every truth and load, and at the six seeded draws within 0.000 (Table 3), so the curves of Figure 14 are TERRA's as much as the reference's. Figure 15 sets the count against the closed forms. The reference's own, (32) on the exact evaluation, reproduces the count (30) to within 0.08 over the grid at every load, the resolution of 200 draws, from one critical difference per load, 1.66 K at 20 kW and 6.90 K at full load. TERRA's discrete rule finds the same critical differences by (31) to within 0.0 % and reproduces the count to within 0.08, from fourteen evaluations per load and no draw: the registered expectation that a closed form and the count agree within 0.08 holds for it. TERRA's Laplace rule accepts a smaller difference, 0.65 K against 1.66 K at 20 kW and 6.73 K against 6.90 K at full load, so its power lies below the count by up to 0.11 at full load, 0.19 at 40 kW and 0.52 at 20 kW, and reaches 0.9 in the 85 kW cell from a full-load test against the count's 80 kW. That is the rule, not the closed form: at the quoted truth of 70 kW the Laplace rule's own count over 200 realized tests fitted through TERRA is 85 at full load, 40 at 40 kW and 2 at 20 kW, against closed-form powers of 0.43, 0.21 and 0.01, each within the binomial resolution; the exact rule accepted 89, 64 and 50 of the same number at those truths. The Laplace rule is conservative by the tail that parts the design-case curves of Figure 14, the Laplace design case standing at 0.69 against 0.79 exact at 70 kW from a 20 kW test; where the rules part, the report reads the discrete one. Figure 16 is the calibration check (41): 200 draws per test load with the truth taken from the prior, each fitted by (7) and (11), and the fraction whose 90 % interval

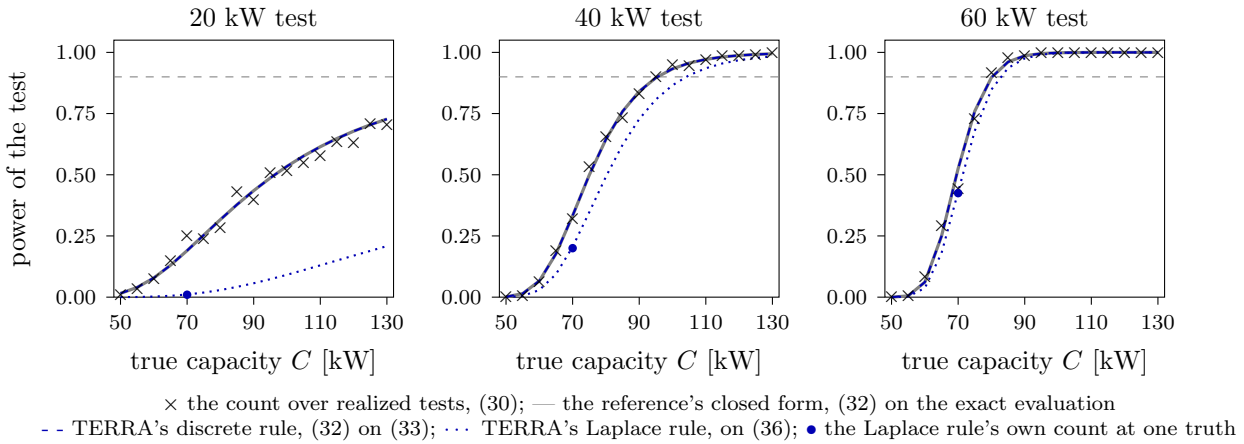


Figure 15: The power of the acceptance test against the true capacity, one panel per test load: crosses, the count over two hundred realized tests (30), scored by the exact evaluation; gray, the reference's own closed form, the critical difference (31) of the exact evaluation and one normal tail per truth (32); blue dashed, TERRA's discrete rule (33) in the same closed form, fourteen evaluations per load; blue dotted, TERRA's Laplace rule (36) in the same closed form, fourteen fits per load; the blue dot, the Laplace rule's own count over two hundred realized tests fitted through TERRA at one truth. The dashed line is 0.9. The gray and the blue dashed curves coincide and reproduce the count within the resolution of the draws at every load. The dotted curve sits below them where the test is weak: the Laplace rule accepts a smaller difference, by the tail that parts the design-case curves of Figure 14 in the same place, and its closed form reproduces its own count.

contains the truth. It is 171 of 199 at 20 kW, 173 of 194 at 40 kW and 176 of 199 at 60 kW, each with its binomial interval, each inside the registered band; 8 draws of 600 did not solve and are counted. The interval is widest at 20 kW, 0.95 in the logarithm of the rise against 0.45 at full load, as the spread (38) predicts from the two test differences. Table 5 ranks the test that would tell after the near case's 20 kW test by the expected information (43), with each probe's hall-B rise by (19) at the fitted conductance and the margin that leaves to the allowable. The stiffening column is the variance ratio of the opening's entry before and after, which for one parameter is the exponential of twice (43). The matched sensor pair adds 1.48 nats, a 19-fold stiffening at the same load and the same rise, because its common bias cancels in the difference and removes the bias term from (24); the full-load test adds 1.12 nats but would raise hall B by 8.8 K, 0.8 K past the allowable; the 40 kW test keeps a margin of 2.2 K; running the 20 kW test eight times longer adds 0.33 nats, because (24) divides only its noise term by the sample count. Of the readings, the differenced channel with any one meter decides accept-or-not with the same plug-in probability of 0.571 under a budget of two; the meters add nothing to that decision. The margin is the number to design to: a design built to the step cannot be shown redundant by any affordable test.

Four comparisons are made here and each tests a different thing. The exact quadrature of (29) against the closed form (40) tests the normal tail (36) that the engine's Laplace spread rests on: the two agree closely at full load and part at low load and wide margin, where (40) flattens and never reaches 0.95 while the exact curve does, so the Laplace spread is a safe summary when the test is informative and an undercount when it is not, which is when the report reads the exact curve instead. The design case against the power (30) tests expectation against realization: at the cell where the expected full-load test crosses 0.95, 89 of 200 realized tests do, so an acceptance margin read from the design case must be widened by the scatter before it is promised. The coverage (41) tests the posterior itself: an interval that claims 90 % and covers 86 to 89 % over 200 draws with

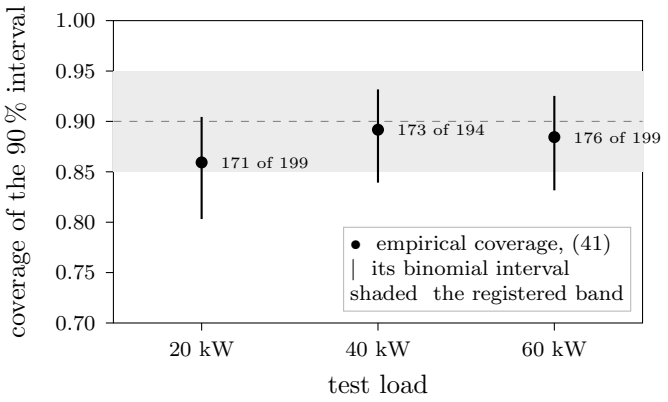


Figure 16: Coverage of the 90 % credible interval on the full-load rise over two hundred draws per test load, by (41) with each draw fitted by (7) and (11), with the Clopper–Pearson interval of each count and the registered band shaded. The truth of each draw is drawn from the prior, so a calibrated posterior covers at the nominal level in expectation; all three sit inside the band, the 20 kW test lowest.

Table 5: The test that would tell, after the near case's 20 kW test ($\hat{G} = 6851$ W/K, $s = 0.24$): each probe's expected information on the conductance by (43); the factor by which it stiffens the posterior, the variance ratio of the entry θ before and after, the exponential of twice (43) for one parameter; the hall-B temperature rise it causes by (19) at the fitted conductance; and the margin that leaves to the allowable difference. A negative margin is a probe that breaches the allowable.

probe	information [nats] (43)	stiffening [–]	hall B rise [K]	margin to the allowable [K]
matched sensor pair	1.48	19.2	2.9	5.1
trip test at 60 kW	1.12	9.4	8.8	-0.8
trip test at 40 kW	0.78	4.7	5.8	2.2
longer test (240 samples)	0.33	1.9	2.9	5.1

the truth drawn from the prior is calibrated to within its binomial error at every load, which is the statement that the engine's uncertainty bars are honest, and it is the one check in the report that could have failed without any number looking wrong. The count of realized tests against the closed form (32) tests whether the simulation was needed: the reference's critical difference and TERRA's discrete rule's both reproduce the count within the draws' resolution, so it was not, and the Laplace rule's closed form reproduces that rule's own count, so the power of an acceptance rule is a property of the rule and is computable from fourteen evaluations per load before any test is run. A closed form affine in the readings with the posterior spread held at the design case, registered on 2026-10-25 and kept in the results file, is not: the Laplace spread of a realized weak test moves with its reading, and that form missed the Laplace rule's own count by up to 0.34 at 20 kW. The probe ranking by (43) then shows the method used forward: it names the matched sensor pair over the hotter test because the floor (24) is bias, a conclusion that follows from the factor graph's own weights and would not be visible in a dashboard of the same readings.

4.2 The service availability from the outage record

Table 6 gives the posterior on each component's failure rate from the three records: the Laplace mode and spread of (54) with the lognormal prior, the mean, spread and 5–95 % interval of the exact posterior of (53) on its grid, their difference in location, and the unavailability by (50) at

Table 6: The posterior on the log failure rate per year from each record (unit A and the feeder; unit B's record is unit A's). Columns: the count; the Laplace mode and standard deviation of (54) with the lognormal prior; the exact posterior's mean, standard deviation and 5–95 % interval, (53) times the prior on a grid; their difference in location; and the unavailability by (50) at the exact posterior-mean rates. The two agree in spread at every record and differ in location by an amount that falls with the count.

record [yr]	unit	k [count]	Laplace mode [ln/yr]	sd [ln/yr]	exact mean [ln/yr]	sd [ln/yr]	exact 5–95 % [ln/yr]	mode − mean [ln/yr]	$\hat{q}$ [–]
1	UA	2	0.76	0.58	0.64	0.59	[-0.40, 1.54]	0.119	0.0421
1	feeder	4	1.37	0.46	1.29	0.47	[0.47, 2.01]	0.086	0.0107
3	UA	7	0.85	0.36	0.80	0.36	[0.17, 1.36]	0.057	0.0461
3	feeder	11	1.30	0.29	1.26	0.30	[0.75, 1.72]	0.040	0.0100
10	UA	25	0.92	0.20	0.90	0.20	[0.56, 1.21]	0.019	0.0489
10	feeder	37	1.31	0.16	1.30	0.16	[1.02, 1.56]	0.013	0.0100

the exact posterior-mean rates. The two posteriors agree in spread at every record, at the $1/\sqrt{k}$ of Section 3.5, and differ in location: the mode sits above the mean by 0.12 in the log-rate at two failures, 0.06 at seven and 0.02 at twenty-five, the skew of (53) at small counts.

Figure 17 and Table 7 carry those posteriors into the service availability by (51): draws of the rates from the exact posterior, turned into unavailabilities by (50) and pushed through the forward enumeration. Table 7 lists, per record and structure, the asserted point value, the 5 %, median and 95 % of the draws, how many draws fell below the asserted value, and the Laplace endpoints relative to the exact ones in unavailability. For the parallel pair with its structure settled, three unit-years put the service availability in [0.9938, 0.9993] with the median at 0.9979, around the asserted 0.9975, which 760 of 2000 draws fall below: the asserted value is a point inside a band that spans a factor of about 9 in unavailability. The band's width falls from 0.0102 at one year to 0.0031 at ten, the $1/\sqrt{k}$ spread of (54) seen through (50) and (51). The Laplace endpoints differ from the exact ones by factors of 0.975 and 0.964 at three years and 0.988 and 1.006 at ten, inside the registered ceiling of 1.10. With the feeder inside the boundary the band sits lower by the feeder's factor, as in the forward model.

Table 8 sets the interval (68), one enumeration per structure and record with no draw, beside the exact band. For the settled parallel pair at three years it is [0.9955, 1.0003] against the exact [0.9938, 0.9993]: its lower end sits 0.0017 above the exact one and its upper end passes one, since $\sigma_A = 0.0014$ is of the order of $1 - \hat{A} = 0.0021$, the normal form (68) puts 7 % of its mass above one, and the exact posterior is skewed toward low values, which a symmetric interval cannot say; at one year the same pair's upper end is 1.0015 with 19 % above one. At ten years the pair's interval is [0.9961, 0.9991] against the exact [0.9956, 0.9987], inside one, and over the four structures no end of the interval is further than 0.007 from the exact band's in the availability, against 0.085 at one year: the linearization and the normal form both close with the record. The mixture's ends by (73) at three years are [0.8729, 0.9998] against the exact mixture's [0.8530, 0.9991], the lower end again above the exact one by the same skew.

The near case after its 20 kW test carries the claim's probability 0.534 from Section 4.1, so its band is the mixture (69) of the parallel and series bands at that weight (72). Its 5 % at three years is 0.853 and its median 0.9950, and no length of record closes the gap between them, because the mixture keeps the series structure at a weight near one half whatever the record says about the rates; in Figure 17 its upper end sits two decades above the settled band. The record narrows the rates; only a test narrows the structure.

The Laplace against the exact posterior on the rates is the same check as the normal tail (36)

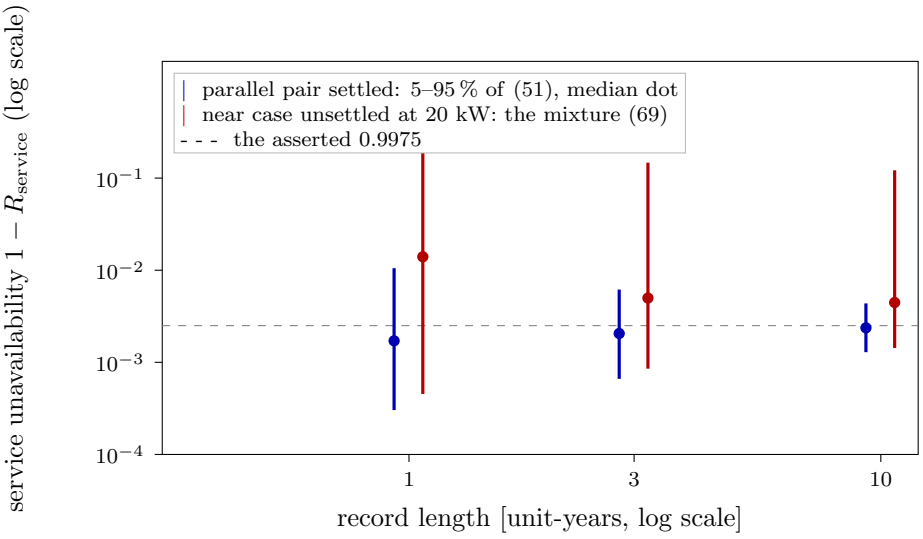


Figure 17: The band on the service availability from the plant's own record, against the record's length, drawn as the service unavailability on a log scale (a band's 5 % end is its upper mark). Blue, the parallel pair with its structure settled: the 5–95 % band of (51) from the exact posterior draws, the median as a dot, narrowing with the record around the asserted value. Red, the near case with its structure unsettled after a 20 kW test: the mixture (69) of the two structures' bands at the claim's probability, whose upper end sits two decades higher because the series structure keeps a probability near one half whatever the record says.

against the exact evaluation of (29), made on the second arm: the engine's mode-and-curvature summary against a grid that is exact. The two agree in spread at every record and differ in location by an amount that falls with the count, 0.12 in the log-rate at two failures and 0.02 at twenty-five, so the Laplace summary is good for a record of a few years and biased high for a record of a few failures, which is why the interval (68) is the derivation and the exact band the check, reported beside it inside the registered ceiling. The settled band against the mixture band (69) is the comparison that carries the arm's result: the record narrows the settled band from 0.0102 to 0.0031 over ten years, and leaves the mixture's 5 % at 0.85 at any length, because the structure is not a rate and no count of failures moves it. That is the demonstration: the uncertainty in a availability number has two sources, the rates and the structure, and only the first is reduced by the record the plant already keeps.

4.3 The value of each candidate measurement

Tables 9 and 10 rank the candidates by the variance they take off the band, in the two regimes, and Figure 18 shows what each leaves. The variance before is (76), with the gradient (66) read off the enumeration and, in the unsettled regime, assembled by (79); the variance after is (76) with the candidate's information added to the posterior and, for a trip test in the unsettled regime, the structure term recomputed by (80); the last column is the candidate's information on the parameters by (43), a different ranking.

With the structure unsettled, the near case after its 20 kW test, the band's spread is 0.0485, and the structure term of (79), 1.9e-03, is two orders above the two structures' own variances, 1.9e-05 and 9.3e-04. Every trip test, which moves the structure probability, therefore beats every record, which moves only the rates: the matched sensor pair takes the spread to 0.029, a year of any component's record only to 0.048. With the structure settled, the far case at full load,

Table 7: The band on the service availability from each record, per structure. Columns: the asserted point value of the forward model; the 5 %, median and 95 % of the draws of (51) from the exact posterior through (50); how many of those draws fell below the asserted value; and the endpoints of (51) from the Laplace draws relative to the exact ones in unavailability (the registered ceiling is 1.10 at the longest record). The mixture rows are the near case after its 20 kW test, by (69) at the claim's probability (72), the two structures' draws redrawn and mixed; draws outside the enumerator's domain are counted in the results file and are at most one per band.

record [yr]	structure	asserted A [–]	5 % A [–]	median A [–]	95 % A [–]	below [count]	Laplace/exact $(1-A)$ at 5 / 95 % [–]
1	parallel pair	0.9975	0.9895	0.9983	0.9997	753 of 1999	0.920 / 0.844
1	parallel pair with the feeder	0.9875	0.9648	0.9865	0.9948	1125 of 1999	0.994 / 1.027
1	series pair	0.9025	0.7606	0.9067	0.9643	946 of 1999	0.987 / 1.014
1	near case unsettled (mixture)	0.9975	0.8070	0.9860	0.9995	1371 of 1998	–
3	parallel pair	0.9975	0.9938	0.9979	0.9993	760 of 2000	0.975 / 0.964
3	parallel pair with the feeder	0.9875	0.9770	0.9874	0.9930	1026 of 2000	0.989 / 0.993
3	series pair	0.9025	0.8331	0.9044	0.9461	945 of 2000	0.987 / 0.980
3	near case unsettled (mixture)	0.9975	0.8530	0.9950	0.9991	1331 of 2000	–
10	parallel pair	0.9975	0.9956	0.9976	0.9987	888 of 2000	0.988 / 1.006
10	parallel pair with the feeder	0.9875	0.9827	0.9874	0.9910	1030 of 2000	0.987 / 1.004
10	series pair	0.9025	0.8687	0.9031	0.9295	972 of 2000	1.001 / 1.006
10	near case unsettled (mixture)	0.9975	0.8785	0.9955	0.9986	1424 of 2000	–

the spread is 0.0044, the conductance carries no gradient in (66), and every trip test's reduction is round-off, stated as a bound. Among the records the feeder's wins, to 0.0039, because its Birnbaum importance by (62) is the largest: its outage takes both units, against the loss of one unit that the other covers. The two unit-year candidates tie by symmetry and are ordered by name. The information column ranks differently in both regimes: the matched pair adds the most information in the settled regime too and takes nothing off the band, because the parameter it informs has no gradient there. The repair log already in the record took the settled variance from 1.4e-04 to 1.9e-05; a plant that logs failures and not repairs has a band about 2.7 times wider. This is the arm that spends the plant's money, and (76) says where.

Two rankings are shown because they answer different questions, and their disagreement is the point. The information (43) says how much a candidate narrows the parameters; the variance taken off (76) says how much it narrows the number the owner acts on. The matched sensor pair tops the first in both regimes and tops the second only when the structure is unsettled: once the structure is settled it adds 1.5 nats on a parameter with no gradient and takes nothing off the band. A ranking by information alone would have bought the wrong measurement in the settled plant. The two regimes against each other show the same thing from the other side: the best spend switches from a trip test to a year of the feeder's record the moment the structure is known, because the structure term of (79) is two orders above the rates' own variances while it is open and vanishes when it is closed. What this demonstrates about the method is that the value of a measurement is a property of the whole graph at its current posterior, not of the instrument, and the same instrument is the best or the worst buy depending on what is already known.

4.4 The shared disturbance from three years of log

Table 11 gives the three years of the disturbance log as they fell under the registered seed: 24 months at severity 0, 6 at one half and 6 at one, with 9, 15 and 16 failures of unit A, unit B and the

Table 8: The interval (68) from one enumeration per structure and record, beside the exact band. Columns: the service availability at the fitted rates; its spread by the delta method (61); the two ends of the interval at $z = 1.645$; and the 5 % and 95 % of the exact draws of (51). An upper end above one is printed as the normal form gives it. The mixture rows are the ends of (73) at the claim's probability (72), beside the exact mixture's.

record [yr]	structure	$\hat{A}$ [–]	σ_A [–]	$\hat{A} - z\sigma_A$ [–]	$\hat{A} + z\sigma_A$ [–]	exact 5 % [–]	exact 95 % [–]
1	parallel pair	0.9982	0.0020	0.9950	1.0015	0.9895	0.9997
1	parallel pair with the feeder	0.9876	0.0071	0.9759	0.9992	0.9648	0.9948
1	series pair	0.9178	0.0453	0.8433	0.9923	0.7606	0.9643
1	near case unsettled (mixture)	–	–	0.8616	1.0012	0.8070	0.9995
3	parallel pair	0.9979	0.0014	0.9955	1.0003	0.9938	0.9993
3	parallel pair with the feeder	0.9879	0.0043	0.9809	0.9950	0.9770	0.9930
3	series pair	0.9100	0.0299	0.8608	0.9592	0.8331	0.9461
3	near case unsettled (mixture)	–	–	0.8729	0.9998	0.8530	0.9991
10	parallel pair	0.9976	0.0009	0.9961	0.9991	0.9956	0.9987
10	parallel pair with the feeder	0.9876	0.0024	0.9836	0.9916	0.9827	0.9910
10	series pair	0.9047	0.0174	0.8761	0.9333	0.8687	0.9295
10	near case unsettled (mixture)	–	–	0.8831	0.9988	0.8785	0.9986

feeder. Under the registered truth a full disturbance month quadruples a unit's rate and doubles the feeder's, so the six full months and six half months, a third of the log, carry 6 of unit A's 9 failures and 8 of unit B's 15.

Table 12 and Figure 19 give the fit. Unit A's sensitivity is $\hat{\beta} = 1.01 \pm 0.61$ and unit B's 0.98 ± 0.52, so the disturbance is named to act on both by (94), at $P(\beta > 0 \mid E)$ of 0.952 and 0.970 against the registered 0.95, unit A by a hair, and the grid reference puts unit A at 0.939, below it: the decision on unit A turns on the posterior approximation, and agreement inside the registered expectation is not agreement on a decision that sits at the threshold, so the record settles unit B and leaves unit A at the line. A full disturbance month multiplies their rates by 2.7 and 2.7 at the mode, inside [1.0, 7.5] and [1.1, 6.3] by (95), intervals that hold the true factor of four and start a little above one: three years of log detect the effect and do not yet size it. The feeder's sensitivity is 0.18 ± 0.55 at $P(\beta > 0 \mid E) = 0.63$, not named, its interval [0.5, 2.9] holding both no effect and the true two; a feeder that fails 3.69 times a year at baseline gives the twelve disturbed months too few extra failures to tell a doubling from none. The correlation of baseline and sensitivity is -0.58 to -0.50, the trade (92) describes, and the probability that the disturbance acts on both units at once is 0.924. The grid reference agrees with the Laplace form: its sensitivity means sit within 0.048 of the modes and its $P(\beta > 0)$ within 0.025, inside the registered 0.1 and 0.03, and below the modes in every case, the skew of a posterior on a few counts that the Laplace form sits above, as in Section 4.2. The reference's resolution is small against those gaps: the wider and finer grid moves the sensitivities' means by at most 0.0000 and their $P(\beta > 0)$ by at most 0.0004.

The fit against the grid is the comparison that carries the arm's check, and the fit against the practice the one that carries its result. The practice multiplies availabilities at handbook rates and, where it admits a common cause at all, asserts its share. Here the disturbance's effect is a parameter with a posterior, named on the units because the disturbed months hold more failures than the calm ones can explain and left open on the feeder because they do not. What the arm demonstrates is that the one input of a redundancy claim the practice can only assert, the coupling of the units' failures, is a quantity the plant's own disturbance log estimates, with a stated confidence and a stated regime, the feeder's, in which three years cannot tell. What it establishes is a relation between a logged condition and the components' failure rates under the stated model, read from

Table 9: The value of each candidate measurement with the structure unsettled (near case, 20 kW test; the record at three unit-years). Columns: the standard deviation of the service availability before, the root of (76) with the gradient (66) and, when the structure is unsettled, the mixture (79); the same after the candidate's information is added, with the structure term recomputed by the preposterior expectation of Appendix D for a trip test; the variance taken off; and the information the candidate adds on the parameters by (43), which is not the same ranking. The structure probability is 0.518 and the structure term of (79) is 1.89e-03, against 1.90e-05 and 9.31e-04 for the two structures' own variances.

candidate	class	sd before [–]	sd after [–]	reduction of var [–]	information on θ [nats]
matched sensor pair	trip test	0.0485	0.0291	0.0015	1.48
trip test at 60 kW	trip test	0.0485	0.0319	0.00133	1.12
trip test at 40 kW	trip test	0.0485	0.0354	0.0011	0.78
longer test (240 samples)	trip test	0.0485	0.0417	0.000612	0.33
one more unit-year, UA	record	0.0485	0.0479	5.53e-05	0.29
one more unit-year, UB	record	0.0485	0.0479	5.53e-05	0.29
one more year of the feeder's record	record	0.0485	0.0484	3.79e-06	0.28

Table 10: The value of each candidate measurement with the structure settled (far case, 60 kW test; the record at three unit-years). Columns: the standard deviation of the service availability before, the root of (76) with the gradient (66) and, when the structure is unsettled, the mixture (79); the same after the candidate's information is added, with the structure term recomputed by the preposterior expectation of Appendix D for a trip test; the variance taken off; and the information the candidate adds on the parameters by (43), which is not the same ranking. With the structure settled the conductance carries no gradient and every trip test's reduction is round-off, stated as a bound; the repair log already in the record took the variance from 1.36e-04 to 1.90e-05.

candidate	class	sd before [–]	sd after [–]	reduction of var [–]	information on θ [nats]
one more year of the feeder's record	record	0.0044	0.0039	4.12e-06	0.28
one more unit-year, UA	record	0.0044	0.0043	2.67e-07	0.29
one more unit-year, UB	record	0.0044	0.0043	2.67e-07	0.29
trip test at 40 kW	trip test	0.0044	0.0044	below 10^{-12}	0.18
trip test at 60 kW	trip test	0.0044	0.0044	below 10^{-12}	0.34
longer test (240 samples)	trip test	0.0044	0.0044	below 10^{-12}	0.34
matched sensor pair	trip test	0.0044	0.0044	below 10^{-12}	1.50

counts; which mechanism carries the stress to the units is not in the record, and a fault that takes both units out at once is a mechanism the model does not hold.

4.5 The change of the service availability under the disturbance

Table 13 and Figure 20 carry the fit into the service availability by (100) and (105). In an ordinary month the availability is 0.9826 with a spread of 0.0056, the interval [0.9734, 0.9917]; in a full disturbance month it is 0.9622, [0.9354, 0.9890], the units' unavailabilities having risen from 0.046 and 0.077 to 0.118 and 0.183; over the schedule the long-run service availability (107) is 0.9781, [0.9677, 0.9884]. Table 14 and Figure 21 give the change, the arm's quantity. A full disturbance month takes 0.0204 off the service availability, $\Delta\hat{A}(1)$ of -0.0204 by (101) with a spread of 0.0157 by (104), the interval [-0.0462, 0.0055] by (106) and a posterior probability of a loss of 0.903 by (106), not named at the registered 0.95; over the year the change (108) is -0.0045, [-0.0101, 0.0011], a loss at 0.906. The loss is likely and not yet established: the interval's upper end passes zero

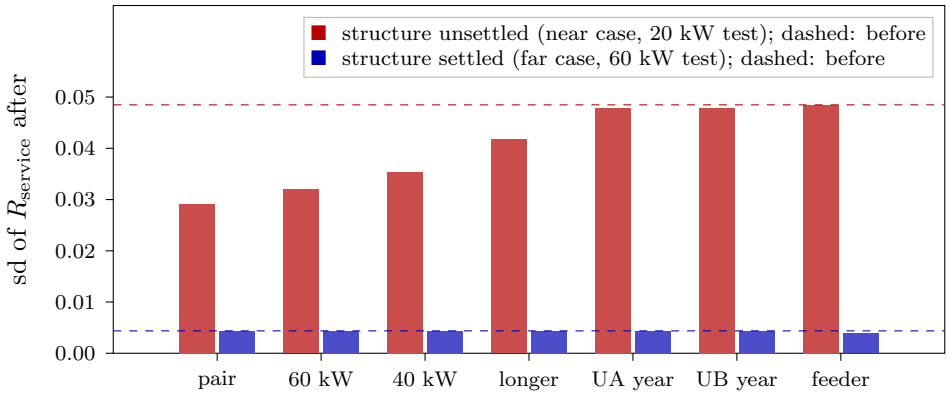


Figure 18: What each candidate measurement leaves of the band on the service availability: the standard deviation after the measurement, the root of (76) with the candidate's information added and the structure term by (80), in the two regimes, with the standard deviation before as a dashed line. The candidates are in the order of the unsettled ranking. With the structure unsettled every trip test beats every record, the matched sensor pair most; with it settled the trip tests take nothing off and the feeder's record takes the most.

Table 11: The disturbance log of three years as it fell, by the schedule's severity levels: the months at each, the failures and operating hours per component, and the registered true rates (83) at that severity (a full disturbance month multiplies a unit's rate by four and the feeder's by two). The windows are 720 h long; each failure costs its component's mean repair time, and the operating hours are the window less that downtime.

severity d [–]	months [count]	failures UA / UB / feeder [count]	operating hours UA / UB / feeder [h]	true rate, unit / feeder [1/yr]
0	24	3 / 7 / 11	16 740 / 16 020 / 17 016	2.50 / 3.69
0.5	6	3 / 3 / 1	3 780 / 3 780 / 4 296	5.00 / 5.22
1	6	3 / 5 / 4	3 780 / 3 420 / 4 224	10.00 / 7.38
all	36	9 / 15 / 16	–	–

because the factors' intervals of Section 4.4 reach down to almost no effect. The change's spread is 0.96 of the availability's in the full month, the coupling (104) keeps: what the two enumerations share, the baselines' uncertainty, cancels in the difference, and what is left is the sensitivities'. The reference's 5–95 % intervals of the change are [-0.0551, 0.0001] and [-0.0115, 0.0003]: the Laplace ends sit within 0.009 of them, inside the registered 0.02, and the availability's own ends within 0.011 of theirs, [0.9701, 0.9895], [0.9244, 0.9816] and [0.9639, 0.9858], inside the registered 0.02 and above them at both ends, the skew of a posterior bounded by one that the symmetric form does not carry. The second draw set moves the reference's ends by at most 0.0007 in the availability and 0.0011 in the change. The dependence (86): the units' joint outage averaged over the schedule is 0.0075 against a product of averaged unavailabilities of 0.0064, a ratio of 1.16; the beta factor of the practice would assert such a number, and here it is read from the record with the spread of the sensitivities behind it.

The practice has no spread to compare: the product rule at handbook rates gives one number, 0.9875 in the forward study, and a beta factor moves it to another. Here the change the disturbance makes carries the spread the schedule's twelve disturbed months set, and the posterior probability that it is a loss at all: 0.9826 in an ordinary month, 0.9622 in a full disturbance month, 0.9781 over the year, each with its interval, and the full month's loss of 0.0204 with its own. What the

Table 12: The fit per component: the posterior mode and standard deviation of the log baseline rate, the sensitivity and the log repair rate from (93); the correlation of baseline and sensitivity; the posterior probability that the disturbance raises the rate, (94) from the Laplace form and from the grid reference; the factor a full disturbance month puts on the rate at the mode with its 90 % interval (95); and whether the disturbance is named at the registered confidence of 0.95. The true factors are four for a unit and two for the feeder.

component	$\hat{\alpha}$ [ln 1/yr]	$\hat{\beta}$ [–]	$\hat{\gamma}$ [ln 1/yr]	corr α, β [–]	$P(\beta > 0)$ Laplace [–]	$P(\beta > 0)$ grid [–]	$e^{\hat{\beta}}$ [lo, hi] [–]	named
unit A	0.86 ± 0.39	1.01 ± 0.61	3.88 ± 0.32	-0.58	0.952	0.939	2.7 [1.0, 7.5]	yes
unit B	1.41 ± 0.32	0.98 ± 0.52	3.88 ± 0.25	-0.60	0.970	0.961	2.7 [1.1, 6.3]	yes
feeder	1.64 ± 0.28	0.18 ± 0.55	5.90 ± 0.24	-0.50	0.629	0.604	1.2 [0.5, 2.9]	no

Table 13: The service availability under the disturbance, the parallel pair with the feeder inside the boundary: the enumeration at the fitted unavailabilities (100), the delta-method spread (102), the two ends of the 90 % interval of (105), and the 5 %, median and 95 % of the grid reference's draws through the same enumeration, with their count inside the enumerator's domain. The schedule row is (107), the levels' values at their frequencies.

severity	$\hat{A}$ [–]	σ_A [–]	$\hat{A} - z\sigma_A$ [–]	$\hat{A} + z\sigma_A$ [–]	reference 5 % [–]	median [–]	95 % [–]	draws [count]
an ordinary month, $d^* = 0$	0.9826	0.0056	0.9734	0.9917	0.9701	0.9821	0.9895	4000
a full disturbance month, $d^* = 1$	0.9622	0.0163	0.9354	0.9890	0.9244	0.9619	0.9816	3997
the plant's year, over the schedule	0.9781	0.0063	0.9677	0.9884	0.9639	0.9774	0.9858	3996

arm demonstrates is that a named effect on a component's rate and its cost to the service are two questions on one posterior: the second needed no new inference, only a second query and its gradient.

5 Discussion

Table 15 is the summary of this report. Each row is one comparison made in Section 4: what is being measured, what the plant has today or an answer that can be computed without the engine, what TERRA computed, and what the comparison shows. The numbers of this toy are not the result; the pattern of the table is. In four rows the engine's standard summary is held against an exact or known answer, and it holds wherever the test is informative and is replaced where it is not: the method can be trusted on the problems where no exact answer exists. In six rows it is held against what the plant has today, a reading beside the drawings' predictions, a single asserted number, an assumed independence, an expected rather than a realized test, and the difference between the columns is the arm's result. What the process buys is not the numbers of this toy but a redundancy claim that carries its own evidence: which record tested which input, at what confidence, and what would strengthen it. A claim without that is a count of units, and a count is what the industry mostly has.

Each of the redundancy claim's three asserted inputs turned out to be testable by a record the plant keeps, and each test has a regime in which it cannot tell. The success structure is tested by a trip test, and the test establishes the claim at 0.97 when the opening's capacity is well above the load and leaves it at 0.53 from a load-bank test when the capacity is within a tenth of it; the decisive quantity is the floor on the hall difference, which is the sensors' bias and not their noise, so the remedy is a matched pair rather than a longer test. The unavailabilities are tested by the

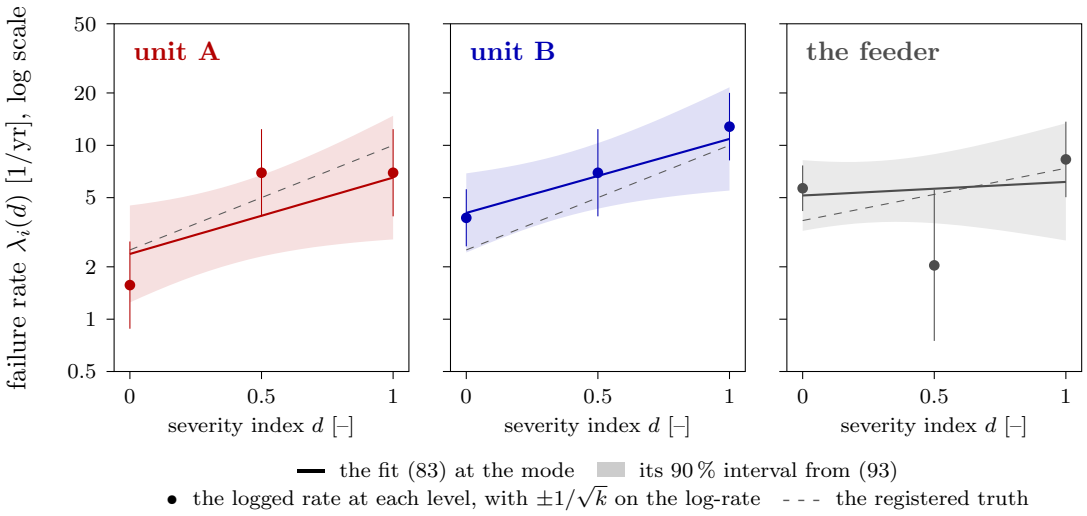


Figure 19: The failure rate of each component against the month's severity index, one panel per component on a log axis, on which the fit (83) is a straight line whose slope is the sensitivity. The band is the 90 % interval of the fitted log-rate at each severity from the covariance (93); the points are the rate the log itself shows at each of the schedule's three levels, the failures over the operating years at that level, with the bar $\pm 1/\sqrt{k}$ on the log-rate, the Poisson spread of a count of k; the dashed line is the registered truth, a factor of four for a unit and two for the feeder. The units' slopes are clear and their bands hold the truth; the feeder's band holds a flat line as well, which is why three years leave its effect open. The bands widen toward the full disturbance, which six months of thirty-six hold.

Table 14: The change of the service availability under the disturbance, from an ordinary month: the difference of two enumerations at the fitted state (101), the delta-method spread on the difference of the gradients (104), the two ends of the 90 % interval of (106), its value at zero, the posterior probability of a loss, and the 5 %, median and 95 % of the grid reference's per-draw differences through the same two enumerations, over the draws inside the enumerator's domain (their count is Table 13's). The year row is (108).

severity	$\Delta\hat{A}$ [–]	σ_Δ [–]	$\Delta\hat{A} - z\sigma_\Delta$ [–]	$\Delta\hat{A} + z\sigma_\Delta$ [–]	$P(\Delta A < 0)$ [–]	reference 5 % [–]	median [–]	95 % [–]
a full disturbance month, $d^* = 1$	-0.0204	0.0157	-0.0462	0.0055	0.903	-0.0551	-0.0193	0.0001
the plant's year, over the schedule	-0.0045	0.0034	-0.0101	0.0011	0.906	-0.0115	-0.0041	0.0003

outage record, and a short record gives a skewed posterior whose Laplace approximation is right in width and wrong in location by an amount that falls with the count; the band on the service availability must be built from the exact draws, and a plant that logs failures without repairs has a band several times wider. The coupling of the units' failures is tested by the disturbance log, and three years of it name the disturbance on one unit, put the other at the threshold where the two posterior forms part, and leave its effect on the feeder, whose doubling adds too few failures to twelve disturbed months, open; what the coupling costs the service is a second query on the same posterior, a loss of 0.020 in a full disturbance month with its interval.

The limits of the study are its data and its model. Every reading and record is synthetic, drawn from the stated models with the stated noise; what the report establishes is what such records can decide, with the honesty of a registration committed before any number, and not what a particular plant's records decide. The capacity of the opening is read as the heat passed at the allowable difference, and that reading is the one modeling decision the report makes: steady readings determine a conductance, and a hard cap on the flow, which the forward model declares,

Table 15: The comparisons of this report. For each, the quantity in plain words with its equation, the alternative or reference it is set against, what TERRA computed, and what the comparison shows. The wide drawing's hypothesis in Appendix C is the AND tree, H_{AND}; the narrow drawing's is the OR tree, H_{OR}.

§	what is measured	alternative or reference	TERRA	what the row shows
4.1	the hall difference a trip test produces	the reading itself, $y_B - y_A$, with its floor (24)	the difference at the conductance fitted to the whole graph (7), the reading with its sensor biases estimated out	the fitted difference sits within the floor of the reading in every test; the distance between them is the bias the fit attributed to the two sensors
C.3	which of the forward study's two drawings a trip test's readings favor, in nats	a one-line formula with each drawing's full-load rise taken as exact and only the hall difference used	the full model fitted once per drawing, biases and meters included, and the two evidences compared	the same verdict in 5 cases of 6; TERRA's number is lower by what the simplification leaves out, so it errs toward caution
4.1	the probability that the opening passes one hall's load	a normal-curve shortcut from the fit's mean and spread (40)	the posterior integrated exactly (29)	they agree when the test is informative and part when it is not; the shortcut flatters a weak test
4.1	how often a real test passes the plant at a given true capacity	what the expected, noise-free test says: the design case	the fraction of 200 noisy tests that pass, the power (30)	the expected test passes well before most real tests do; a margin must be widened by the scatter
4.1	how often a real test passes, with no test run	the count over 200 simulated tests (30) and its own closed form	each rule's critical difference by bisection, fourteen evaluations per load, then a normal tail (32)	TERRA's discrete rule matches the count within 0.08; its Laplace rule matches its own count and sits below the exact rule's by up to 0.11 at full load: conservative where the test is weak
4.1	whether the stated uncertainty is honest	the true conductance, drawn 200 times from the prior	how often the 90 % interval contained it, the coverage (41)	86 to 89 %, as a 90 % interval should: the uncertainty bars can be trusted
4.2	a component's failure rate learned from its outage log	the exact posterior on a fine grid (53)	the normal summary at the peak (54)	the same width; the summary sits high for a log of few failures, so the band uses the exact draws
4.2	the service availability	the single number the claim asserts	a band from the plant's own record (51), wider while the structure is in doubt (69)	the asserted number is one point in a band a factor of 9 wide in unavailability; only a test narrows the structure
4.2	the band on the service availability, with no draw	2000 draws of the rates through the enumeration (51)	one enumeration, its gradient and the fits' spreads (68)	the interval sits within 0.007 of the exact band in the availability at ten years and passes one for the parallel pair on a short record: the draws are the check, not the answer
4.3	which measurement to buy next	how much each candidate narrows the model's parameters (43)	how much each narrows the availability the owner acts on (76)	different rankings; the same instrument is the best or the worst buy depending on what is already known
4.4	the effect of a shared disturbance on the failure rates	a common-cause fraction asserted from a table, never tested against the record	the sensitivity of each component's rate to the logged severity, with its posterior (94)	named on unit B at $P(\beta > 0) \geq 0.95$, unit A at the threshold where the Laplace form and the grid part, the feeder left open; a factor of 2.7 at the mode with an interval three years leave wide
4.5	the change of the service availability	4000 draws of the grid posterior through two	one posterior, two enumerations per	the interval sits within 0.009 of the reference in the

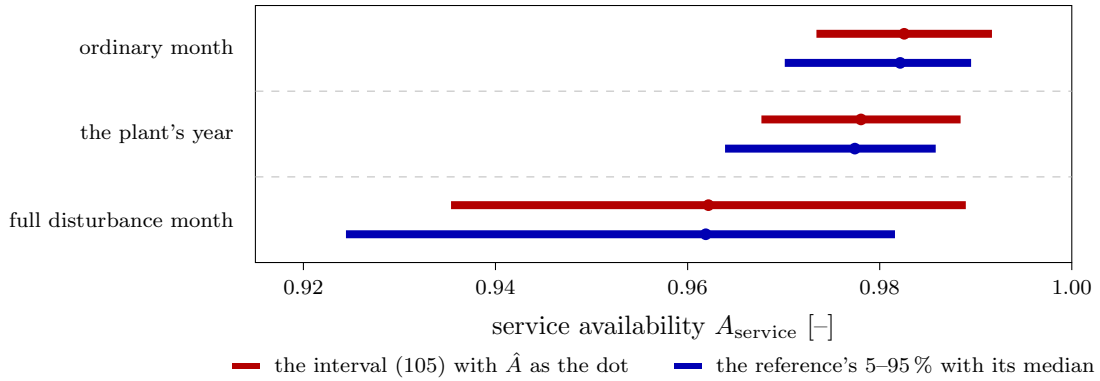


Figure 20: The service availability under the disturbance, at each reported severity and over the schedule. Red, the interval $\hat{A} \pm 1.645\,\sigma_A$ of (105) from one fit and one enumeration per severity, with the mode as the dot; blue, the 5–95 % of the grid reference's draws through the same enumeration, with the median. The Laplace interval sits a little above the reference at both ends, the skew of a posterior bounded by one that the symmetric form does not carry, and within the registered 0.02 of it.

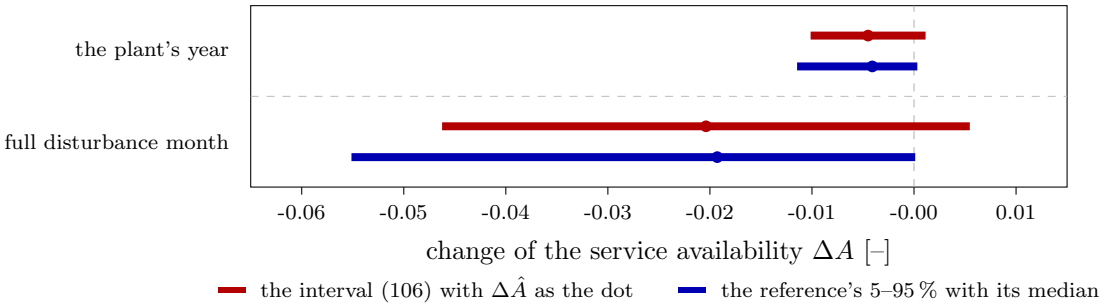


Figure 21: The change of the service availability under the disturbance, in a full disturbance month and over the plant's year, from an ordinary month. Red, the interval $\Delta\hat{A} \pm 1.645\,\sigma_\Delta$ of (106) from one posterior and two enumerations per severity, with the estimate as the dot; blue, the 5–95 % of the grid reference's per-draw differences through the same enumerations, with the median. The dashed line is no change.

is unobservable in any test below it and binds without warning in a test above it. The trip test is read after settling, so the time a hall takes to reach its difference, which a transient model with the halls' thermal mass would give, is an input here. A sensor's bias is taken constant over a test and the prior on it is what sets the floor; a bias that drifts, or a prior that is too narrow, would make the posterior overconfident, which the coverage check would show on a real record as it did not on these. The rate posteriors assume Poisson failures in operating time and a stationary rate; a plant whose rate changes with age or season has a record that says less than its count suggests. The feeder stays a declared support, and the record updates its unavailability, not its topology. The disturbance's effect is a factor exponential in a logged severity index, and the log tests that form only at the three severities the schedule holds; a direct common-cause event, one fault that takes both units out at once, is a mechanism this model does not carry, and would need its own event on the graph. The delta method is linear and the preposterior plug-in, so the third arm's rankings are design-time quantities, not the band a real measurement would deliver. And the system has three components, so that every structure and every band could be checked against a closed form; the constructions are the same on a plant with hundreds, where the enumeration is truncated and carries its own bound.

How much of the process is particular to this plant is a fair question, since every arm ends

in a decision. The decisions are rules on a posterior quantity: a threshold on the probability of acceptance (29) and on the posterior probability of a sensitivity (94), a loop that inverts one monotone statistic, and a branch on whether the claim is settled; none of them belongs to cooling. What each arm computes beneath its rule is general in the same way: a fit and its curvature, an evidence, a posterior over discrete hypotheses, an exceedance, an information matrix, the gradient of an enumeration, and a Poisson count row on a log-rate tied to a shared severity, each of which takes any graph TERRA can write. What is particular is the shape in which the rules were composed here, and it is narrower than Figures 10 to 13 suggest. One scalar, the conductance, carries the whole structural question, so the claim's relation crosses it once and the success structure has two values; the probability of acceptance is then a mass on one side of that crossing on a one-dimensional grid (33), and the power follows from a single critical difference because that probability falls monotonically in one statistic.

A plant whose structure turned on two quantities at once, two openings, or an opening and a damper, would carry a region in place of the point. The grid posterior extends to two or three dimensions at a cost that grows with each; the mixture over structures becomes a mixture over as many hypotheses as the region separates, with the structure term of (79) summed over them; and the power loses its one-dimensional shortcut and reverts to a linearized rule or to simulated tests. The covariance of Section 3.6.1 is block-diagonal because the structural parameter and the rates were learned from different records; a plant whose failure rates depend on the operating point would couple them, and the gradient of the enumeration would then carry a term the step in the conductance hides here. The record arm reads counts of failures and repairs, which every plant keeps, and the service verdict is read through the heat-flow enumeration of the companion study, which is the one piece bound to cooling. The honest statement is that the questions and the operations are general, that the shape in which they were composed here, one scalar, two structures, independent records, is the simplest the questions admit, and that a second system of a different shape is the test the composition has not yet had.

What a trip test catches follows from what it measures. The test reads the opening's conductance in the outage state, so it catches everything that changes that conductance: a grille or a filter in the opening that has fouled, an obstruction stacked against it, a damper left part-closed, a door propped or shut, and any re-fit of the halls that moved the airflow path. And because the test trips a unit, it reads the opening as the outage will find it, with the surviving unit's airflow pattern and the whole of hall B's heat crossing, rather than in the normal state in which the opening carries little and a partial blockage shows nothing. What the test cannot catch is a change that moves the allowable instead of the opening: a tighter IT inlet limit, written into a tenant's contract or a new equipment class, is a decision that lowers the allowable rise in (2) without touching the plant, so the last test must be re-read against the new allowable rather than re-run; and a hard cap on the flow, which a steady test below it cannot see, as said above. The re-test interval follows from the mechanisms: a trip test is due whenever the opening has been touched or the halls re-fitted, and after any outage in which a hall ran hot, since those are the events that change the conductance; a changed inlet limit calls for arithmetic, not a test; and the calendar interval between them need only be short enough to catch the slow mechanism, fouling, before it takes the margin the last test showed.

What the report does not do is as much a part of its answer as what it does. It does not accept the near design from a load-bank test, because no test at that load can; it does not name the disturbance in its own registered month, because that month looked calm; and it does not rank the trip tests above the record once the structure is settled, because they then move nothing the engineer cares about. A redundancy claim that carried these three statements beside its number would say what the number rests on. Table 16 collects the answers of the report.

Table 16: The answers of this report, one row per question a reliability engineer would ask of this plant and its records, in the order of the Conclusion. Every entry is read from the committed results file.

question	answer	why, with the deciding number	what would change it
Is my N+1 claim true, from a trip test, when the opening's capacity is well above the load?	yes	the probability of acceptance (29) is 0.97 from a 20 kW test and 0.999 at full load	nothing identified
Is my N+1 claim true when the capacity sits within ten percent of the load?	not from a load-bank test	0.53 at 20 kW, 0.82 at 40 kW and 0.99 at full load, against 0.95	a matched sensor pair (Table 5), or a hotter test where the allowable permits
Is an 80 kW opening acceptable at 95 % from a 20 kW load-bank test?	no	its probability of acceptance is 0.88 exact and 0.79 by the closed form, (29); a 40 kW test accepts it at 0.978	a 40 kW or full-load test
How much margin above the step does an acceptable design need?	conditional, on the test	the exact 0.95 crossing is in the 70 / 80 / 100 kW cells for 60 / 40 / 20 kW tests (Table 4); power 0.9 needs more	a tighter floor on the hall difference
Is the posterior on the full-load rise calibrated?	yes	coverage 171 / 199, 173 / 194, 176 / 199 of the 90 % interval by (41), inside the registered band	nothing identified
How uncertain is my service availability given three unit-years of record?	yes, as a band	[0.9938, 0.9993] around the asserted 0.9975 by (51); a factor of about 9 in unavailability	a longer record
Does the Laplace posterior on the rates suffice?	conditional, on the count	spread within a few percent of the exact at every record; location off by 0.06 in the log-rate at seven failures, 0.12 at two (Table 6)	the exact draws, which the band uses
What does an unsettled structure do to the band?	yes, it dominates	the mixture's 5 % falls to 0.853 against 0.9938 settled, by (69)	settling the structure (the first question)
Which measurement tightens the band most while the structure is unsettled?	yes	the matched sensor pair, sd 0.048 to 0.029 by (80); every trip test beats every record	nothing identified
And once it is settled?	yes	one more year of the feeder's record, sd 0.0044 to 0.0039; the trip tests take off nothing (a bound)	nothing identified
Did a shared disturbance act on the units?	yes on unit B; unit A at the threshold	$P(\beta > 0 \mid E)$ is 0.970 for unit B by (94), against 0.95; unit A's is 0.952 in the Laplace form and 0.939 on the grid, either side of the line; a full disturbance month multiplies their rates by 2.7 and 2.7 at the mode	more disturbed months for unit A
And on the feeder?	cannot tell	$P(\beta > 0 \mid E)$ is 0.63; its factor's 90 % interval [0.5, 2.9] by (95) holds no effect and the true two	more disturbed months, or a sharper severity index
What does the disturbance do to the service availability?	it lowers it	$\Delta\hat{A}(1) = -0.0204$ by (101), [-0.0462, 0.0055] by (106), $P(\Delta A < 0 \mid E) = 0.903$ by (106); from 0.9826 in an ordinary month to 0.9622 in a full disturbance month, -0.0045 over the year	nothing identified
Is the opening's heat-flow capacity itself estimated?	no	steady readings determine the conductance; the capacity is read as $G\,\Delta T_{\text{allow}}$ and a hard cap is unobservable until it binds	a test that drives the opening to its cap

6 Conclusion

This report set out to show that the inputs a cooling-redundancy claim asserts, the success structure, the component unavailabilities and their independence, can each be tested against a record the plant keeps, on a system small enough that every answer could be checked against a closed form. The aim was the process, not the numbers, and the reason for the process is that an asserted input fails silently while a tested one carries its confidence and its remedy: the two-hall system is a toy, and the same four steps apply on a plant with hundreds of components, with the enumeration truncated by probability and the mass left out bounded, as the forward study's enumeration already provides; that is a generality of construction, and not a scalability these studies demonstrate.

The process was the same for every question. The question was written as a reliability engineer asks it. The plant's record was written as readings with stated noise and priors on a factor graph of the network, or as counts over operating time. The engine then did the one thing it does: it combined the physics rows, the reading rows and the prior rows into a posterior and read the answer off it, a fitted conductance with its uncertainty, the probability that a claim on it holds, a band on a failure rate, a sensitivity to a shared disturbance with its spread. The availability number followed from the enumeration of the forward study, re-run on the posterior instead of on asserted values, so that every uncertainty the record carries reached the service availability.

What the engine contributed at each step. For the $N + 1$ claim, the trip test was a maximum-a-posteriori fit of the opening's conductance with a Laplace width, and the claim's probability was read as the posterior mass of the full-load rise within the allowable, which turned a yes-or-no assertion into a number with its confidence, the regime in which a test cannot tell, and the test that would resolve it. For the outage record, a Poisson factor on each log-rate gave a posterior that the enumeration propagated into a band on the service availability, with the Laplace width checked against exact quadrature where the count was small. For the choice of measurement, the preposterior variance of the service availability, by the delta method on the same posterior, ranked candidate tests and records by what each would take off the band. For the shared disturbance, the nine entries of the three components' rate models were fitted at once from the disturbance log's count rows, and the posterior probability that each sensitivity is positive was read from the covariance. For the change of the service availability under the disturbance, the same posterior went through a second query, two enumerations and the difference of their gradients, so that the cost of the named effect to the service came with its spread and no new inference.

What the process found, in words: a trip test establishes the claim when the opening's capacity is well above the load and cannot establish it from a load-bank test when the capacity is within a tenth of it; an unsettled structure is a wider uncertainty than any record the plant keeps; the measurement worth taking depends on whether the structure is settled; three years of disturbance log name the disturbance on one unit, leave the other at the threshold and the feeder open, and a full disturbance month takes 0.020 off the service availability, a loss at the posterior probability 0.90. Each verdict carries the regime in which it cannot tell, and Table 16 holds the numbers. Table 15 is the form of the result: twelve comparisons, six of which hold the engine against an exact answer and six of which hold it against what the plant has today.

The next step is the record this study did not have: a real trip test and a real outage log, on which the same four steps run without a generator behind the readings.

A Reproduction

```
PYTHONPATH=src python docs/terra_graph_engine/case_studies/foundation_hvac/\
    two_hall_reliability_from_readings/run.py
```

```
PYTHONPATH=src python docs/terra_graph_engine/case_studies/foundation_hvac/\
    two_hall_reliability_from_readings/make_report.py --pdf
```

The first command refuses to run unless `preregistration.json` is committed and unmodified, computes the four arms under it and writes `results.json` with the record of the registration it ran under, registered on 2026-10-16:

> path `docs/terra_graph_engine/case_studies/foundation_hvac/two_hall_reliability_from_readings/preregistration.json`
> LF sha256 `c6268b6e0151423d055fd3000ad09e407943b24cf8070a07632ee88dff3e54f5`
> git blob `d942768fd4344cfcdbf93a21a86c37a7d4772b3f`
> commit `b56c3827f68c8fbed2817b11a48704bd9504e630`

The root seed is 20261016, spawned into five streams in the order trip-test draws, coverage, records, the disturbance log, propagation. The run takes about two minutes; the second command renders this document from the results file, byte-identical on re-render, and builds the PDF with latexmk. The gate test under `tests/plugins/hvac/papers/` pins the verdict strings and bands, the registration record and commit order, two closed-form checks, one fresh full run, and the rendered document.

B Data and problems

Every reading and record is synthetic: generated by `readings_model.py` and `arms.py` from the registered design, noise, rates and seeds, with the generator committed beside the results under the repository's license. No external dataset is used. The forward reliability model of the two halls is the one of the study's companion, loaded by path and not edited, with its wall conductance set to $C/\Delta T_{\text{allow}}$. The study leaves no corpus problems.

C The two drawings as hypotheses

This appendix carries the registered comparison of the forward study's two drawings, a question apart from the claim of Section 3.4: which of two drawings the readings favor. A design pair is a case's true opening and a drawing on the other side of the load: the far case pairs its 80 kW with a 30 kW drawing, the near case its 66 kW with a 54 kW drawing; the comparison is between the two drawings of a pair. A redundancy claim is a statement about structure: that the plant survives the loss of one unit. The drawing asserts it, and nothing on the drawing can test it, because the drawing is the hypothesis. The question this arm answers is whether one test, a trip of unit B at a load the plant can run, can tell the claimed structure from its alternative, and with what confidence. It matters because the claim is signed off once and relied on for years: a false confirmation leaves hall B to overheat on the first real outage, and a false refutation buys a unit the plant did not need. In general terms, two hypotheses H_1 and H_0 each predict the readings y a test would produce, and the readings decide between them through the ratio of the probabilities each assigns to what was read, the Bayes factor,

$$B = \frac{p(y \mid H_1)}{p(y \mid H_0)}, \tag{109}$$

so that the posterior odds on H_1 are the prior odds multiplied by B and, at equal prior odds, the posterior probability of H_1 is $B/(1 + B)$. Its logarithm, in nats, is the evidence the test carries; the claim is confirmed when it clears a decision gap fixed before the test, refuted when the

alternative clears it, and undecided between. Here the hypotheses are the two success structures of the forward model, drawn as fault trees in Figure 11, each with its validity condition: H_{OR}, the opening cannot pass one hall's load and either unit's loss loses the service; H_{AND}, it can, and both must fail. A design pair gives each tree the full-load rise its drawing implies by (1) and (3), the narrow drawing's ΔT_0^{OR} for the OR tree and the wide drawing's $\Delta T_0^{\mathrm{AND}}$ for the AND tree, and each tree predicts the hall difference a test at load Q will show; $p(y \mid H)$ is the evidence (12) of the graph of Section 3.3 under that tree's prior, and the rest of this section specializes (109) to the two. The ratio is evaluated in two ways, through TERRA and by a closed form that exists for this case alone, and the two are set beside each other in Section C.3. Both are computed three ways: on the *design case*, the readings the model predicts at the truth with the biases at their prior mean, which is seed-free and decides what a test is expected to say; on the registered seed's single realization, whose gap scatters around the design case by the predictive spread the closed form gives; and over 200 realizations per case, whose verdict counts say how often one test says it. Every realization is a fresh fit of the same graph; nothing is reused between them but the graph.

C.1 The evaluation through TERRA

The verdict on the design's claim is *confirmed* when the tree the design claims leads by at least a decision gap Δ^* of 1 nat, odds of e to one, about 2.7 to one; *refuted* when the other tree leads by the gap; *cannot tell* between. Figure 22 lays the steps out. What the verdict reads is the log Bayes factor, the evidence for the AND tree over the OR tree,

$$\Delta = \log Z_{\mathrm{AND}} - \log Z_{\mathrm{OR}}, \tag{110}$$

in nats, with $\log Z$ the log evidence (12) of a graph: the probability of the readings under that graph once the freedom of its unknowns has been paid for, which TERRA returns from the fit (7) and the curvature (9) at its mode. So (110) needs two graphs, one per tree. They are two copies of the graph of Section 3.3, identical in every row of Table 2 but one: the prior on the opening, the lognormal law on the full-load rise centered at the rise the drawing implies, with the drawing's tolerance τ of 2 % as its spread,

$$\ln \Delta T_{\mathrm{full}} \sim \mathcal{N}(\ln \Delta T_0,\ s_G^2), \tag{111}$$

$$s_G = \log(1+\tau), \tag{112}$$

with ΔT_0 each drawing's rise, one hall's load times the allowable rise of 8 K over the drawing's capacity by (1) and (3), which for the near case is

$$\Delta T_0^{\mathrm{AND}} = 7.27 \text{ K}, \tag{113}$$

$$\Delta T_0^{\mathrm{OR}} = 8.89 \text{ K}, \tag{114}$$

and 6.00 and 16.0 K for the far case. The rows and the prior (111) make a graph per tree, TERRA fits each and returns its evidence, (110) is their difference, and the verdict is its reading against Δ^*. The tolerance is the one lever in it: a wider τ would put prior spread on top of the reading floor (24) and halve the gap, so the graphs carry the drawings' own tolerance and nothing looser. The closed form of the next section is the check: it exists because the two graphs differ in a single number.

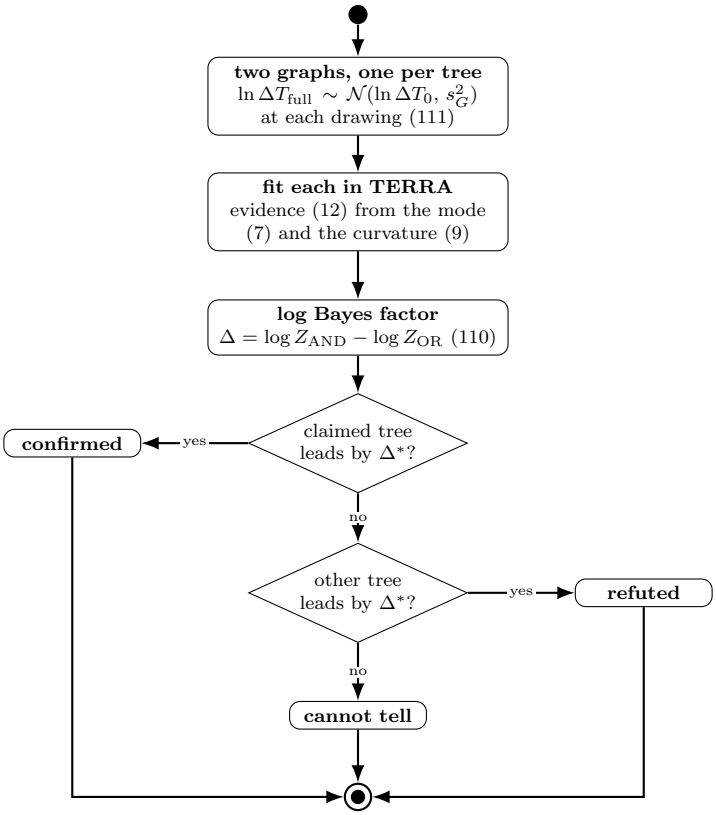


Figure 22: The steps from the two drawings to the verdict on the claim. Two graphs are built, one per tree, differing only in the prior on θ (111); TERRA fits each and returns its evidence; the log Bayes factor (110) is then read against the decision gap twice, and the verdict is confirmed when the claimed tree leads by the gap, refuted when the other tree does, and cannot tell between. The closed form of Section C.2 checks the factor.

C.2 The closed-form reference

A trip test at load Q sends Q through the opening, so each tree, being a fixed full-load rise, predicts one hall difference, its rise scaled by the test load over the hall load,

$$\mu_{\text{OR}} = \frac{Q}{Q_{\text{hall}}}\,\Delta T_0^{\text{OR}}, \tag{115}$$

$$\mu_{\text{AND}} = \frac{Q}{Q_{\text{hall}}}\,\Delta T_0^{\text{AND}}. \tag{116}$$

The observed difference ΔT_{obs} is the true one plus the two sensors' noise and biases, so under either tree it is normal about that tree's prediction with the floor (24) as its spread,

$$p(\Delta T_{\text{obs}} \mid H) = (2\pi\,\sigma_{\Delta T}^2)^{-1/2} \exp\!\left(-\frac{(\Delta T_{\text{obs}} - \mu_H)^2}{2\,\sigma_{\Delta T}^2}\right). \tag{117}$$

The meters beside the differenced channel predict the same numbers under both trees, so their factors cancel in any ratio, and the comparison lives on the differenced channel alone. The evidence for the AND tree over the OR tree is the logarithm of the ratio of the two probabilities, the log

Table 17: The design case, with the truth the wide drawing. Columns: the expected hall difference by (116); the evidence for the AND tree by (119) at that difference; the verdict at the decision gap of 1 nat; the Laplace evidence (12) of the two graphs that differ only by the drawing's prior on the conductance, with its verdict. With the truth the narrow drawing the gaps change sign and the verdicts are the same words.

pair [kW]	load [kW]	ΔT [K]	gap (119) [nats]	verdict	Laplace form [nats]	verdict
far (30 / 80)	20	2.00	10.98	confirmed	10.05	confirmed
far (30 / 80)	40	4.00	43.92	confirmed	40.01	confirmed
far (30 / 80)	60	6.00	98.81	confirmed	82.21	confirmed
near (54 / 66)	20	2.42	0.29	cannot tell	0.10	cannot tell
near (54 / 66)	40	4.85	1.15	confirmed	0.95	cannot tell
near (54 / 66)	60	7.27	2.58	confirmed	2.29	confirmed

Bayes factor (Kass and Raftery 1995), in which the normalizing factors are identical and vanish,

$$\Delta = \log p(\Delta T_{\mathrm{obs}} \mid H_{\mathrm{AND}}) - \log p(\Delta T_{\mathrm{obs}} \mid H_{\mathrm{OR}}), \tag{118}$$

$$= \frac{(\Delta T_{\mathrm{obs}} - \mu_{\mathrm{OR}})^2 - (\Delta T_{\mathrm{obs}} - \mu_{\mathrm{AND}})^2}{2\,\sigma_{\Delta T}^2}, \tag{119}$$

in nats: positive when the observation sits closer to the AND prediction than to the OR prediction, measured in units of the floor, and its size is how much closer; the verdict is read at the same decision gap as in Section C.1. Expanding the squares cancels the $\Delta T_{\mathrm{obs}}^2$ terms and leaves Δ linear in the observation,

$$\Delta = \frac{(\mu_{\mathrm{AND}} - \mu_{\mathrm{OR}})\,(2\,\Delta T_{\mathrm{obs}} - \mu_{\mathrm{OR}} - \mu_{\mathrm{AND}})}{2\,\sigma_{\Delta T}^2}, \tag{120}$$

with slope $(\mu_{\mathrm{AND}} - \mu_{\mathrm{OR}})/\sigma_{\Delta T}^2$ in the observation; the observation scatters with standard deviation $\sigma_{\Delta T}$, so one realized test scatters about its expected gap with the predictive standard deviation

$$\sigma_{\Delta} = \frac{|\mu_{\mathrm{OR}} - \mu_{\mathrm{AND}}|}{\sigma_{\Delta T}}. \tag{121}$$

At the design case the observation equals the AND prediction and the expected gap is $(\mu_{\mathrm{AND}} - \mu_{\mathrm{OR}})^2/(2\sigma_{\Delta T}^2)$, which is $\sigma_{\Delta}^2/2$: the expected verdict and its own scatter are one number seen twice, so a test with a small expected gap is also a test whose single realization is unreliable. The closed form (119) is arithmetic on three numbers, the two predictions and the floor, and passes through no engine; it is the reference for Section C.1. The two forms are computed for a reason. The closed form (119) exists only because the hypotheses differ in a single number and every other reading predicts the same under both; change the problem a little and no closed form exists, and the evidence must come from (12), the integral over the full graph with the biases, the meters and the tolerance all in it. Agreement between the two where both exist is what licenses (12) where only it does, and their difference is the price of the realism (119) leaves out.

C.3 The comparison on the registered tests

Each of the six registered tests is one realization of the trip test of Section 3.3, its readings generated by the measurement model (20) to (22) from the truth, the wide drawing, with the registered noise, biases and seed. Figure 23 sets the readings against the two hypotheses. The dot is the observed

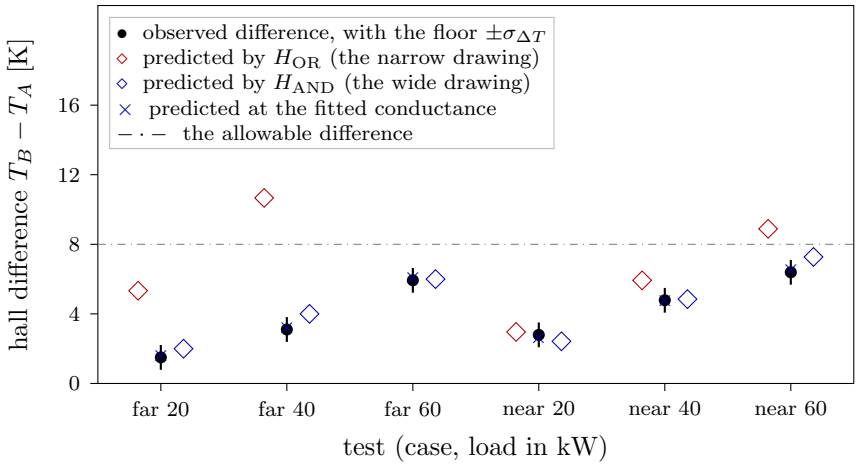


Figure 23: Readings against predictions for the six trip tests of the registered seed. Dot, the observed difference $y_B - y_A$ of (20) and (21), with the floor (24) as its bar; diamonds, the differences the two trees predict by (115) and (116); cross, the difference at the conductance fitted by (7). On the far case the two predictions are many floors apart at every load; on the near case they are within a floor at 20 kW and two apart at 60 kW. The truth is the wide drawing in every test.

difference $y_B - y_A$, with the floor (24) as its bar; the diamonds are the hall differences the two trees predict, (115) and (116), with each drawing's full-load rise from (1) and (3); the cross is the difference at the conductance fitted by (7), the observation with the biases estimated out. On the far case the two predictions are many floors apart at every load; on the near case they are within one floor at 20 kW and two at 60 kW.

Table 17 is the design case, the readings the model predicts at the truth with the biases at their prior mean. Its evidence column is (119) evaluated at the predicted difference, where by (120) it equals half the square of the predictive spread (121); the verdict follows the rule of Appendix C at the decision gap of 1 nat; and the Laplace column is the evidence (12) of the two graphs that differ only in their prior on the opening's entry. For the far case the gap is 11.0 nats at the 20 kW load-bank test and 99 at full load, confirmed at every load; with the truth the narrow drawing the gaps change sign and the verdicts are the same words, since (119) only changes sign when the two predictions are exchanged. For the near case the gap is 0.29 nats at 20 kW, cannot tell, and 2.6 nats at full load, confirmed. The Laplace form gives 10.1 and 2.3 nats for the same two cases, lower by the cost of the drawing's 2 % prior spread, which the volume terms of (12) charge and (119) does not; the verdicts agree except at the near case's 40 kW, where 1.15 nats becomes 0.95 and crosses the gap. Figure 24 plots the design-case evidence of Table 17 against the test load, with the registered seed's realization as a cross, and Table 18 lists that realization: the observed difference; the realized gap, (119) at the observed rather than the predicted difference; its predictive spread (121); the verdict; the fitted conductance and its posterior spread by (7) and (11); and the probability of acceptance (29), as the normal tail (36) and as the exact quadrature. The near case's 20 kW realization is the case to read: a gap of -0.11 nats against a predictive spread of 0.76, a fitted conductance of 7590 W/K, a full-load rise by (1) of 7.91 K, 1 % under the allowable with a posterior width of 26 %, and a probability of acceptance of 0.53. That test sat on the fence. Its full-load realization confirmed the design by 5.4 nats with a probability of acceptance of 0.99.

Table 19 counts the verdicts of (119) over 200 realizations per case. The far case is confirmed in 200 of 200 at 40 and 60 kW and in 198 at 20 kW. The near case at 20 kW cannot tell in 151, confirms the design in 41 and refutes a true claim in 8, the scatter (121) predicts for an expected

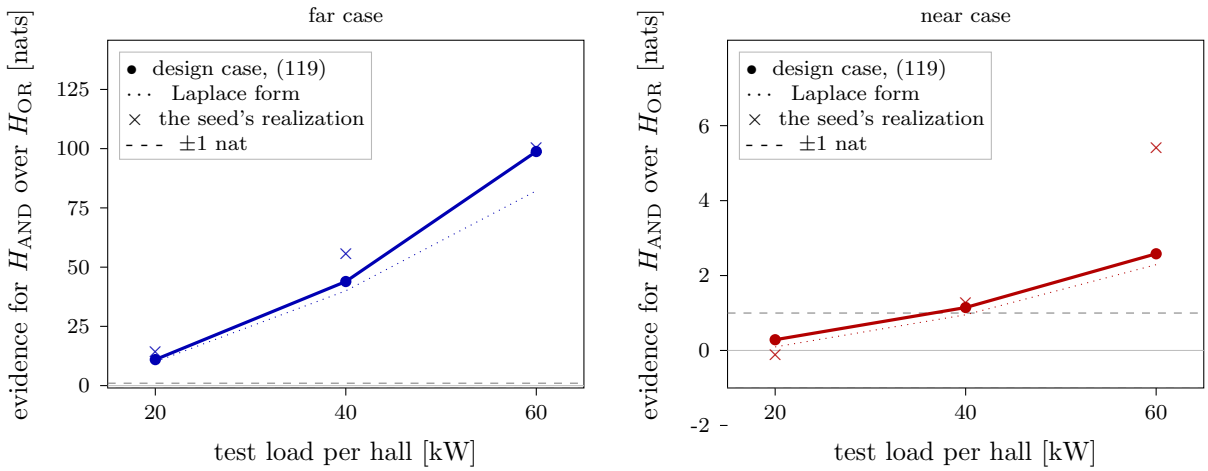


Figure 24: The evidence a trip test yields for the AND tree over the OR tree against the test load, for the far case (left) and the near case (right, on its own scale), with the truth the wide drawing. Dots, (119) on the design case; dotted, the Laplace form (12) at the drawing's tolerance; crosses, (119) at the registered seed's observed difference; dashed, the decision gap. The far case clears the gap by an order of magnitude at every load; the near case sits inside the gap at 20 kW and clears it at 60 kW on the design case, while its realizations scatter across the gap.

gap of 0.29; at full load, where the design case confirms by 2.6 nats, 39 of 200 single tests still cannot tell and 10 refute, because the predictive spread of 2.3 nats is nearly the expected gap itself.

The honest statement for a design within a tenth of the step is that the design separates and one realization may not. For the owner, the trip test is the cheapest way to learn whether the claim's structure is the plant's, and the design case of (119), computable from the drawings and the sensor specification before any test is run, says in advance whether one test will be enough.

What the two columns of Table 17 demonstrate is the method on a case with a known answer. The closed form (119) is arithmetic any reader can repeat; the Laplace form (12) is the engine scoring two nine-unknown graphs. They give the same verdict in five rows of six and numbers that differ by what the fuller model adds: (12) charges the drawing's 2 % tolerance as prior spread and marginalizes the biases instead of fixing the floor, which is why it is lower in every row and by more at higher load, 1 nat at 20 kW and 17 at 60 kW on the far case, where the fixed floor of (119) is most generous. The method is therefore conservative relative to the shortcut, and trustworthy where the shortcut cannot be written down. The near case's 40 kW row shows the other side: 1.15 against 0.95 nats, one verdict each side of the gap, so a verdict that clears the gap by a hair is not a verdict, and the report goes on to the probability of acceptance rather than resting on the word. The three ways the evidence is computed answer three different questions: the design case says what a test is expected to say, the seed says what one test did say, and the 200 draws say how often one test says what the design case expects. The gap by (119) grows as the square of the test load, 0.29, 1.15 and 2.58 nats at 20, 40 and 60 kW on the near case, so the load is the lever, and it buys evidence quadratically until it reaches the allowable rise.

Table 18: The registered seed's single realization of each test, the truth the wide drawing. Columns: the observed difference of (20) and (21); the realized gap, (119) at that difference; its predictive standard deviation (121); the verdict; the fitted conductance by (7) and the posterior spread s of its entry θ by (11); the probability of acceptance (29) as the normal tail (36), as the discrete form (33) and as the exact one-dimensional quadrature.

pair	load [kW]	$y_B - y_A$ [K]	gap [nats]	$\pm$ [nats]	verdict	$\hat{G}$ [W/K]	s [–]	Laplace P [–]	discrete P [–]	exact P [–]
far	20	1.50	14.28	4.69	confirmed	12614	0.406	0.900	0.969	0.969
far	40	3.10	55.75	9.37	confirmed	12395	0.215	0.990	0.999	0.999
far	60	5.93	100.22	14.06	confirmed	9900	0.116	0.991	0.999	0.999
near	20	2.79	-0.11	0.76	cannot tell	7590	0.261	0.518	0.534	0.534
near	40	4.78	1.29	1.51	confirmed	8428	0.146	0.787	0.820	0.820
near	60	6.39	5.40	2.27	confirmed	9206	0.106	0.973	0.991	0.991

Table 19: The verdict over two hundred realized tests per case, the truth the wide drawing: each realization is a draw of (20) to (22), each gap is (119) at its own observed difference, and each verdict is read at the decision gap. The design case decides what a test is expected to say; these counts say how often one test says it.

pair	load [kW]	confirmed [count]	cannot tell [count]	refuted [count]	tests [count]
far	20	198	0	2	200
far	40	200	0	0	200
far	60	200	0	0	200
near	20	41	151	8	200
near	40	119	65	16	200
near	60	151	39	10	200

D Two reductions kept out of the main text

D.1 The preposterior structure term

The structure term of (80) needs the structure probability a candidate will leave, which depends on readings not yet taken; what is known is its distribution. The candidate narrows the posterior of the opening from the spread s, its entry of Σ, to s', the same entry of Σ_a; by the law of total variance applied to that entry, the present variance s^2 is the expected future variance s'^2 plus the variance of the future mean, so the future mean m' is normal about the present mean m with variance $s^2 - s'^2$, and the future structure probability at that mean is the tail (36) with the future spread and the future rise at that mean,

$$p'(m') = \Phi\left(\frac{\ln \Delta T_{\text{allow}} - \ln \widehat{\Delta T}'_{\text{full}}}{s'} \right), \tag{122}$$

$$\widehat{\Delta T}'_{\text{full}} = Q_{\text{hall}}\, e^{-m'}. \tag{123}$$

The structure term after the candidate is the expectation of $p'(1-p')$ over where m', and with it the rise (123), may land,

$$\mathbb{E}[p'(1-p')] = \int p'(m')\,\big(1 - p'(m')\big)\,\mathcal{N}(m' \mid m,\ s^2 - s'^2)\,dm', \tag{124}$$

evaluated as a one-dimensional quadrature: m' on a grid of 2 001 points over eight spreads either side of m, each weighted by its normal density, and the products summed. A candidate that adds nothing on the conductance leaves $s' = s$ and the term at $p(1-p)$.

D.2 The gradient of the service availability in the disturbance's entries

The spread (102) takes the gradient of (97) in the nine entries at the fitted state. The enumeration gives the Birnbaum importance (63) of each component, and the chain rule of (84) carries it to the entries: per component,

$$g_{\alpha,i} = -I_B(i)\,\hat{q}_i\,(1-\hat{q}_i), \tag{125}$$

$$g_{\beta,i} = d^*\,g_{\alpha,i}, \tag{126}$$

$$g_{\gamma,i} = -g_{\alpha,i}, \tag{127}$$

the sensitivity's entry the baseline's scaled by the severity, zero at $d^* = 0$, and the repair entry's the baseline's with the sign reversed, a faster repair lowering the unavailability that a higher failure rate raises. The change's gradient (103) is the difference of two such vectors, one per enumeration; at $d^* = 0$ the sensitivity's entries vanish, so the change's sensitivity entries are the disturbed enumeration's alone. Appendix E evaluates them for unit A.

E Three worked examples

This appendix carries three of the report's chains by hand: the far case's 20 kW trip test through Sections 3.3 and 3.4 and Appendix C, one component's three-year record through Section 3.5, and unit A's three years of disturbance log through Section 3.7, so that a reader can see each number of the tables arise from the inputs of Table 1. Every number is read from the committed results and the registration; the derivations are unnumbered.

E.1 The trip test

The far case is the 30 and 80 kW openings, and by (1) and (3) each drawing gives its tree a full-load rise, one hall's load times the allowable difference of 8 K over the capacity,

$$\begin{aligned}
\Delta T_0^{\mathrm{OR}} &= Q_{\mathrm{hall}}\,\Delta T_{\mathrm{allow}}/C_{\mathrm{OR}},\\
&= 60\,000 \times 8/30\,000,\\
&= 16.000\ \mathrm{K},\\
\Delta T_0^{\mathrm{AND}} &= Q_{\mathrm{hall}}\,\Delta T_{\mathrm{allow}}/C_{\mathrm{AND}},\\
&= 60\,000 \times 8/80\,000,\\
&= 6.000\ \mathrm{K}.
\end{aligned}$$

At a test load Q of 20 kW per hall, all of hall B's load crosses the opening by (18), and the closure law (19) makes each tree's predicted hall difference its rise scaled by the test load over the hall

load,

$$\begin{aligned}\mu_{\mathrm{OR}} &= (Q/Q_{\mathrm{hall}})\,\Delta T_0^{\mathrm{OR}},\\ &= (20\,000/60\,000)\times 16.000,\\ &= 5.333\ \mathrm{K},\\ \mu_{\mathrm{AND}} &= (Q/Q_{\mathrm{hall}})\,\Delta T_0^{\mathrm{AND}},\\ &= (20\,000/60\,000)\times 6.000,\\ &= 2.000\ \mathrm{K}.\end{aligned}$$

The floor on the observed difference, (24) in its hand form, comes from the sensor noise σ of 0.3 K per sample averaged over $N = 30$ samples and the bias prior s_b of 3.0 K on each sensor,

$$\begin{aligned}\sigma_{\Delta T} &= \sqrt{2\,(\sigma^2/N + s_b^2)},\\ &= \sqrt{2\,(0.3^2/30 + 3.0^2)},\\ &= \sqrt{2\,(0.0030 + 9.0000)},\\ &= 0.711\ \mathrm{K},\end{aligned}$$

of which the bias term is 100 %. The registered seed's test averaged the hall sensors at $y_A = 21.567$ and $y_B = 23.066\,^{\circ}\mathrm{C}$ (Table 18), so the observed difference is

$$\begin{aligned}\Delta T_{\mathrm{obs}} &= y_B - y_A,\\ &= 23.066 - 21.567,\\ &= 1.498\ \mathrm{K},\end{aligned}$$

closer to the wide drawing's 2.000 K than to the narrow drawing's 5.333 K. The evidence for the AND tree over the OR tree by (119) is the difference of the two squared misfits over twice the squared floor,

$$\begin{aligned}\Delta &= \frac{(\Delta T_{\mathrm{obs}} - \mu_{\mathrm{OR}})^2 - (\Delta T_{\mathrm{obs}} - \mu_{\mathrm{AND}})^2}{2\,\sigma_{\Delta T}^2},\\ &= \frac{(1.498 - 5.333)^2 - (1.498 - 2.000)^2}{2\times 0.711^2},\\ &= \frac{14.707 - 0.252}{1.012},\\ &= 14.28\ \mathrm{nats},\end{aligned}$$

positive and above the decision gap Δ^* of 1 nat, so the verdict is *confirmed*. On the design case the observed difference is the wide drawing's own prediction, the second square vanishes, and the same arithmetic gives the entry of Table 17, while by (121) one realization scatters about it with the standard deviation σ_Δ,

$$\begin{aligned}\Delta &= (2.000 - 5.333)^2/1.012,\\ &= 10.98\ \mathrm{nats},\\ \sigma_\Delta &= |5.333 - 2.000|/0.711,\\ &= 4.69\ \mathrm{nats},\end{aligned}$$

so the seed's realization sits 0.7 such deviations above its design case. The probability of acceptance by the Gauss–Newton closed form (40) is a design-case quantity: with the true capacity C at the wide drawing's 80 kW, the true full-load rise is the drawing's 6.000 K and the test's own difference at 20 kW is 2.000 K,

$$\begin{aligned} P_{\mathrm{GN}} &= \Phi\!\left(\frac{\Delta T_{\mathrm{test}}\,\ln(\Delta T_{\mathrm{allow}}/\Delta T_{\mathrm{true}})}{\sigma_{\Delta T}}\right), \\ &= \Phi\!\left(\frac{2.000\times\ln(8/6.000)}{0.711}\right), \\ &= \Phi\!\left(\frac{0.575}{0.711}\right), \\ &= \Phi(0.809), \\ &= 0.791, \end{aligned}$$

where the log clearance of the allowable is 0.288 and the spread s the test leaves is $\sigma_{\Delta T}/\Delta T_{\mathrm{test}} = 0.356$, so the fitted full-load rise clears the allowable by 0.81 standard deviations. The exact value by quadrature at the same truth, the last column of Table 4, is 0.883: the closed form treats the posterior of the rise as normal in its logarithm where it is skewed toward small rises, and understates the probability by 0.092. The seed's own realization fitted the conductance at 12 614 W/K, a full-load rise of 4.757 K by (1), with a posterior spread s of 0.406 (Table 18), and the same normal form on that fit, the Laplace column of the table, reads

$$\begin{aligned} P_{\mathrm{acc}} &= \Phi\big((\ln\Delta T_{\mathrm{allow}} - \ln\widehat{\Delta T}_{\mathrm{full}})/s\big), \\ &= \Phi\big(\ln(8/4.757)/0.406\big), \\ &= \Phi(1.281), \\ &= 0.900, \end{aligned}$$

against 0.969 by the exact quadrature on the same readings.

E.2 The record

Unit A's three-year record is $k = 7$ failures over $T = 26\,280$ operating hours, 3.00 years (Table 6). The prior on the log failure rate $\theta = \log\lambda$ is normal with mean $m = \ln 2.5 = 0.916$, the rate the claim was built on, and spread $s_\lambda = \ln 3 = 1.099$. Adding the prior's quadratic $(\theta - m)^2/(2s_\lambda^2)$ to (54) and setting the derivative to zero gives the condition the posterior mode $\hat\theta$ satisfies, with T in years and λ per year,

$$\begin{aligned} Te^{\hat\theta} - k + (\hat\theta - m)/s_\lambda^2 &= 0, \\ 3.00\times 2.350 - 7 + (0.855 - 0.916)/1.207 &= 0.051 - 0.051, \\ &= 0, \end{aligned}$$

at $\hat\theta = 0.855$: the posterior mode of the rate is

$$\begin{aligned} \hat\lambda &= e^{\hat\theta}, \\ &= 2.35 \text{ per year}, \end{aligned}$$

between the record's own $k/T = 2.33$ per year and the prior's 2.5. The curvature of the negative log posterior at the mode is $Te^{\hat{\theta}} + 1/s_\lambda^2$, and the Laplace spread is its inverse square root,

$$\begin{aligned}\sigma_\theta &= \left(Te^{\hat{\theta}} + 1/s_\lambda^2\right)^{-1/2},\\ &= (7.051 + 0.829)^{-1/2},\\ &= 0.356,\end{aligned}$$

the entries of Table 6; the exact posterior's mean sits 0.057 below the mode with the same spread to three decimals. The repair log holds 7 repairs over 1 260 h of downtime, 0.144 years, so the record's own repair rate is $7/0.144 = 48.7$ per year, which is the prior's center of $8760/180 = 48.7$ per year; the same condition then puts the mode at the center, $\hat{\mu} = 48.7$ per year. By (50) the unavailability at the modes is

$$\begin{aligned}q &= \frac{\hat{\lambda}/\hat{\mu}}{1 + \hat{\lambda}/\hat{\mu}},\\ &= \frac{2.350/48.67}{1 + 2.350/48.67},\\ &= \frac{0.0483}{1.0483},\\ &= 0.0461,\end{aligned}$$

against the asserted 0.05. One draw of the service availability by (51) is the enumeration of the parallel pair at that unavailability, the feeder outside the boundary, with both units carrying the same q: the states that deliver are nothing out, unit A out and unit B out, and the state that does not is both out,

$$\begin{aligned}A_{\text{service}} &= (1-q)^2 + q\,(1-q) + q\,(1-q),\\ &= 0.9100 + 0.0439 + 0.0439,\\ &= 0.9979,\end{aligned}$$

which is $1 - q^2$. The asserted value is the same sum at $q = 0.05$, 0.9975. The band of Table 7 is this sum over 2000 draws of the exact posterior rather than the one draw at the mode: its 5–95 % interval is [0.9938, 0.9993], and its median, 0.9979, is where the draw at the mode lands.

E.3 The disturbance

Unit A's three years of log (Table 11) hold 3 failures in 16,740 operating hours at severity 0, 3 in 3,780 at one half and 3 in 3,780 at one. The fit's two conditions on the baseline and the sensitivity, (89) and (90), balance the fitted counts against the logged ones over all months and over the severity-weighted months; at the mode $\hat{\alpha} = 0.864$, $\hat{\beta} = 1.011$ the fitted counts by level are

$$\begin{aligned}a_0 &= 1.911 \times e^{0.864},\\ &= 4.53,\\ a_{1/2} &= 0.432 \times e^{0.864 + 1.011/2},\\ &= 1.70,\\ a_1 &= 0.432 \times e^{0.864 + 1.011},\\ &= 2.81,\end{aligned}$$

against 3, 3 and 3 logged. The two balances (89) and (90) then read, fitted against logged with the priors' pulls,

$$\begin{aligned} 9.04 - 9 + \frac{0.864 - 0.916}{1.207} &= 0.000, \\ 3.66 - 4.5 + \frac{1.011}{1.207} &= 0.000, \end{aligned}$$

both zero to the rounding shown. The information block (92) adds the prior weights $1/s^2 = 0.829$ to the fitted counts and their severity moments,

$$\begin{aligned} \Lambda_{\mathrm{A}} &= \begin{pmatrix} 0.829 + 9.044 & 3.662 & 0 \\ 3.662 & 0.829 + 3.238 & 0 \\ 0 & 0 & 9.829 \end{pmatrix}, \\ &= \begin{pmatrix} 9.872 & 3.662 & 0 \\ 3.662 & 4.066 & 0 \\ 0 & 0 & 9.829 \end{pmatrix}, \end{aligned}$$

the last entry the repair row's $1/s^2 + De^{\hat{\gamma}}$, and its inverse is the covariance,

$$\begin{aligned} \Sigma_{\mathrm{A}} &= \Lambda_{\mathrm{A}}^{-1}, \\ &= \begin{pmatrix} 0.1521 & -0.1370 & 0 \\ -0.1370 & 0.3693 & 0 \\ 0 & 0 & 0.1017 \end{pmatrix}, \end{aligned}$$

so that

$$\begin{aligned} \sigma_\beta &= \sqrt{0.3693}, \\ &= 0.608, \\ \operatorname{corr}(\alpha, \beta) &= \frac{-0.1370}{0.390 \times 0.608}, \\ &= -0.58, \end{aligned}$$

the entries of Table 12. The disturbance's effect by (94) is

$$\begin{aligned} P(\beta_{\mathrm{A}} > 0 \mid E) &= \Phi(\hat{\beta}/\sigma_\beta), \\ &= \Phi(1.011/0.608), \\ &= \Phi(1.664), \\ &= 0.952, \end{aligned}$$

at the registered confidence of 0.95: named, by a hair. The factor a full disturbance month puts on the rate, and its 90 % interval (95), are

$$\begin{aligned} e^{\hat{\beta}} &= e^{1.011}, \\ &= 2.75, \\ \left[e^{1.011 - 1.645 \times 0.608},\ e^{1.011 + 1.645 \times 0.608}\right] &= [1.01,\ 7.47], \end{aligned}$$

an interval that holds the true factor of four and almost no effect alike. In a full disturbance month unit A's unavailability by (84) is

$$\begin{aligned}\hat{q}_{\mathrm{A}}(1) &= \frac{e^{\hat{\alpha}+\hat{\beta}-\hat{\gamma}}}{1+e^{\hat{\alpha}+\hat{\beta}-\hat{\gamma}}},\\ &= \frac{e^{0.864+1.011-3.885}}{1+e^{0.864+1.011-3.885}},\\ &= 0.1182,\end{aligned}$$

and with unit B's 0.1828 and the feeder's 0.0166 the enumeration (100) gives the service availability of the parallel pair behind the feeder, $(1-\hat{q}_{\mathrm{F}})\,(1-\hat{q}_{\mathrm{A}}\hat{q}_{\mathrm{B}}) = 0.9622$, the entry of Table 13. Its gradient's unit-A entries by (125), with the Birnbaum importance $I_B(\mathrm{A}) = 0.1797$, are

$$\begin{aligned}g_{\alpha,\mathrm{A}} &= -I_B(\mathrm{A})\,\hat{q}_{\mathrm{A}}\,(1-\hat{q}_{\mathrm{A}}),\\ &= -0.1797 \times 0.1182 \times 0.8818,\\ &= -0.0187,\\ g_{\beta,\mathrm{A}} &= d^{*} g_{\alpha,\mathrm{A}},\\ &= -0.0187,\\ g_{\gamma,\mathrm{A}} &= -g_{\alpha,\mathrm{A}},\\ &= 0.0187,\end{aligned}$$

the other six likewise, and the delta method (102) over the block-diagonal covariance gives

$$\begin{aligned}\sigma_A &= \sqrt{g^{\top}\Sigma\, g},\\ &= \sqrt{2.654 \times 10^{-4}},\\ &= 0.0163,\end{aligned}$$

and the interval (105) is $[0.9354, 0.9890]$, against the reference's $[0.9244, 0.9816]$. The change the disturbance makes, the fifth arm's quantity, needs the ordinary month beside it: at $d^{*} = 0$ the same enumeration gives $\hat{A}(0) = 0.9826$, so by (101)

$$\begin{aligned}\Delta\hat{A}(1) &= \hat{A}(1) - \hat{A}(0),\\ &= 0.9622 - 0.9826,\\ &= -0.0204.\end{aligned}$$

Its gradient's unit-A entries by (103) are the differences of the two enumerations' entries, the ordinary month's sensitivity entry being zero,

$$\begin{aligned}g_{\Delta,\alpha,\mathrm{A}} &= g_{\alpha,\mathrm{A}}(1) - g_{\alpha,\mathrm{A}}(0),\\ &= -0.0187 - (-0.0034),\\ &= -0.0153,\\ g_{\Delta,\beta,\mathrm{A}} &= g_{\beta,\mathrm{A}}(1) - 0,\\ &= -0.0187,\\ g_{\Delta,\gamma,\mathrm{A}} &= -g_{\Delta,\alpha,\mathrm{A}},\\ &= 0.0153,\end{aligned}$$

the other six likewise, and the delta method (104) on the same covariance gives

$$\begin{aligned}\sigma_\Delta &= \sqrt{g_\Delta^\top \Sigma\, g_\Delta},\\ &= \sqrt{2.471 \times 10^{-4}},\\ &= 0.0157,\end{aligned}$$

smaller than $\sigma_A = 0.0163$ because the baselines' uncertainty cancels in the difference; the interval (106) is $[-0.0462, 0.0055]$ against the reference's $[-0.0551, 0.0001]$, and the probability of a loss by (106) is

$$\begin{aligned}P(\Delta A(1) < 0 \mid E) &= \Phi(-\Delta\hat{A}(1)/\sigma_\Delta),\\ &= \Phi(1.297),\\ &= 0.903.\end{aligned}$$

F Derivations of the three operations

The chain from (5) to (12) is three steps of calculus, given here so that nothing in Section 3.2 has to be taken on trust. Bayes' rule at the level of the state says that the posterior is the likelihood of the readings times the prior on the state, divided by the probability of the readings,

$$p(z \mid y) = \frac{p(y \mid z)\, p(z)}{p(y)}. \tag{128}$$

Each row is one Gaussian factor in its own residual: a reading row says the reading sits within σ_k of the quantity it reads, a prior row that an unknown sits within s_k of its expected value, a physics row that its imbalance is within the stated tolerance of zero. With $w_k = 1/\sigma_k^2$, one row's factor is

$$p_k = \Big(\frac{w_k}{2\pi}\Big)^{1/2} \exp\big(-\tfrac{1}{2}\, w_k\, r_k(z)^2\big), \tag{129}$$

and independent rows multiply, which is (5) with its normalizing factors written out; $p(y)$ does not depend on z and is dropped. Taking the negative logarithm turns the product into a sum,

$$-\log p(z \mid y) = \tfrac{1}{2}\sum_k w_k\, r_k(z)^2 - \tfrac{1}{2}\sum_k \log\frac{w_k}{2\pi} + \log p(y), \tag{130}$$

and only the first term moves when z moves: it is the cost (6), and the two constants are what (12) later restores. The logarithm is monotone, so the state that maximizes the posterior is the state that minimizes the cost, and at a minimum the derivative with respect to every unknown vanishes,

$$\frac{\partial C}{\partial z_i} = \sum_k w_k\, r_k(z)\, \frac{\partial r_k}{\partial z_i}, \tag{131}$$

$$J^\top W\, r(z) = 0, \tag{132}$$

the first being the derivative and the second, in matrix form, the condition that it vanish: the normal equations of weighted least squares. When every residual is linear in z they are solved in one step; when a residual is nonlinear, as the closure law is in e^θ, Gauss–Newton linearizes at the current guess, takes that step and repeats, which is (8), and its stopping test is literally that the

derivative is near zero. Expanding the cost to second order about the minimum, where its gradient is zero, leaves a quadratic whose curvature is the Hessian,

$$C(z) \approx C(z^*) + \tfrac{1}{2}\,(z - z^*)^\top \Lambda\,(z - z^*), \tag{133}$$

up to terms proportional to the residuals, which are small at a good fit; exponentiating gives the Gaussian (9) with the information matrix (10). Finally the evidence is the integral of the unnormalized posterior over every state the model allows, and a Gaussian integrates in closed form,

$$\int \exp\!\big(-\tfrac{1}{2}\,(z - z^*)^\top \Lambda\,(z - z^*)\big)\,dz = \frac{(2\pi)^{d/2}}{\sqrt{\det \Lambda}}, \tag{134}$$

whose logarithm, added to $-C(z^*)$ and to the row constants dropped above, is (12). The first term rewards fit; the volume terms charge a model for every unknown it had to pin down; the row constants make models with different numbers of rows comparable as densities. A difference of two log evidences is the logarithm of a ratio of probabilities, a pure number; a nat is the name of that number when the logarithm is natural, as a bit is when it is base two, and one nat is odds of e to one. On the scale of Kass and Raftery (1995), a difference below one nat is not worth more than a bare mention, one to three is positive, three to five strong, and above five very strong; the decision gap of one nat used in the Methods is the lower edge of that scale.

G Notation

The symbols used in the report are collected in Table 20.

References

[Rausand and Hoyland 2004] Rausand, M., Høyland, A. (2004). System Reliability Theory: Models, Statistical Methods, and Applications, 2nd ed. Wiley, Hoboken. ISBN 978-0-471-47133-2.

[Billinton and Allan 1992] Billinton, R., Allan, R. N. (1992). Reliability Evaluation of Engineering Systems: Concepts and Techniques, 2nd ed. Plenum Press, New York. doi:10.1007/978-1-4899-0685-4.

[Vesely et al. 1981] Vesely, W. E., Goldberg, F. F., Roberts, N. H., Haasl, D. F. (1981). Fault Tree Handbook. NUREG-0492, U.S. Nuclear Regulatory Commission, Washington. https://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr0492/.

[Mosleh et al. 1988] Mosleh, A., Fleming, K. N., Parry, G. W., Paula, H. M., Worledge, D. H., Rasmuson, D. M. (1988). Procedures for Treating Common Cause Failures in Safety and Reliability Studies, vol. 1. NUREG/CR-4780, U.S. Nuclear Regulatory Commission, Washington.

[IEEE 2007] IEEE (2007). IEEE Std 493-2007, Recommended Practice for the Design of Reliable Industrial and Commercial Power Systems. IEEE, New York. doi:10.1109/IEEESTD.2007.380668.

[IEEE 2013] IEEE (2013). IEEE Std 3006.7-2013, Recommended Practice for Determining the Reliability of 7x24 Continuous Power Systems in Industrial and Commercial Facilities. IEEE, New York. doi:10.1109/IEEESTD.2013.6493367.

[IEEE 2014] IEEE (2014). IEEE Std 3006.5-2014, Recommended Practice for the Use of Probability Methods for Conducting a Reliability Analysis of Industrial and Commercial Power Systems. IEEE, New York.

[IEEE 2018] IEEE (2018). IEEE Std 3006.8-2018, Recommended Practice for Analyzing Reliability Data for Equipment Used in Industrial and Commercial Power Systems. IEEE, New York.

[IEEE 2025] IEEE (2025). IEEE Std 3006.1-2025, Recommended Practice for Reliability Planning and Design of Industrial and Commercial Power Systems. IEEE, New York.

[ASHRAE 2021] ASHRAE Technical Committee 9.9 (2021). Thermal Guidelines for Data Processing Environments, 5th ed. ASHRAE, Peachtree Corners. ISBN 978-1-947192-89-4.

[Birnbaum 1969] Birnbaum, Z. W. (1969). On the importance of different components in a multicomponent system. In: Krishnaiah, P. R. (ed.), Multivariate Analysis II, Academic Press, New York, pp. 581–592.

[Kass and Raftery 1995] Kass, R. E., Raftery, A. E. (1995). Bayes factors. Journal of the American Statistical Association, 90(430), 773–795. doi:10.1080/01621459.1995.10476572.

[Tierney and Kadane 1986] Tierney, L., Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. Journal of the American Statistical Association, 81(393), 82–86. doi:10.1080/01621459.1986.10478240.

Table 20: Notation used in the report.

symbol	meaning	unit
T_A, T_B	hall temperatures	°C
y_A, y_B, b_A, b_B	the hall sensors' averaged readings and their offset biases	°C, K
ΔT_{allow}, ΔT_{obs}	the allowable hall difference; the observed one, $y_B - y_A$	K
Q_{hall}, Q	one hall's IT load in service; the test load per hall	W
G, ΔT_{full}	conductance of the opening; the rise of hall B over hall A with its unit out and the full load applied, Q_{hall}/G	W/K, K
ΔT_0^{OR}, ΔT_0^{AND}, $\hat{G}$, s	the full-load rise each drawing implies for its tree; the fitted conductance and the posterior spread of its entry θ	K, K, W/K, –
C	capacity of the opening, read as $G\,\Delta T_{\text{allow}}$	W
q, Q_A, P_A, P_B	heat on the opening, the surviving unit's duty, the IT loads	W
σ, N, σ_Q, σ_P	a temperature sensor's noise per sample, the samples per test, the duty and PDU meters' noise as a fraction of the reading	K, –, –, –
s_b, s_T, s_G, s_λ, s_μ	prior spreads: a sensor bias, the set point, the opening's entry, a log failure rate, a log repair rate	K, K, –, –, –
$\sigma_{\Delta T}$	the floor on the hall difference, the delta-method spread $g^\top \Sigma_0 g$ of $y_B - y_A$; $\sqrt{2(\sigma^2/N + s_b^2)}$ up to the meters' part	K
H_{AND}, H_{OR}	the two fault trees as hypotheses	–
H_1, H_0, B	two hypotheses in general and the Bayes factor between them	–
Θ_{claim}, α	the region of the parameter a claim asserts; the risk of a false acceptance	–
V_a, y_a	the value of a candidate measurement a, the readings it would return	–, –
a	in the record arm's question, a threshold on the service availability	–
Δ, σ_Δ, Δ^*	evidence for H_{AND} over H_{OR}, the predictive spread of one realization of it, the decision gap at which a verdict is read	nats
τ	a drawing's tolerance on its conductance	–
λ, μ, r	failure rate, repair rate, mean repair time	1/yr, 1/yr, h
$\hat{\lambda}$, $\hat{\mu}$, σ_θ	the rates at the posterior mode; the posterior standard deviation of the log failure rate	1/yr, 1/yr, –
k, T	failures and operating hours of a record	–, h
q_i, $q_i(d)$, A_{service}	unavailability of component i, the same at severity d, availability of the service	–
d, d_t, d^*	the disturbance severity index, of month t, and the severity a result is reported at	–
M, s_β, β	the months of the disturbance log; the spread of the prior on a sensitivity; a sensitivity in general	–, –, –
$\lambda_{i,0}$, β_i, α_i, γ_i	baseline failure rate, sensitivity to the disturbance, log baseline rate and log repair rate of component i	1/yr, –, –, –
k_{it}, T_{it}, m_i, D_i	failures and operating hours of component i in month t; its repairs and summed downtime	–, h, –, h
a_{it}, Λ_i, w_d, $\bar{A}$	the fitted count of a month; a component's information block; a severity's frequency in the schedule; the service availability averaged over it	–
$\Delta A(d)$, u, g_Δ, σ_Δ, $\Delta\bar{A}$	the change of the service availability at severity d from an ordinary month; a threshold on it; the change's gradient and spread; the year's change	–
$I_B(i)$	Birnbaum importance of component i	–
θ, Σ, H_a	the opening's entry $\log G$ in the trip-test state, and in the third arm the parameter vector; its posterior covariance; a candidate's added information	–
z, $r_k(z)$, w_k, J, W, Λ	the state vector, a row's residual, its weight, the Jacobian of the rows, the diagonal matrix of weights, the information matrix	–
$x \sim \mathcal{N}(m, s^2)$	x is normally distributed with mean m and standard deviation s	–
$P(A \mid y)$, $p(k \mid \theta)$	the probability of A given the readings y; the probability of the count k at the log-rate θ	–
$\propto$, $\int \cdots dz$	equal up to a constant that does not depend on z; the integral over every value of the state	–
Φ, $\mathbb{E}[\cdot]$, $\#\{\cdot\}$	the standard normal distribution function; the expected value; the count of cases meeting a condition	–
a, $a^\top$, $\det$, I	the gradient of A in the parameters, its row form, the determi-	–

[Lindley 1956] Lindley, D. V. (1956). On a measure of the information provided by an experiment. The Annals of Mathematical Statistics, 27(4), 986–1005. doi:10.1214/aoms/1177728069.

[Raiffa and Schlaifer 1961] Raiffa, H., Schlaifer, R. (1961). Applied Statistical Decision Theory. Division of Research, Graduate School of Business Administration, Harvard University, Boston.

[Gelman et al. 2013] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., Rubin, D. B. (2013). Bayesian Data Analysis, 3rd ed. Chapman and Hall/CRC, Boca Raton. doi:10.1201/b16018.