Detection of efficiency degradation in a staged water-cooled chiller plant from two windows of hourly power telemetry by a fitted loss scale

Ricardo Marquez, Ph.D.
rmarquez123@gmail.com

September 2026

Abstract

Cooling equipment degrades slowly, and the degradation is invisible in the quantity such equipment is metered on, because electrical power also moves with the weather and with the load the plant serves. This report watches a 200 kW water-cooled chiller plant — two positive-displacement units of 100 kW, each the least efficient machine the energy code allows for its size — over two windows of 240 hourly power readings, between which the machine has lost 21 % of its efficiency while the window itself has become milder and lighter. It is an estimation study: one parameter per window, a loss scale multiplying the three loss coefficients of the plant's thermodynamic law, is fitted to that window's readings by weighted least squares, and the difference between the two fits is tested. The mean metered power falls from 22.69 to 18.81 kW, −17.1 %, so a flat power dashboard reads an improvement; the fits read 0.99829 ± 0.00349 early and 1.30726 ± 0.00388 late against a placed 1.30861, a rise of 30.95 % at a statistic of +59.18, an efficiency 21.04 % lower at the late window's own conditions, and an extra draw of 3.887 ± 0.066 kW, $4,086 \pm 69$ a year. On the only measure the two share, the direction of the change, the metered power is worse than the fitted comparison: it reports the change with the wrong sign. The statistic measures meter noise alone. Fourteen healthy machines whose shape, evaporator behavior, unloading or unit count differs from the fitted law show apparent drifts up to 16.89 % at statistics past the detection threshold, so the claim this design supports is that the measured 30.95 % exceeds the largest such false drift by a factor of 1.83, not that it stands 59 standard errors from zero. Across 48 plant models no verdict turns, the margin falling from 0.922171 to 0.868766.

1 Introduction

A data center pays for electricity twice, once to run its computers and again to remove the heat they produce, and the second bill is usually dominated by a chiller plant. Masanet, Shehabi, Lei, Smith and Koomey (2020) put global data-center electricity use at about one percent of world demand and show that efficiency gains, cooling among them, have held that share nearly flat while computing has grown; Barroso, Hölzle and Ranganathan (2019) describe the facility overhead of warehouse-scale computing and the ratio by which it is reported. A chiller that fouls, loses refrigerant charge or develops a fault in its heat exchangers does not fail: it goes on delivering the cooling asked of it and simply draws more electricity to do so. The loss is therefore never announced, and it accumulates for as long as nobody looks. Looking is harder than it sounds, because the quantity the plant is metered on is not its efficiency but its power, and power moves with the weather and with the load.

A plant whose condition worsens through an autumn can draw less power in November than it did in August and read, on every dashboard in the building, as an improvement.

The detection of such faults has a long literature. Katipamula and Brambley (2005) survey the methods used for building systems and divide them into those that compare measurements against a physical model of the equipment and those that learn the equipment's normal behavior from data. On the physical side, the standard description of a vapor-compression chiller is the thermodynamic model of Gordon and Ng (1995), which derives from the first and second laws a relation between the compressor power, the cooling delivered and the two water temperatures, with three coefficients standing for internal entropy generation, an equivalent heat leak and the resistance of the heat exchangers; Chua, Ng and Gordon (1996) tested it against measurements on reciprocating machines and Foliaco, Bula and Coombes (2020) revisited it on a field-measured centrifugal chiller. Gordon and Ng already proposed reading a shift in those coefficients as a diagnosis, since each stands for a distinct physical loss. On the data side, Reddy, Niebur, Andersen, Pericolo and Cabrera (2003) compared the performance models available for on-line training and Reddy (2007) evaluated four fault detection methods against a common protocol, both using the instrumented 90-ton chiller of Comstock and Braun (1999), which remains the reference data set for the field; Zhao, Wang and Xiao (2013) built a statistical detector for centrifugal chillers from control charts on the residuals of a regression. None of these constructions is new here, and none is claimed as new.

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 the boundary and component laws that close them, attached to the same graph. Each of these is a relation the state of the network must satisfy, and the engine assembles them and solves them together. Domains enter through plugins that supply their laws, units and file formats; the solving and the bookkeeping are shared. This study uses its chiller-plant and data-center plugins to write the law of a staged plant, to solve it forward at every hour of two telemetry windows, and to recover one coefficient of that law per window from the metered power by least squares. Section 3 states the framework and each of these constructs precisely.

This report asks a narrow question and answers it with a stated bound on how far the answer may be trusted. The question is whether a change in a chiller plant's condition can be read from nothing but its hourly power, when the two windows being compared differ in load and in heat-rejection temperature as much as they differ in the machine. The construction is one number per window: the factor by which that window's machine is lossier than a reference machine of the same size and shape, fitted to the window's readings by weighted least squares, with the difference of the two fits divided by their combined standard error as the test. It is an estimation study, and the only uncertainty it reports is the one the fit yields. What it adds to the constructions above is the second half of the answer: the same detector is run on fourteen machines that did not degrade at all but whose shape or configuration differs from the law the fit assumes, and the drift it falsely reports on those is what bounds the claim. Section 2 states the plant, the windows and the model, Section 3 the estimator and the measures, Section 4 the results, Section 5 their limits and Section 6 the answers; the appendices hold the reproduction recipe and the provenance of the inputs.

2 System and data

The plant. Figure 1 draws the plant and the one quantity it is watched through. It is a water-cooled chiller plant of two identical positive-displacement units of 100 kW each, 200 kW in all, sized at 28.43 tons of refrigeration per unit, which places it in the smallest size class of the energy code. Each unit is the least efficient machine that meets both minimum efficiencies that class requires, the

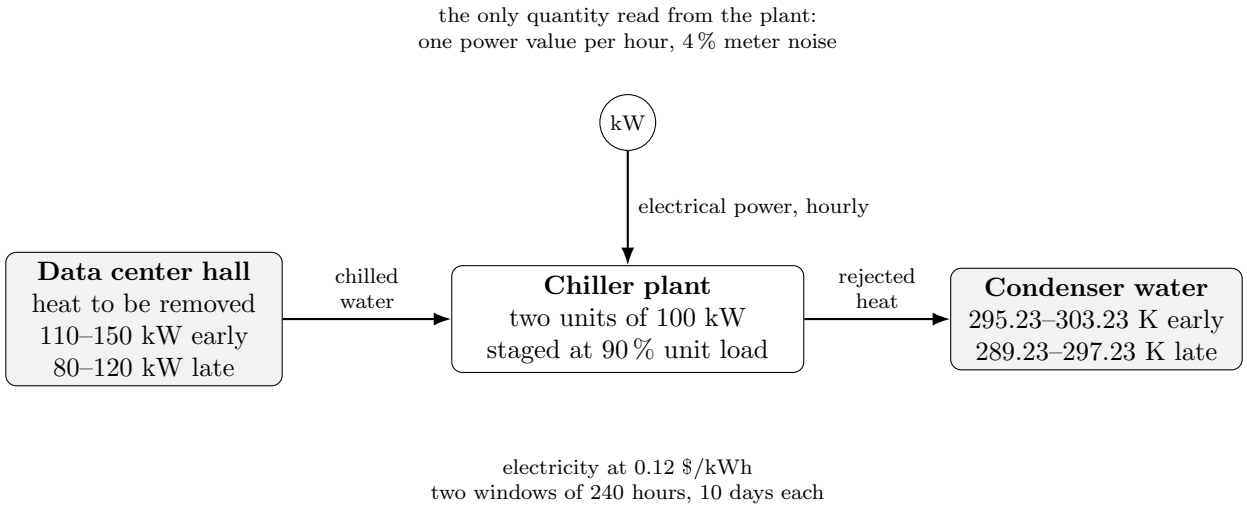


Figure 1: The plant of this report and the one quantity it is watched through. The hall hands the plant a heat load, the plant removes it into condenser water, and a meter records the plant's electrical power once an hour. The two windows differ in both of the conditions that set that power: the later window is lighter (50 against 65 percent of rated duty on average) and cooler on the condenser side by 6 K. Both changes push the metered power down, which is why the meter alone cannot say whether the machine itself has changed.

full-load and the integrated part-load value, rated at the standard conditions: its full-load coefficient of performance at the rating point is 5.577 and its integrated part-load value is 5.861. Chilled water enters the evaporator at 285.37 K and leaves at 279.82 K, and the entering temperature is held there whatever the load. Heat is rejected to condenser water, whose temperature entering the condenser is the second driving condition; head-pressure control would floor that temperature at 285.93 K, which is below the coldest hour of either window, so the floor never binds and the plant sees the condenser water as it is. A second unit is staged on when the running unit passes 90 % of its capacity.

The telemetry. Table 1 and Figure 3 state the two windows. Each is 240 consecutive hours, 10 days, of three numbers per hour: the heat the plant removed, the temperature of the condenser water entering it, and the plant's electrical power. Duty and temperature follow a diurnal cycle, peaking in the afternoon; the early window averages 130 kW of duty at 299.23 K and the late window 100 kW at 293.23 K, so the later window is lighter by 30 kW and cooler by 6 K on average. The power is the plant's true power for that hour multiplied by a random factor of unit mean and 4 % standard deviation, which is the meter noise; no other quantity is noisy and no hour is missing. The windows are not measurements of a real plant. They are generated from the plant model itself, the early window healthy and the late window degraded, which is what makes the true answer available to check the estimate against; the price of that choice is stated in Section 5. Their temperatures are placed so that the warmer window's maximum coincides with the upper bound of the range the plant law was calibrated over, 303.23 K, and the cooler window's minimum stays above that range's lower bound, 288.64 K, so that no hour of either window asks the law to extrapolate.

Table 1: The two telemetry windows. The later window is both lighter and cooler on the condenser side, and its mean metered power is lower by 17.1 % in consequence. Both windows lie inside the range of condenser-water temperature over which the plant law was calibrated, 288.64–303.23 K: the warmer window's maximum coincides with the upper bound of that range and the cooler window's minimum stays 0.59 K above the lower bound, so no hour of either window is priced outside the calibration.

quantity	early window	late window
hours	240	240
mean heat removed [kW]	130.0	100.0
heat removed, range [kW]	110.0–150.0	80.0–120.0
mean condenser water [K]	299.23	293.23
condenser water, range [K]	295.23–303.23	289.23–297.23
meter noise, one standard deviation	4 %	4 %
mean metered power [kW]	22.689	18.813

The degradation. Between the windows the machine loses 21 % of its mean efficiency at the late window's own conditions. That loss is imposed as one number: the factor multiplying all three loss coefficients of the unit's law together, which moves the machine's efficiency level while leaving the shape of its response to load and to temperature exactly as measured. The factor that produces the stated efficiency loss is 1.30861, an increase of 30.86 %; the design intended an increase between 20 % and 35 %, and it falls in that range. Nothing else changes: the units, the staging, the chilled water and the control are the same in both windows.

The model. Figure 2 shows what is known and what is not. At each hour the heat removed and the condenser-water temperature are given, and the plant's law turns them into a power, once the loss scale is fixed. The loss scale is the one unknown, one value per window, and the 240 power readings of the window are what determine it.

Assumptions.

- Units: temperatures in kelvins, heat and power in watts in the relations and in kilowatts in the text, money in United States dollars at a flat 0.12 $/kWh with no demand charge and no time-of-use structure.
- The hour is the resolution, every quantity is constant within it, and the plant is at steady state at each hour, so no heat and no compressor transient is carried from one hour to the next.
- The plant's electrical power is the compressor power of the running units alone. No chilled-water pump, condenser-water pump, cooling-tower fan or distribution fan is metered with it; that auxiliary share is passed to the fit as known and equal to 0, and no hour is served without refrigeration, so each window fits the loss scale alone.
- The meter noise is multiplicative, of unit mean, independent between hours and of known size, and it is the only source of error in the readings; the duty and the temperature are read exactly.
- The units are identical, share their load equally, and are priced by a law whose shape is calibrated on one laboratory machine and carried to this size by similarity.

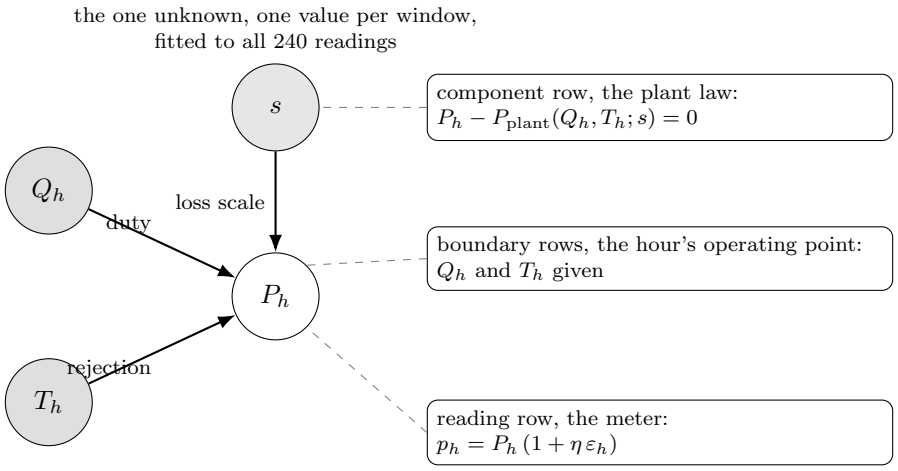


Figure 2: The model behind the estimate. Gray circles are the quantities given at each hour: the heat the plant removes and the temperature it rejects that heat to. The white circle is the plant's electrical power, which the component law determines from those two once the shaded circle, the loss scale, is fixed. The rows are of three kinds — the component law that prices the plant, the boundary values that fix the hour's operating point, and the reading row that ties the meter to the power. Solving the first three forward gives the power at any loss scale; the estimate runs the same map backward, choosing the one loss scale per window that makes the 240 predicted powers fit the 240 readings best.

- The degradation is a change of efficiency level only. A fault that changed the shape of the machine's response, rather than its level, is not what is placed here, and is what the bound of Section 4 measures.

The questions. Does the metered power say what happened to this plant? Can a fit of one parameter per window recover the change that did happen, and how large is the uncertainty of that recovery? How much drift would the same procedure report on a plant that did not degrade but is not the plant the law describes, and does the measured drift stand clear of it? What does the degradation cost, and how far do any of these answers depend on the choices made in modeling the plant?

3 Methods

3.1 The plant law

A chiller removing heat Q from chilled water entering its evaporator at T_{chi} and rejecting it to water entering its condenser at T_{ci} draws compressor power P, and its coefficient of performance is the ratio of the two. The thermodynamic model of Gordon and Ng (1995) relates them through three loss coefficients: b_1 in W/K, the internal entropy generation; b_2 in W, an equivalent heat leak; and b_3 in K/W, the total resistance of the two heat exchangers. With all three at zero the relation is the reversible limit. Solved for the power, the law (1) gives (2), and below a minimum stable load $Q_{\min}$ the machine cycles on and off, so its time-averaged power is the linear (3),

$$\frac{T_{\text{chi}}}{T_{\text{ci}}}\Big(1 + \frac{1}{\text{COP}}\Big) = 1 + \frac{b_1 T_{\text{chi}}}{Q} + \frac{b_2\,(T_{\text{ci}} - T_{\text{chi}})}{T_{\text{ci}}\,Q} + \frac{b_3\,Q}{T_{\text{ci}}}\Big(1 + \frac{1}{\text{COP}}\Big), \tag{1}$$

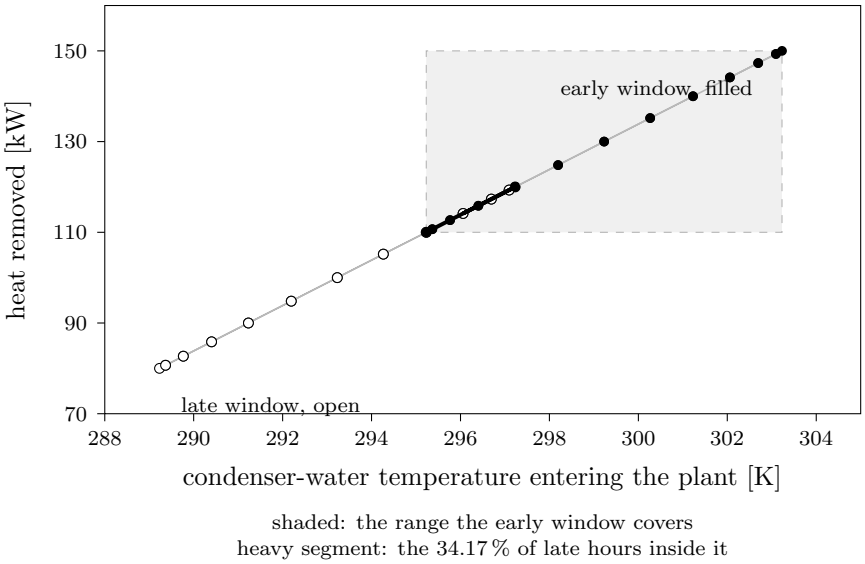


Figure 3: The two windows in the plane of the two conditions that set the plant's power. Each marker is one hour of the diurnal cycle, filled for the early window and open for the late one; duty and temperature rise and fall together through the day, so each window traces a segment rather than a cloud, and because the two swings are in the same proportion in both windows the two segments lie on one line. The shaded rectangle is the range of duty and temperature the early window covers. Only the heavy part of the late segment lies inside it, 34.17 % of the late window's hours, so a comparison of the two machines at the late window's own operating points is in the main an extrapolation of the early machine beyond the conditions it was watched under — although an extrapolation along the line the early window itself lies on, which is the mildest kind.

$$P_{\mathrm{GN}}(Q) = \frac{T_{\mathrm{ci}}\,Q + b_1 T_{\mathrm{chi}} T_{\mathrm{ci}} + b_2\,(T_{\mathrm{ci}} - T_{\mathrm{chi}})}{T_{\mathrm{chi}} - b_3\,Q} - Q, \tag{2}$$

$$P_{\mathrm{unit}}(Q) = \frac{Q}{Q_{\min}}\,P_{\mathrm{GN}}(Q_{\min}) \quad \text{for } Q < Q_{\min}. \tag{3}$$

The exchanger term makes the losses grow with load, so the efficiency rises from low load toward a maximum and falls again beyond it.

The shape of the law — the ratios among the three coefficients, which set how the power responds to load and to both water temperatures — is taken from a fit to all the running steady states of an instrumented 90-ton laboratory chiller (Comstock and Braun, 1999), and is carried to a unit of this capacity by similarity, the entropy generation and the heat leak growing with capacity and the exchanger resistance falling with it. That gives three reference coefficients β_1, β_2 and β_3 for a unit of 100 kW. One factor, the *loss scale* s, multiplies all three together and is the only thing that sets the machine's efficiency level,

$$b_1(s) = s\,\beta_1, \tag{4}$$

$$b_2(s) = s\,\beta_2, \tag{5}$$

$$b_3(s) = s\,\beta_3. \tag{6}$$

This is the coordinate the whole report works in. A machine at $s = 1$ is the reference machine; a machine at a larger s has the same response to load and temperature and is simply lossier everywhere, which is what a fouled exchanger, a worn compressor or a short charge produces. The

factor is anchored to the energy code: each of the two minimum efficiencies alone fixes the value of s at which the unit just meets it, and the least efficient unit meeting both takes the smaller. For this unit that anchor is 0.597426, and the machine it defines is the reference machine of (4)–(6), so that a fitted scale near one means a plant in the condition the code requires of a new one. The plant stages by threshold and shares the duty equally among the running units, with n units of capacity C and a stage-up fraction σ,

$$n_{\mathrm{run}} = \min\big(n,\ \lceil Q/(\sigma C)\rceil\big), \tag{7}$$

$$P_{\mathrm{plant}}(Q, T_{\mathrm{ci}}; s) = n_{\mathrm{run}}\, P_{\mathrm{unit}}\big(Q/n_{\mathrm{run}}\big). \tag{8}$$

with $n = 2$, $C = 100$ kW and $\sigma = 0.90$. Equations (7) and (8) are the forward map this report estimates through: given the hour's duty, the hour's condenser-water temperature and a loss scale, they return a power.

3.2 The network model

The plant law of the previous subsection is one instance of a general construction, in which the unknowns of a system and the relations they must satisfy are written on a graph and solved together; this subsection states it, because the estimator of the next one is a search over solutions of exactly this kind. 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, such as a temperature or a head, one flow per edge, and whatever else the physics determines. Every relation the state must satisfy — a conservation law at a node, a constitutive law on an edge, a boundary condition, a component law — is written as a residual $r_k(z)$ that vanishes when the relation holds. With as many independent relations as unknowns and no readings to reconcile, the state is the unique solution of

$$r(z) = 0, \tag{9}$$

found by Newton iteration on the sparse Jacobian J of the residuals, $J_{ki} = \partial r_k/\partial z_i$, from a stated start until the residual norm falls below a tolerance ϵ:

$$z \leftarrow z - J^{-1} r(z) \quad \text{until } \|r(z)\| < \epsilon. \tag{10}$$

A transient model adds a storage term to the conservation relations and advances the state in time with an implicit integrator, solving a system of the same kind at every step. The relations of this report are solved forward in that sense at every hour of both windows: the duty and the condenser-water temperature are given, and the plant's power follows. One quantity is not given, the loss scale of the machine, and it is recovered afterward from the metered power of each window by least squares on that same forward map, which the next subsection states on its own terms. Everything that follows is a choice of unknowns and relations on this graph.

3.3 The telemetry

Each window is 240 hours indexed by h, on a diurnal cycle peaking at the fourteenth hour. The duty and the condenser-water temperature swing with that cycle about their window means, and the reading is the true power of the hour multiplied by a noise factor,

$$\theta_h = 2\pi\,(h - 14)/24, \tag{11}$$

$$Q_h = \big(x_0 + x_1 \sin\theta_h\big)\, Q_{\mathrm{rated}}, \tag{12}$$

$$T_h = T_0 + T_1 \sin\theta_h, \tag{13}$$

$$p_h = P_{\text{plant}}(Q_h, T_h; s^\star)\,\big(1 + \eta\,\varepsilon_h\big), \tag{14}$$

$$\varepsilon_h \sim \mathcal{N}(0, 1). \tag{15}$$

with $Q_{\text{rated}} = 200$ kW and $\eta = 0.04$. The early window uses $x_0 = 0.65$ and $x_1 = 0.10$ with $T_0 = 299.23$ K and $T_1 = 4.0$ K, the late window $x_0 = 0.50$ and $x_1 = 0.10$ with $T_0 = 293.23$ K and $T_1 = 4.0$ K. The true loss scale $s^\star$ of (14) is one in the early window and 1.30861 in the late one, that being the value at which the plant's mean efficiency over the late window's conditions falls by 21 %. Because duty and temperature in (12) and (13) are driven by one phase (11), the hours of a window lie on a segment in the plane of the two conditions rather than filling a region of it, which is what Figure 3 shows and what makes the overlap between the windows small.

3.4 The estimator

One loss scale is fitted to each window on its own. The meter noise of (14) is proportional to the power, so each hour's misfit is divided by the power the model predicts there rather than by the reading, which would bias the fit downward; the estimate is the value of the loss scale minimizing the sum of those relative misfits, and its standard error follows from the curvature of that sum at the minimum,

$$r_h(s) = \frac{p_h - P_{\text{plant}}(Q_h, T_h; s)}{P_{\text{plant}}(Q_h, T_h; s)}, \tag{16}$$

$$\hat{s} = \arg\min_{s} \ \textstyle\sum_{h=1}^{H} r_h(s)^2, \tag{17}$$

$$J_h = \frac{1}{P_{\text{plant}}(Q_h, T_h; \hat{s})}\,\frac{\partial P_{\text{plant}}}{\partial s}(Q_h, T_h; \hat{s}), \tag{18}$$

$$\hat{\sigma}^2 = \frac{\sum_h r_h(\hat{s})^2}{\nu} \Big/ \sum_h J_h^2, \tag{19}$$

$$\nu = H - 1. \tag{20}$$

with $H = 240$ hours and $\nu = 239$ degrees of freedom, one parameter having been fitted. The minimization of (17) is carried out by Gauss–Newton iteration from $s = 1$, each step kept positive and below the load at which the denominator of (2) would vanish. The standard error (19) is the usual one of a nonlinear least-squares fit: it scales the inverse curvature by the scatter actually left in the residuals, so it answers the question of how far the fitted scale would move if the same plant were metered again, and no other question. The drift between the windows is that difference measured in units of the two standard errors combined, and the relative change of the fitted scales is reported beside it,

$$z = \frac{\hat{s}_{\text{late}} - \hat{s}_{\text{early}}}{\sqrt{\hat{\sigma}^2_{\text{early}} + \hat{\sigma}^2_{\text{late}}}}, \tag{21}$$

$$\delta = \hat{s}_{\text{late}} / \hat{s}_{\text{early}} - 1. \tag{22}$$

A positive z means the late machine is the lossier one. The verdict is read off $|z|$ against two fixed thresholds: at or above 5 the drift is called detected, at or above 2 borderline, and below that there is not enough evidence.

3.5 The consequences of the drift

Two quantities turn the fitted scales into something an operator can act on, and both are evaluated at the late window's own hours, since those are the conditions the plant is actually working at. The

first is the efficiency lost: the mean over the window of the hourly ratio of heat removed to power drawn, computed once with the early fit and once with the late fit, and their relative difference. The second is the waste: the mean extra power the late machine draws over what the early machine would have drawn at the same hours,

$$\overline{\mathrm{COP}}(s) = \frac{1}{H}\sum_{h}\frac{Q_h}{P_{\mathrm{plant}}(Q_h, T_h; s)}, \tag{23}$$

$$\lambda = 1 - \overline{\mathrm{COP}}(\hat{s}_{\mathrm{late}})/\overline{\mathrm{COP}}(\hat{s}_{\mathrm{early}}), \tag{24}$$

$$W = \frac{1}{H}\sum_{h}\Big(P_{\mathrm{plant}}(Q_h, T_h; \hat{s}_{\mathrm{late}}) - P_{\mathrm{plant}}(Q_h, T_h; \hat{s}_{\mathrm{early}})\Big). \tag{25}$$

The waste (25) carries a standard error obtained by propagating the two fits' standard errors through the same derivative (18), and it is multiplied by the hours in a week and in a year and by the tariff to give a cost. Both quantities compare the two *fitted* machines at the late window's operating points, and the early machine was never metered at most of those points, so the share of late hours lying inside the range of duty and temperature the early window covers is reported with them; below one, the comparison is partly an extrapolation.

3.6 The held-out check and the metered baseline

Two further quantities are computed, one to test the estimate and one to stand against it. The test is a held-out prediction: the machine fitted on the early window, which never saw the late window's readings, is asked to predict them, and the root mean square of its error is divided by that of the late window's own fit. The baseline is what a flat power dashboard reports, the relative change in mean metered power between the windows,

$$\rho = \frac{\mathrm{RMS}_h\big(p_h - P_{\mathrm{plant}}(Q_h, T_h; \hat{s}_{\mathrm{early}})\big)}{\mathrm{RMS}_h\big(p_h - P_{\mathrm{plant}}(Q_h, T_h; \hat{s}_{\mathrm{late}})\big)}, \tag{26}$$

$$D = \bar{p}_{\mathrm{late}}/\bar{p}_{\mathrm{early}} - 1. \tag{27}$$

A ratio ρ near one would say that the early machine explains the late readings as well as a machine fitted to them, which is what a plant that did not change would give. The baseline (27) uses only the metered power, exactly as a facility's own trend chart does, and is the comparison this report is written against.

3.7 The false drift of a healthy plant

The standard errors of (19) describe the meter noise and nothing else. They are silent about the error that matters more here, which is that the machine the telemetry came from may not be the machine (1)–(8) describes. That error is measured directly. Fourteen alternative plants are built, each differing from the fitted law in one respect: each of the three loss coefficients halved and doubled in turn, with the machine re-anchored to the code minimums around the changed ratio; a part-load efficiency curve that is a parabola in load and linear in temperature rather than the thermodynamic law; an evaporator whose sensitivity to chilled-water temperature is divided by 1.7 and by 3.1, which is the range by which the law's chilled-water sensitivity is known to exceed the laboratory data it was calibrated on; chilled water that falls with load, fitted as though it were held fixed; a unit unloading in three and in four equal compressor stages instead of running as one machine; and the same total capacity split into four units of 50 kW or gathered into one of 200 kW.

For each, both windows are generated from that plant with no degradation whatever between them, and both are fitted with the law of Section 3. Whatever drift comes back is false, and the largest of those false drifts is the floor below which this design cannot distinguish a degraded machine from a mismodeled one.

3.8 Sensitivity to the plant model

Separately from the false-drift bound, the verdicts are scored again under every plant choice a reader could make differently. This is a sensitivity analysis over modeling choices, not a statement about how uncertain the estimate is: each combination is a different plant, priced exactly, and the table says which of them keep the verdicts and which overturn them. Five choices are varied: loss scale, staging rule, unit count, heat-sink floor, chilled-water model. The loss scale of the plant itself is set to 0.8, 1, 1.2, so that the machine being watched is more or less lossy to begin with. The staging rule is the threshold rule (7) or least power, which runs the number of units drawing least among those that can carry the load. The unit count is 2 units of the stated capacity or twice as many of half of it, each re-anchored to the code minimums for its own size. The condenser floor is 55°F or 65°F of entering condenser water. The chilled-water model holds the entering water at its rating temperature or lets it fall with load under constant evaporator flow. Their product is 48 combinations. In each, the plant is rebuilt, each case's degradation is re-derived on that plant by the same efficiency-loss rule, and both windows are refitted, so nothing is carried across. The verdicts are scored by one clause per case, each written so that it is positive while that case keeps the verdict it has under the plant of this report, and the margin is the smallest of them,

$$c_k = \zeta_k - \zeta_{\mathrm{d}} \quad \text{(case } k \text{ reads detected)}, \tag{28}$$

$$c_k = \min\left(\zeta_k - \zeta_{\mathrm{b}},\ \zeta_{\mathrm{d}} - \zeta_k\right) \quad \text{(borderline)}, \tag{29}$$

$$c_k = \zeta_{\mathrm{b}} - \zeta_k \quad \text{(not enough evidence)}, \tag{30}$$

$$g = \min_k c_k. \tag{31}$$

with ζ_k the statistic of case k under the combination, $\zeta_{\mathrm{d}} = 5$ and $\zeta_{\mathrm{b}} = 2$ the two thresholds. The table reports the plant of this report, every combination differing from it in one choice, and the joint worst; ties would go to the combination with the lowest level indices in the order the choices are listed here.

3.9 Metrics

The recovery question is answered by the fitted loss scale (17) of each window, dimensionless, with the standard error (19) of that fit; its reference is the value the window was built with, one for the early window and 1.30861 for the late one, and the estimate is read against it in units of its own standard error, where a well-behaved fit should sit within about two. The detection question is answered by the statistic (21), in units of the combined standard error, read against the two thresholds 2 and 5 — and, more importantly, by the relative drift (22) in percent, whose reference is not zero but the largest false drift of Section 4, since a drift smaller than that is indistinguishable from an error in the plant model. The consequence questions are answered by the efficiency lost (24) in percent, whose reference is the 21 % placed, and by the waste (25) in kilowatts and dollars a year, whose reference is zero and which is also worth reading as a share of the late window's mean metered power. The plausibility of the fitted machines is measured by the held-out ratio (26), dimensionless, whose reference is one, the value a plant that did not change would give. The baseline is the metered change (27) in percent, read for its sign: it is right or wrong about the

direction of the change, and nothing else about it matters here. The robustness of the verdicts is measured by the margin (31), in units of the statistic, read against zero: its sign says whether all four verdicts survive a combination of plant choices and its size says by how much.

4 Results

4.1 What the meter says

The mean metered power is 22.689 kW in the early window and 18.813 kW in the late one, a change of −17.08 % by (27). Read as a trend, that is an improvement of roughly a sixth, and it is what every chart drawn from the power meter alone would show. It is wrong in sign. The machine in the late window is 21.04 % less efficient than the machine in the early one, and the fall in metered power is entirely accounted for by the two conditions that also changed: the late window asks for 30 kW less heat removal on average and rejects that heat to condenser water 6 K cooler. Both make a chiller draw less power, and together they more than cover the extra the degraded machine draws. This is the failure the rest of the section is written against, and it is not a defect of the meter: the meter reports the plant's power correctly, and the plant's power is simply not the plant's condition.

4.2 The fitted loss scales

Table 2 gives the estimate. The early window fits a loss scale of 0.99829 ± 0.00349, which is 0.49 standard errors from the healthy value of one; the late window fits 1.30726 ± 0.00388, which is 0.35 standard errors from the 1.30861 the window was built with. Both sit comfortably inside the two standard errors a well-behaved fit should give, so the estimator recovers what was there. The residual scatter of the fits, 953 W early and 808 W late, is of the order the 4 % meter noise implies at these power levels, so the one-parameter model leaves no structure behind in either window. The drift between the fits is 30.95 % of the early scale by (22), against the 30.86 % placed, and the statistic (21) is +59.18. Its sign says the late machine is the lossier one, and its size is far past the 5 at which the drift is called detected. At the late window's own conditions the two fitted machines have mean efficiencies of 6.949 and 5.487, a loss of 21.04 % by (24) against the 21 % placed — the estimate recovers not only the direction of the change but its size, to within a twentieth of itself.

4.3 What the statistic does not measure

The statistic of +59.18 should not be read as a confidence. Its denominator is the standard error of the two fits, and by (19) that quantity answers one question only: how far the fitted scale would move if the same plant were metered a second time. It says nothing about the possibility that the machine the telemetry came from is not the machine the law describes, and on this design that possibility is the larger one by an order of magnitude. Table 3 and Figure 4 measure it. Each row is a healthy plant — no degradation whatever between its two windows — whose shape, evaporator behavior, unloading or unit count differs from the fitted law, and the drift column is what the detector nevertheless reports, from noise-free traces so that only model error contributes; the statistic beside it comes from the same meter-noise draw as the primary case. The largest false drift is 16.89 % at a statistic of +13.4, from a plant whose part-load efficiency is a part-load parabola in place of the thermodynamic law; the next is 11.35 % at +10.6, from one unit of 200 kW in place of two of 100 kW; and the largest in the other direction is −4.50 % at −8.7, from a machine with its equivalent heat leak doubled. Every one of those is a perfectly healthy plant reading as

Table 2: The estimate on the primary case. The upper block is what each window's fit returns: a loss scale with the standard error of that fit, read against the value the window was built with, and the scatter of the fitted power about the readings. The lower block compares the two windows. The meter says the plant now draws 17.1 % less power; the fits say its losses are 30.95 % higher, that it converts 21.04 % less heat per unit of electricity at the late window's own conditions, and that the difference costs $4,086 a year at 0.12 $/kWh. The two standard errors are those of the fits and describe the meter noise alone; Table 3 bounds the error the plant law itself can contribute.

quantity	early window	late window
loss scale placed in the late window	—	1.30861
loss scale fitted	0.99829 ± 0.00349	1.30726 ± 0.00388
distance from the value placed [standard errors]	0.49	0.35
root-mean-square fit residual [W]	953	808
mean metered power [kW]	22.689	18.813
change in metered power		−17.08 %
change in the fitted loss scale		+30.95 %
detection statistic $\lvert z \rvert$		59.18
mean efficiency, late conditions		6.949 → 5.487, 21.04 % lower
extra power drawn [kW]		3.887 ± 0.066
extra cost [$/yr]		$4,086 \pm 69$

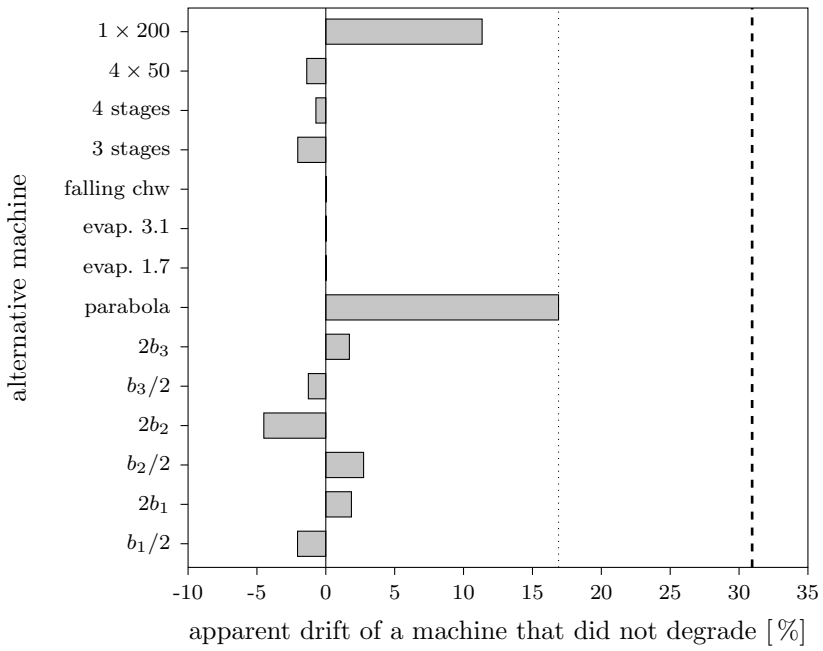
drift, and two of them read as drift at a statistic past the detection threshold. So the honest claim is not that the difference between the windows is 59 standard errors from zero. It is that the measured drift of 30.95 % exceeds the largest false drift this design admits by a factor of 1.83. The detection is real, and the margin is a factor of about two, not a factor of sixty. The rows that go the other way are worth naming, because they are the ones that support the result: the two rows that change the evaporator's sensitivity to chilled water by the 1.7 to 3.1 factor by which the law is known to exceed the laboratory data, and the row that lets chilled water fall with load while fitting it as fixed, move the answer by at most 0.05 %. Whatever else limits this detection, it is not how the chilled-water side is modeled.

4.4 The held-out prediction and the cost

The early machine, fitted on readings it will not be tested on, predicts the late window's power with a root mean square error of 4,031 W against the 808 W of the late window's own fit, a ratio of 4.99 by (26) where a plant that had not changed would give about one. The miss is not scatter: its mean is 3,876 W, so 96 % of the error is a single-signed under-prediction. The healthy machine consistently expects the late window to draw less power than it does, which is the degradation seen from the other side. The cost of doing nothing follows from (25). At the late window's own hours the fitted late machine draws 3.887 ± 0.066 kW more than the fitted early machine would have, which is 20.7 % of the late window's mean metered power and comes to 78.37 ± 1.33 a week and $4,086 \pm 69$ a year at 0.12 $/kWh. That figure is an extrapolation and is labeled as one: only 34.17 % of the late window's hours lie inside the range of duty and condenser-water temperature the early window covers, as the heavy segment of Figure 3 shows, so for the remaining 65.83 % the early machine is being evaluated at conditions it was never metered at. The extrapolation is a smooth one — the law is continuous in both conditions and the distance outside the box is small — but the number is a model's answer, not a measurement's, and it should be read as the order of the loss rather than as an invoice.

Table 3: What the detector reports when the plant is healthy but is not the plant the fit assumes. Each row generates both windows from the machine named, with no degradation between them, and fits them with the law of this report. The drift column is computed from noise-free traces, so it is model error alone; the statistic beside it comes from the same meter-noise draw as the primary case, so it is what this design would have reported. The apparent drifts run from −4.50 to 16.89 %, and the largest of them carries a statistic well past the detection threshold, which is the point of the table: a large statistic is not by itself evidence of degradation. The measured drift of 30.95 % exceeds the largest entry by a factor of 1.83. The three rows that change the evaporator side move the fit by at most 0.05 %, so the result does not rest on how chilled water is modeled.

machine the telemetry came from	what differs	drift [%]	z	waste [kW]
internal entropy generation halved	loss ratios	−2.06	−3.02	−0.266
internal entropy generation doubled	loss ratios	+1.86	+3.33	+0.226
equivalent heat leak halved	loss ratios	+2.75	+5.75	+0.339
equivalent heat leak doubled	loss ratios	−4.50	−8.70	−0.569
heat-exchanger resistance halved	loss ratios	−1.27	−2.00	−0.156
heat-exchanger resistance doubled	loss ratios	+1.71	+2.32	+0.219
a part-load parabola in place of the thermodynamic law	part-load shape	+16.89	+13.41	+2.207
evaporator sensitivity divided by 1.7	evaporator sensitivity	+0.02	+0.16	+0.003
evaporator sensitivity divided by 3.1	evaporator sensitivity	+0.03	+0.18	+0.004
chilled water falling with load, fitted as fixed	chilled-water model	+0.05	+0.22	+0.007
a unit unloading in three equal compressor stages	unit unloading	−2.04	−2.63	−0.277
a unit unloading in four equal compressor stages	unit unloading	−0.72	−0.95	−0.099
four units of 50 kW in place of two of 100 kW	unit count	−1.38	−1.89	−0.163
one unit of 200 kW in place of two of 100 kW	unit count	+11.35	+10.60	+1.419



each bar is a healthy machine fitted with the report's law, so the drift it appears to show is model error alone
dotted: the largest false drift, 16.89 %; dashed: the drift measured in the study, 30.95 %

Figure 4: What this detector reports for machines that did not degrade. Each bar is one alternative plant whose shape, evaporator behavior, unloading or unit count differs from the law the fit assumes; both of its windows are generated healthy, with no degradation at all, and the bar is the drift the fit nevertheless reports, from noise-free traces so that only model error contributes. The dotted line is the largest of those false drifts, 16.89 %, and the dashed line is the drift measured in the study, 30.95 %. The measured drift clears the largest false drift by a factor of 1.83, and that factor, not the number of standard errors the meter noise alone would support, is the strength of the detection.

Table 4: The four cases, from the primary one to the thinnest. The efficiency column is the mean efficiency the late window loses, and the two columns beside it are the loss scale placed in that window and the loss scale fitted to it. Each case shortens the window, raises the meter noise and shrinks the degradation, and each is one draw of the noise. The fitted scale tracks the value placed in every row, but the statistic falls by a factor of 46 from the first row to the last, and with it the verdict. The last row is a single draw; over 1,000 draws of the same case the statistic has a median of 2.922 and the most common verdict is borderline, which is the answer Table 5 scores.

case	hours	noise	efficiency	placed	fitted	z	verdict
primary	240	4 %	21 %	1.30861	1.30726 ± 0.00388	+59.18	detected
Easy	240	4 %	21 %	1.30861	1.30520 ± 0.00385	+58.62	detected
Moderate	96	6 %	10 %	1.12976	1.12333 ± 0.00815	+10.27	detected
Hard	48	8 %	5 %	1.06161	1.04315 ± 0.01370	+1.28	not enough evidence

4.5 Detection as the evidence thins

Table 4 and Figure 5 repeat the estimate on three further cases, each shorter, noisier and less degraded than the last. Panel (a) of the figure plots the statistic of each case on a logarithmic scale against the two thresholds; panel (b) is explained below. With the same 240 hours and 4 % noise but a different noise draw, the statistic is 58.62 and the drift is detected, as in the primary case. Halving the window to 96 hours, raising the noise to 6 % and shrinking the degradation to 10 % of efficiency brings the statistic to 10.27, still detected but by a factor of two rather than a factor of ten. At 48 hours, 8 % noise and 5 % of efficiency lost, the statistic is 1.28 and the verdict is that there is not enough evidence — although the fitted scale, 1.04315 ± 0.01370 against a placed 1.06161, is still in the right place. The estimator does not break on thin evidence; it reports that the evidence is thin. That last row is one draw of the meter noise, and on a case this thin one draw is not the answer. Repeating it over 1,000 draws, which is panel (b) of Figure 5, gives a median statistic of 2.922 and a mean of 2.954, with 79.2 % of the draws borderline, 17.9 % reporting not enough evidence and 2.9 % detected. The most common verdict is borderline, and the single draw of Table 4 fell in the 17.9 % that does not reach it. A case whose answer changes with the noise draw has to be reported as a distribution, and that is why the thinnest case is.

4.6 Sensitivity to the plant model

Table 5 and Figure 6 score the four verdicts again under every plant a reader could reasonably build instead. The figure plots the margin (31) of all 48 combinations, ordered by margin, against the zero line below which a verdict would have turned; the table gives the plant of this report, each single change, and the joint worst, with the primary and the thinnest case's statistics beside the margin. Under the plant of this report the margin is 0.922171, and it is the clause of the thinnest case: that case's statistic, 2.92217, sits 0.92217 above the borderline threshold, while the primary case's clause is 54.1799. The binding clause is the thinnest case staying borderline, which is as it should be: the fragile case is the one that decides whether a conclusion is robust, and a margin set by the primary case would measure nothing. No combination turns a verdict. The margin ranges from 0.868766 at the worst combination — the unit's three loss coefficients scaled by 0.8, least-power staging, chilled water falling with load — to 0.989336 at the best, and the single change that costs most is the unit's three loss coefficients scaled by 0.8. One point of reading needs stating, because the two tables disagree on the face of it. In Table 4 the thinnest case reads that there is not enough evidence; in Table 5 its clause is written against the borderline verdict. Both are right.

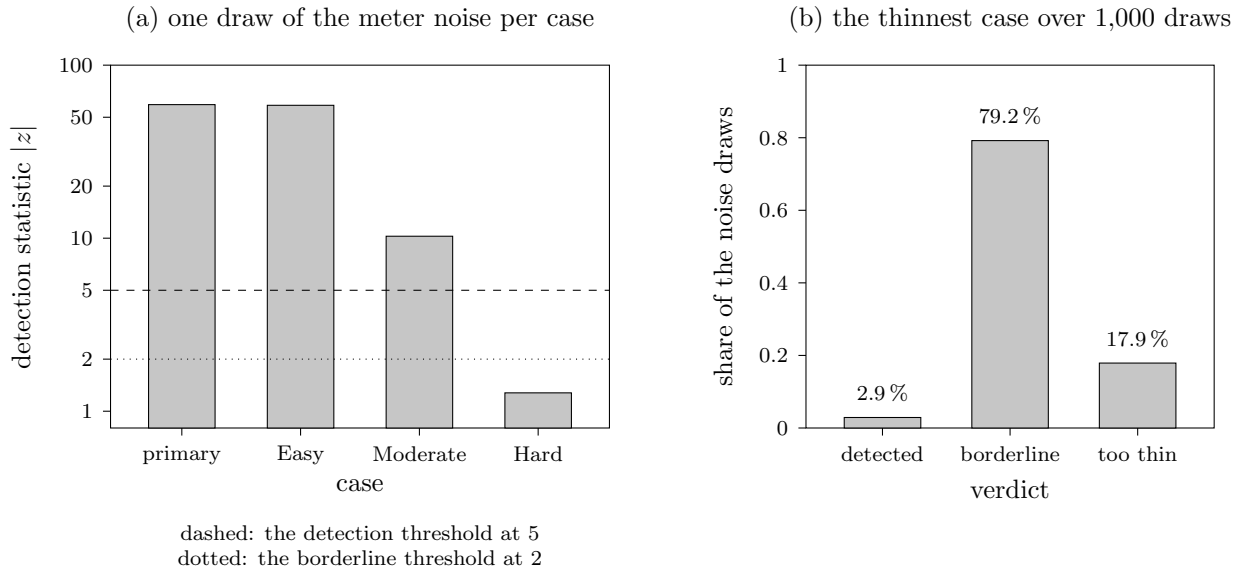


Figure 5: How the detection weakens as the evidence thins. (a) The statistic of the primary case and of the three further cases, on a logarithmic scale, with the two decision thresholds drawn; each bar is one draw of the meter noise. (b) The thinnest case repeated over 1,000 draws of that noise, as the share of draws falling in each verdict; the third label abbreviates the verdict that there is not enough evidence. The single draw of panel (a) reads 1.28 and reports too little evidence, while the median of the distribution in panel (b) is 2.92 and reports a borderline result; on a case this thin one draw is not the distribution, which is why the distribution is what the report gives.

The first is the answer at the single noise draw of that row, whose statistic is 1.28. The second is the answer at the median statistic of the 1,000 draws, 2.922, which is the value the survival table scores every plant on, so that combinations are compared with each other and not with each other's noise. A single draw is not the distribution, and on a case this thin the two fall on opposite sides of a threshold.

5 Discussion

Two findings of this report have different standing. That the metered power is the wrong quantity to watch is robust and does not depend on any of the modeling here: the plant's power fell 17.1 % while its efficiency fell 21.04 %, and the two conditions that moved between the windows are enough on their own to produce that. Any site whose load and weather drift together over a season is exposed to the same reversal. That a 30.95 % degradation was detected is also true, but the strength of the detection is the factor of 1.83 over the largest false drift of Table 3, not the 59 of the statistic. A reader who takes the statistic at face value will believe this method can resolve a one-percent change; the table says it cannot resolve anything below about 17 % on this design, because a healthy plant of the wrong shape already produces that much.

The largest limitation is that the telemetry is generated rather than measured. The windows come from the same plant model the fit assumes, with only the loss scale and the meter noise separating truth from fit, so the recovery of the placed scale to within half a standard error is a statement about the estimator and not about a real machine. That is exactly why the false-drift table exists: it is the part of the report that puts a plausible model error back in, and its

Table 5: The four verdicts under every plant choice a reader could make differently. The two middle columns are the statistic ζ of the primary case and of the thinnest one under that plant. The first row is the plant of this report, the middle rows change one choice at a time, and the last is the worst of all 48 combinations. Each row rebuilds the plant, re-derives each case's degradation on that plant and refits both windows, so nothing is carried across. The margin is the smallest of the four clauses and is positive while all four verdicts still hold; it is set in every row by the thinnest case, whose statistic sits closest to a threshold, and never by the primary case, whose statistic is more than ten times the detection threshold. No row turns a verdict, and the margin falls only from 0.922171 to 0.868766 at the worst combination. The thinnest case enters this table at the median statistic of its 1,000 noise draws, not at the single draw of Table 4.

plant choice changed	primary	thinnest	margin	verdicts
none (the plant of this report)	59.18	2.922	0.922171	holds
the unit's three loss coefficients scaled by 0.8	58.51	2.886	0.885668	holds
chilled water falling with load	59.19	2.920	0.919835	holds
a condenser floor at 65 °F	59.74	2.957	0.956643	holds
twice as many units of half the capacity	59.30	2.928	0.927510	holds
least-power staging	59.22	2.925	0.924814	holds
the unit's three loss coefficients scaled by 1.2	59.67	2.950	0.949836	holds
the unit's three loss coefficients scaled by 0.8, least-power staging, chilled water falling with load	58.30	2.869	0.868766	holds

answer is that model error dominates meter noise here by an order of magnitude — the largest false drift, 16.89 %, is 23 times the 0.74 % that the two fits' standard errors amount to. Even so, the fourteen alternatives are alternatives someone chose; a real machine may differ in a way none of them represents, and the bound is a floor on the error, not a ceiling.

The model stops in several further places. The plant law's shape comes from one instrumented laboratory chiller of 90 tons, carried to a 100 kW unit by a similarity rule that no second machine has tested, and the law's sensitivity to chilled-water temperature is known to exceed that machine's own data by a factor between 1.7 and 3.1. Here that matters little, because the chilled water is held at its rating temperature and the rows that vary it move the answer by at most 0.05 %, but it would matter in a study of chilled-water reset. Only the compressors are metered: pumps, tower fans and distribution fans are excluded, and since a real plant meter usually includes some of them, a real application would have to fit that share or measure it. The degradation placed is a pure change of efficiency level, which is what fouling and wear mostly produce; a fault that changed the machine's response to load — a failing unloader, a stuck valve — would show up in this estimator as a poor fit rather than as a clean shift, and the report does not test that case. The windows are short and adjacent in condition; a longer baseline covering more of the operating plane would shrink the extrapolation behind the waste figure, of which 65.83 % currently lies outside the early window's range. Finally, the economics are deliberately plain: a flat tariff, no demand charge, and no account of what the repair itself would cost or of how long the degradation took to accumulate.

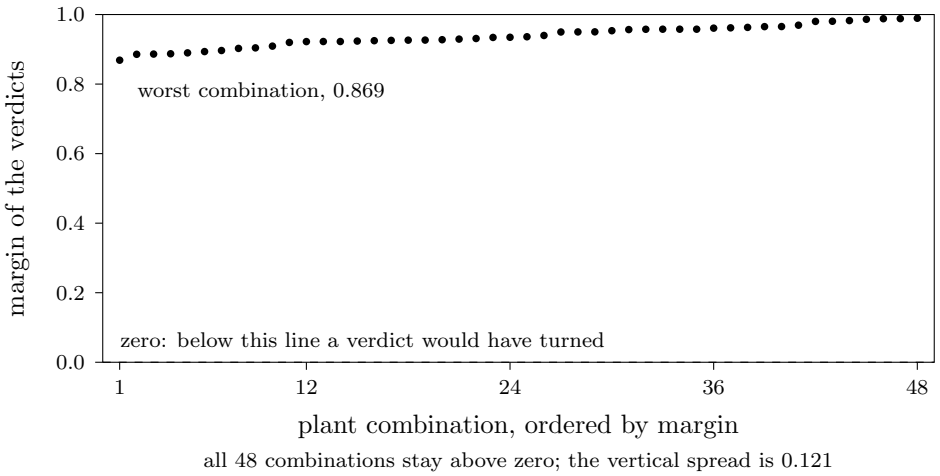


Figure 6: The margin of the four verdicts across every plant the report scores, one point per combination, ordered by margin. The margin is the smallest of the four clauses of the scoring rule, one clause per case, each written to be positive while that case keeps the verdict it has under the plant of this report. Every one of the 48 combinations stays above zero, from 0.868766 at the worst to 0.989336 at the best, so no verdict turns anywhere in the table. This is a sensitivity analysis over choices in the plant model, not a statement about how uncertain the estimate is.

6 Conclusion

This report asked whether the electrical power a chiller plant is metered on says what happened to the plant, and the answer is that it does not: across two windows in which the machine lost 21.04 % of its efficiency, the mean metered power fell from 22.69 to 18.81 kW, −17.1 %, because the later window was lighter by 30 kW and cooler by 6 K. It asked whether fitting one parameter per window recovers the change, and it does: the loss scale comes back as 0.99829 ± 0.00349 early and 1.30726 ± 0.00388 late against a placed 1.30861, within 0.49 standard errors in both windows, and the machine fitted on the early window mispredicts the late window's power by 4.99 times the late fit's own error, almost all of it a single-signed under-prediction. It asked what the degradation costs, and the two fitted machines differ by 3.887 ± 0.066 kW at the late window's hours, $4,086 \pm 69$ a year, a figure partly extrapolated because only 34.17 % of those hours lie inside the early window's operating range. It asked how much drift the same procedure reports on a plant that did not degrade, and the answer is the quantity that bounds everything else: up to 16.89 %, at statistics past the detection threshold, on fourteen healthy machines whose shape or configuration differs from the law the fit assumes. The measured 30.95 % therefore clears the largest false drift by a factor of 1.83, and that factor — not the 59 of the statistic, which measures the meter noise alone — is what the detection is worth. Finally it asked how far these answers depend on the plant model: over 48 combinations of five modeling choices no verdict turns, the margin falling from 0.922171 to 0.868766 at the worst, where it is set throughout by the thinnest of the four cases rather than by the primary one. The next step is to shorten the gap between what the statistic claims and what the method can deliver, by fitting the shape of the machine as well as its level so that a mismodeled part-load curve stops masquerading as drift, and to run the same estimator on metered telemetry from a plant whose maintenance history is known.

A Reproduction

```
cd docs/terra_graph_engine/case_studies/level2_room_hall/cop_drift
PYTHONPATH=../../../../../src python run.py
PYTHONPATH=../../../../../src python make_report.py --pdf
cd ../../../../..
python -m pytest -q tests/plugins/datacenter/test_cop_drift_case_study.py
```

`run.py` builds both windows, fits each, scores the four cases, repeats the thinnest over 1,000 noise draws, refits the fourteen alternative machines and re-scores the verdicts over the 48 plant combinations, and writes `results.json`; `make_report.py` reads only `results.json` and writes `report.tex`. Every random draw comes from a counter-based generator seeded per case: 0 for the primary case and the alternative machines, 100, 101, 102 for the three further cases, and 0 to 999 for the distribution of the thinnest. Each seed draws the early window's noise first and the late window's second, so a window's noise depends on the seed and on the window length alone. The whole run is deterministic: the same command writes the same bytes. The test in the last line regenerates the results, compares them with the committed file, and rebuilds the tex.

B Data and problems

No measured data enter this report. The telemetry is generated by `run.py` from the plant model of Section 3, and `results.json` is the only input the report reads. The plant law's three reference coefficients are a nonlinear least-squares fit, in compressor power, to all the running steady states of the 18 benchmark runs of the instrumented 90-ton chiller of Comstock and Braun (1999); the efficiency minimums the unit is anchored to are those of ASHRAE Standard 90.1-2019 for water-cooled positive-displacement chillers below 75 tons, rated at the conditions of AHRI Standard 550/590. The calibration covers condenser water from 288.64 to 303.23 K, which is the range both windows are placed inside. The study leaves two open problems. The first is the gap this report measures but does not close: the fit has one parameter, the machine's efficiency level, so any change of the machine's *shape* is forced into that one parameter and is reported as drift, which is the whole content of Table 3; a fit that identified the three coefficients separately would separate the two, at the cost of an estimator whose standard errors are much larger on windows this short. The second is that the alternative machines of that table were chosen by hand, and no argument is offered that they span the ways a real chiller can differ from the law.

References

[Masanet et al. 2020] Masanet, E., Shehabi, A., Lei, N., Smith, S., Koomey, J. (2020). Recalibrating global data center energy-use estimates. Science 367(6481), 984–986. doi:10.1126/science.aba3758.

[Barroso et al. 2019] Barroso, L. A., Hölzle, U., Ranganathan, P. (2019). The Datacenter as a Computer: Designing Warehouse-Scale Machines. Synthesis Lectures on Computer Architecture, Springer. doi:10.1007/978-3-031-01761-2.

[Katipamula and Brambley 2005] Katipamula, S., Brambley, M. R. (2005). Methods for fault detection, diagnostics, and prognostics for building systems — a review, Part I. HVAC&R Research 11(1), 3–25. doi:10.1080/10789669.2005.10391123.

[Comstock and Braun 1999] Comstock, M. C., Braun, J. E. (1999). Development of Analysis Tools for the Evaluation of Fault Detection and Diagnostics in Chillers (ASHRAE Research Project RP-1043). Report HL99-20, Ray W. Herrick Laboratories, Purdue University, West Lafayette, IN.

[Gordon and Ng 1995] Gordon, J. M., Ng, K. C. (1995). Predictive and diagnostic aspects of a universal thermodynamic model for chillers. International Journal of Heat and Mass Transfer 38(5), 807–818. doi:10.1016/0017-9310(94)00208-D.

[Chua et al. 1996] Chua, H. T., Ng, K. C., Gordon, J. M. (1996). Experimental study of the fundamental properties of reciprocating chillers and their relation to thermodynamic modeling and chiller design. International Journal of Heat and Mass Transfer 39(11), 2195–2204. doi:10.1016/0017-9310(95)00335-5.

[Reddy et al. 2003] Reddy, T. A., Niebur, D., Andersen, K. K., Pericolo, P. P., Cabrera, G. (2003). Evaluation of the suitability of different chiller performance models for on-line training applied to automated fault detection and diagnosis (RP-1139). HVAC&R Research 9(4), 385–414. doi:10.1080/10789669.2003.10391077.

[Reddy 2007] Reddy, T. A. (2007). Application of a generic evaluation methodology to assess four different chiller FDD methods (RP-1275). HVAC&R Research 13(5), 711–729. doi:10.1080/10789669.2007.10390982.

[Zhao et al. 2013] Zhao, Y., Wang, S., Xiao, F. (2013). A statistical fault detection and diagnosis method for centrifugal chillers based on exponentially-weighted moving average control charts and support vector regression. Applied Thermal Engineering 51(1–2), 560–572. doi:10.1016/j.applthermaleng.2012.09.030.

[Foliaco et al. 2020] Foliaco, B., Bula, A., Coombes, P. (2020). Improving the Gordon-Ng model and analyzing thermodynamic parameters to evaluate performance in a water-cooled centrifugal chiller. Energies 13(9), 2135. doi:10.3390/en13092135.

[ASHRAE 2019] ASHRAE (2019). ANSI/ASHRAE/IES Standard 90.1-2019: Energy Standard for Buildings Except Low-Rise Residential Buildings. ASHRAE, Atlanta, GA.

[AHRI 2023] AHRI (2023). AHRI Standard 550/590 (I-P): Performance Rating of Water-Chilling and Heat Pump Water-Heating Packages Using the Vapor Compression Cycle. Air-Conditioning, Heating, and Refrigeration Institute, Arlington, VA.