Calibrated Bayesian state estimation of a water distribution network from sparse flow meters, and what its graph decomposition separates

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

September 2026

Abstract

A water utility reads a few dozen meters and needs the pressures and flows everywhere else; inferring that state from sparse readings, with an honest statement of its uncertainty, is the base on which leak detection and pressure management stand. This study estimates the state of EPA Net3, the North Marin network of 92 junctions, 5 fixed-head nodes and 117 open links, from flow meters on 58 of 116 open pipes (standard deviation 1 kg/s) and one head gauge (1 cm) at the discharge of the running pump, with every junction demand uncertain to 30 % of its nominal value. The network is written as conservation residuals at the junctions and Hazen–Williams residuals on the pipes, and the maximum-a-posteriori state with its Laplace covariance is compared with EPANET 2.2 solutions of demand draws from the same prior. Over 20 draws, 1789 of 1840 junction heads lie within two posterior standard deviations of the truth and 1839 within three (97.2 % and 99.9 %, against 95.45 % and 99.73 % for a calibrated Gaussian); the head RMSE is 5.4 cm and the chi-square misfit is adequate in 20 of 20 draws. With the demands known, the MAP state equals a weighted least-squares solution by an independent Levenberg–Marquardt solver to round-off (below 10^{-9} m). The articulation-point decomposition of the network is reproduced — 31 articulation points, 35 biconnected components of which 32 are bridges — and shown not to separate the information matrix: eliminating the 31 articulation heads leaves 1 interior block, eliminating the 32 bridge flows with them leaves 18. In a break scenario — 12 of 116 pipes broken, 5 kg/s each, 8.8 % of the network's demand in all — a latent outflow at every junction ranked by its posterior z-score recovers 5 of 12 leak nodes in the top 12; ranking one-leak hypotheses by their normalized Laplace evidence recovers 5 in the top 12 and 7 in the top 24, with every false positive within 3 hops of a true leak. The estimator's head mean-square error (0.0029 m^2) is better than the 0.02 m^2 Han et al. report. The estimator's broken-pipe recall (42 % at $k = 12$) is worse than the 80 % Han et al. report. Their sensor set and damage model differ from this study's; both comparisons are on their metric, not on their data. Every number costs 119 engine solves.

1 Introduction

A water distribution network is operated from a few dozen readings: flow meters on some mains, pressure gauges at a few points, the levels of its tanks. Everything else a utility would like to know, the pressure at an unmetered junction, the flow in an unmetered pipe, how much a district is drawing, must be inferred, and the demands that drive the network are themselves known only from patterns. Getting that inference right matters for ordinary operation, where pressure management and leakage control (Puust, Kapelan, Savic and Koppel, 2010) depend on the state at nodes no one measures, and it matters most when a network is damaged and the readings that survive must say

where the water is going. An estimate is useful only with a credible uncertainty, and the honesty of that uncertainty against an independent simulation is something a method can be tested on.

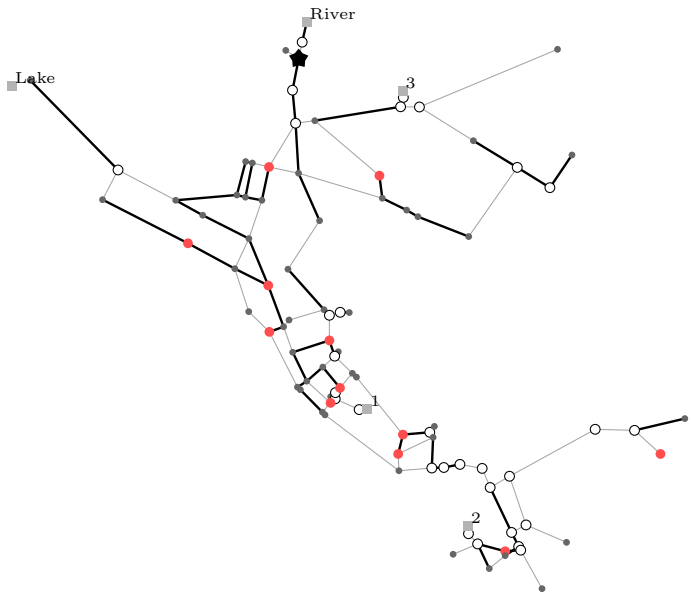
State estimation for networks follows the weighted least-squares tradition that power systems established: the physics as constraints trusted to a small tolerance, the readings as noisy observations, the unknowns fitted to both. Han, Eguchi, Mehrotra and Venkatasubramanian (2018) posed it for a damaged water network as inference on a factor graph, with Hazen–Williams factors between neighboring heads, mass-balance factors at junctions and measurement factors on flow meters and pressure sensors, solved by Gauss–Newton belief propagation after decomposing the network at its articulation points into biconnected components; on EPA Net3 they report a mean-square head error of 0.02 m^2 and 80 % of broken pipes localized under earthquake damage. Irofti, Romero-Ben, Stoican and Puig (2025) localize leaks with chained factor graphs over a time window on the Modena and L-Town benchmarks. Both build on the sparse least-squares machinery that Dellaert and Kaess (2017) describe for robotics, and both use EPANET (Rossman et al., 2020), the reference hydraulic solver, as the source of their truths, as this study does.

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. Domains enter through plugins that supply their laws, units and file formats; the inference and the bookkeeping are shared. This study uses its hydraulics plugin to write Net3's mass balances, pipe laws and pump curve, estimates the state from flow meters and one gauge, checks the honesty of the resulting uncertainties against an independent solver, examines the network's articulation decomposition in the information matrix, and ranks leak hypotheses in a pipe-break scenario. Section 3 states the framework and each of these constructs precisely.

This study takes Net3 in Han et al.'s state-estimation setting and asks four questions of one set of residuals. Whether the posterior uncertainties are honest: a factor-graph solver's variances are a device of the solver, so whether the Laplace covariance covers the truth at the Gaussian rates against an independent hydraulic solver is an open question, answered here over twenty demand draws. Whether the estimator is the familiar one: with no priors the MAP state should be the weighted least-squares solution, and that is verified against an independent Levenberg–Marquardt solver on the same residual vector. What the articulation-point decomposition separates when the unknowns are heads and link flows in the information form. And which junctions leak when a tenth of the pipes are broken, under two localization routes, with the recall reported as observed and the misses explained. No construction here is new; the value is in the checks, the failure modes they expose, and the four exercise-sized problems the study leaves in its corpus. Section 2 states the network, the data and the model, Section 3 the methods, Section 4 the results, Section 5 their limits, Section 6 the answers, and the appendices the reproduction recipe, the data and the references.

2 System and data

The network. Figure 1 draws EPA Net3 as shipped with EPANET 2.2 and WNTR 1.5.0: 92 junctions of which 58 carry a demand (nominal total 680 kg/s at the first pattern step), two reservoirs and three tanks held at fixed heads (River (67.1 m), Lake (50.9 m), 1 (44.2 m), 2 (42.7 m), 3 (48.2 m)), 117 pipes and two pumps. At the snapshot one pipe and one pump are closed by status and drop from the graph, leaving 117 open links on 97 nodes. A head is the water's energy



squares: reservoirs and tanks (fixed heads); open circles: the 31 articulation junctions; star: the head gauge (junction 61); thick links: the 58 metered pipes of the first draw; red: the 12 leak nodes of the break scenario

Figure 1: EPA Net3, the North Marin network shipped with EPANET 2.2, drawn from its file coordinates: 92 junctions, 5 fixed-head nodes (two reservoirs, three tanks) and 117 open links at the snapshot. Marked are the sensors of the study (flow meters on half of the open pipes, one head gauge at the running pump's discharge), the articulation junctions at which the graph decomposes, and the junctions that leak in the break scenario.

per unit weight at a node, in meters; a flow is the mass of water per second along a link; a demand is the outflow a junction delivers to its consumers.

The readings. Flow meters on 58 of the 116 open pipes (50 %, a fresh random set per draw), each with standard deviation 1 kg/s, and one head gauge of 1 cm at junction 61, the discharge of the running pump. Each junction demand is known only approximately, to 30 % of its nominal value with a floor of 0.5 kg/s. The truth of each draw is an EPANET 2.2 solution of the network with the demands drawn from that uncertainty; the readings are the truth plus their noise. Nothing here is field data: the network file is EPANET's public example and the truths are simulations.

The model. Figure 2 shows the equations around one node. The unknowns are a head at each of the 92 junctions, a signed mass flow on each of the 119 links, closed ones included (211 unknowns; the fixed heads are substituted constants and a closed link's flow is held at zero by its own relation), and one demand variable per demand-carrying junction when demands are uncertain. One conservation relation per junction, one pipe law per link and the running pump's curve determine them; the readings and the demand priors constrain them. Section 3 states them.

The questions. (1) Are the posterior standard deviations honest — does the truth fall within 2σ and 3σ at the Gaussian rates? (2) With the demands known, is the MAP state the weighted least-squares state? (3) What does the articulation-point decomposition separate in the information

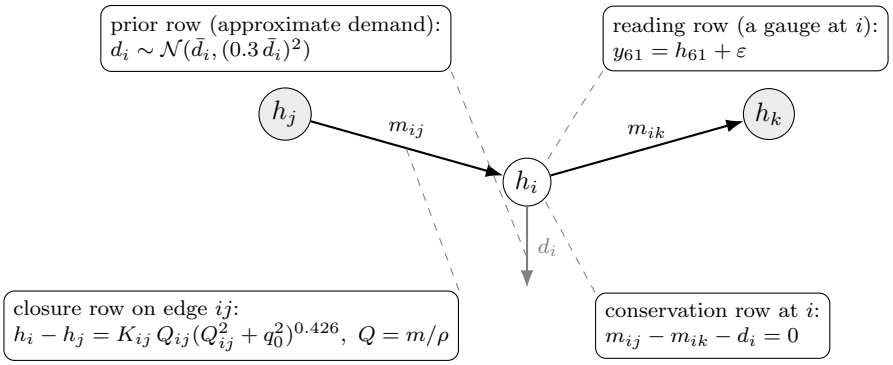


Figure 2: The model around one node. Unknowns are the head at the node, the mass flows on its links and its demand; the equations tying them are conservation at the node, a closure law on each link, a reading row per sensor and a prior row per approximately known quantity. Every row carries a weight; the fit minimizes the weighted squared misfits of all rows at once, and the uncertainty of each unknown follows from the curvature of that misfit. The same picture holds for every node and every domain; only the closure law changes.

matrix, and what does the elimination-cost predictor say? (4) With 12 pipes broken (10 % of the open pipes, a 5 kg/s outflow at a junction end of each), which junctions are leaking?

3 Methods

3.1 The Net3 factor graph

The nodes are junctions, reservoirs and tanks and the edges are pipes and pumps. Each junction carries a head unknown h_i (m), each link a mass-flow unknown m_e (kg/s) and each demand-carrying junction a demand unknown d_i (kg/s). Each physical law is written as a residual, an expression that vanishes when the law holds and carries a weight, the inverse variance of how far it may be off; Section 3.2 makes this the general rule. Conservation at junction i and the Hazen–Williams law on pipe $e = (i, j)$ give,

$$r_i = \sum_{e \to i} m_e - \sum_{i \to e} m_e - d_i, \tag{1}$$

$$r_e = h_i - h_j - K_e\, Q_e\, (Q_e^2 + q_0^2)^{0.426}, \tag{2}$$

$$K_e = \frac{10.667\, L_e}{C_e^{1.852} D_e^{4.871}}, \tag{3}$$

with $Q_e = m_e/\rho$ the volumetric flow in m^3/s, $\rho = 1000$ kg/m^3, K_e from the pipe's length L_e, diameter D_e and roughness C_e in SI units, and $q_0 = v_0 A_e$, $v_0 = 10^{-3}$ m/s, a small regularizing flow that keeps the law smooth through zero flow. The running pump follows its one-point EPANET curve, a head gain that falls quadratically with flow, with an implicit check valve frozen in its running branch from a forward solve; a closed link's residual holds its flow at zero. The physics residuals are weighted at $(1\ \mathrm{g/s})^{-2}$ on conservation and $(1\ \mathrm{mm})^{-2}$ on head, so the physics is trusted far more than any reading. The meters, the gauge and the demand priors attach to the same unknowns as residuals of the same form: a flow meter on link e, the head gauge at junction

61, and the prior that a demand stays close to its nominal value $\bar{d}_i$ each contribute one,

$$r = y_e - m_e, \tag{4}$$
$$r = y_{61} - h_{61}, \tag{5}$$
$$r = d_i - \bar{d}_i, \tag{6}$$

weighted by σ_m^{-2}, σ_h^{-2} and $\sigma_{\bar{d}}^{-2}$ respectively, with $\sigma_m = 1$ kg/s, $\sigma_h = 1$ cm, and $\sigma_{\bar{d}} = 30\,\%$ of the nominal demand with a floor of 0.5 kg/s.

3.2 The probabilistic model

The residuals of the previous subsection are all of one kind, and the engine treats them as such; this subsection states the general model they belong to. 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 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,

$$p(z \mid y) \propto \prod_k \exp\!\big(-\tfrac{1}{2}\, w_k\, r_k(z)^2\big). \tag{7}$$

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:

$$C(z) = \tfrac{1}{2} \sum_k w_k\, r_k(z)^2, \tag{8}$$
$$z^* = \arg\min_z C(z), \tag{9}$$
$$z \leftarrow z - (J^\top W J)^{-1} J^\top W\, r(z) \quad \text{until } \|J^\top W\, r(z)\| < \epsilon. \tag{10}$$

Near the minimum the cost (8) 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:

$$p(z \mid y) \approx \mathcal{N}\big(z^*, \Lambda^{-1}\big), \tag{11}$$
$$\Lambda = J^\top W J, \tag{12}$$
$$\sigma_i = \sqrt{(\Lambda^{-1})_{ii}}. \tag{13}$$

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,

$$\log Z = -C(z^*) + \tfrac{d}{2}\log 2\pi - \tfrac{1}{2}\log\det\Lambda + \sum_k \tfrac{1}{2}\log\frac{w_k}{2\pi}. \tag{14}$$

Differences of $\log Z$ between hypotheses, from (14), 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.

3.3 Testing the fit and its calibration

Every draw is solved to its MAP state (9) from a flat start (heads at the mean fixed head, 1 kg/s on every link) to a gradient norm of 10^{-10}, and the posterior standard deviations (13) are read one marginal at a time from a sparse factorization of Λ. Two checks follow. The misfit test asks whether the fitted model explains the readings and priors to within their stated noise,

$$\chi^2 = \sum_{k\in\text{readings, priors}} w_k\, r_k(z^*)^2, \tag{15}$$

compared with the chi-square distribution on as many degrees of freedom as readings and priors minus fitted unknowns; a fit whose residuals exceed what the noise allows is flagged as inadequate at $\alpha = 0.01$. The calibration test asks whether the posterior standard deviations are honest: for every junction i of every draw, whether the EPANET truth h_i^{true} lies within k posterior standard deviations of the estimate,

$$\left|h_i^{\text{true}} - h_i^*\right| \le k\,\sigma_i, \qquad k \in \{2, 3\}, \tag{16}$$

counted over all 1840 (draw, junction) pairs and read against the Gaussian rates 95.45 % and 99.73 %.

3.4 Weighted least squares by an independent solver

With the demands known there are no prior residuals, and the cost (8) is the classical weighted least-squares objective, the physics residuals scaled by $1/\sigma_{\text{phys}}$ and the meter residuals by $1/\sigma_m$. The same residual vector and Jacobian are handed to SciPy's Levenberg–Marquardt solver from the same flat start at tolerances 10^{-14}; the two minimize the same function and should agree to round-off, which is checked on the heads and the flows of the first draw and against the EPANET truth of that draw.

3.5 What the decomposition separates

Articulation points and biconnected components of the open-link graph are found by Tarjan's algorithm. Whether such a graph decomposition is also a decomposition of the inference is a question about the information matrix Λ (12): the marginal covariance of an interface set of unknowns B is the inverse of its Schur complement over the interior I,

$$\Sigma_{BB} = \left(\Lambda_{BB} - \Lambda_{BI}\Lambda_{II}^{-1}\Lambda_{IB}\right)^{-1}, \tag{17}$$

computed by factoring Λ_{II} block by block over the connected components of its sparsity graph; the number of those blocks is what the interface separates, and the marginal is exact whatever the interface. A dense elimination-cost predictor compares the whole with the parts,

$$\rho_{\text{cost}} = \frac{\sum_r n_r^3 + s^3}{n^3}, \tag{18}$$

with n the unknowns of the whole, n_r those of component r (its nodes plus its links) and s the interface size. Three interfaces are tried: the articulation heads alone, the heads with the bridge flows, and the heads with every flow incident to an articulation junction.

3.6 Break localization

In the break scenario 12 of the 116 open pipes are broken by adding a 5 kg/s outflow at a junction end of each, and two routes look for the leaking junctions from one set of readings. The continuous route adds a latent outflow s_i with a zero-mean prior of standard deviation 10 kg/s to every junction's conservation residual at once, solves one MAP, and ranks junctions by how far their outflow stands from zero in units of its posterior standard deviation,

$$z_i = \frac{|s_i^*|}{\sigma_{s_i}}. \tag{19}$$

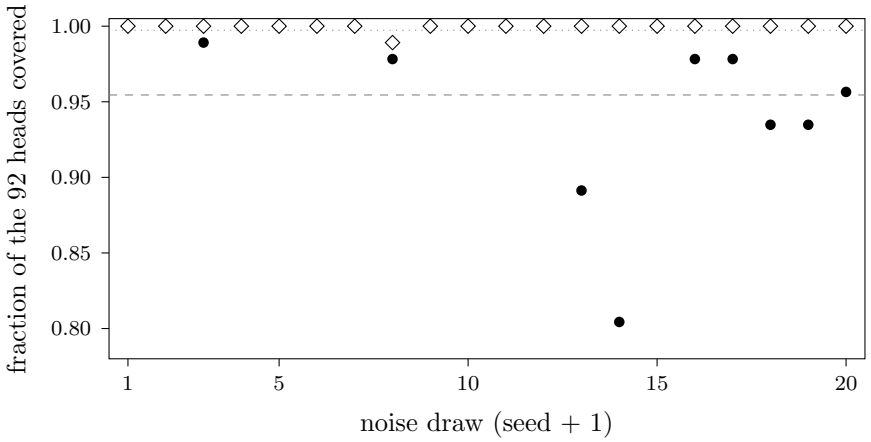
The hybrid route builds one hypothesis per junction, a single latent outflow there, plus the leak-free model, solves each to its MAP state and scores it by the normalized evidence (14), ranking by the gain over the leak-free model,

$$\Delta \log Z_j = \log Z_j - \log Z_0. \tag{20}$$

The engine's latent fault is an additive source on the node's conservation residual, so a leak is a negative fault; the report states outflows. Both routes keep the demand priors (6), so a leak competes with demand uncertainty at its node. Hypotheses whose evidence agrees to 10^{-6} nats are one hypothesis to the data and are ordered by name.

3.7 Metrics

The questions of Section 2 are answered with measures chosen so that each can be read against a reference. Calibration is the number of (draw, junction) pairs satisfying (16) at $k = 2$ and $k = 3$, read against the 1756 and 1835 of 1840 a calibrated Gaussian would give, together with the head RMSE against the EPANET truth, whose published reference is Han et al.'s mean-square error of 0.02 m^2, and the number of draws whose misfit (15) is adequate, read against the false-rejection rate $\alpha = 0.01$. The least-squares equivalence is the largest difference in head and in flow between the MAP and the Levenberg–Marquardt solutions, whose reference is round-off, stated as a bound. The decomposition is judged by the number of interior blocks each interface leaves, one meaning no separation, and by the cost ratio (18), below one meaning the parts are cheaper than the whole. Localization is judged by recall, the number of the 12 true leak nodes in each route's top 12 and top 24, read against the 80 % Han et al. report on their own damage scenario, and by the graph distance of every false positive from the nearest true leak, which says whether a miss lands in the right neighborhood. Cost is counted in engine solves rather than timed, so that it does not depend on the machine.



filled circles: within 2σ; open diamonds: within 3σ; dashed and dotted lines: the Gaussian expectations 0.9545 and 0.9973

Figure 3: Interval coverage per noise draw. In each of the 20 draws, the fraction of the 92 junction heads whose EPANET truth lies within two (circles) and three (diamonds) posterior standard deviations of the MAP estimate. Over all 1840 pairs, 1789 (97.2 %) are within 2σ and 1839 (99.9 %) within 3σ, against Gaussian expectations of 95.45 % and 99.73 %.

4 Results

Calibration. Figure 3 plots, for each of the 20 draws, the fraction of junction heads whose truth falls within two and three posterior standard deviations (16); Table 1 gives the counts with the head errors and the misfit. Over the 1840 (draw, junction) pairs, 1789 are within 2σ (97.2 %) and 1839 within 3σ (99.9 %), against 1756 and 1835 for a calibrated Gaussian. The excess at 2σ is on the conservative side, and its source is visible in the table: the draws with the lowest counts are those with the largest single errors, a head whose posterior width sits below the error at one poorly metered node while the rest of the network is over-covered. The head RMSE over all pairs is 5.4 cm (0.0029 m^2); the misfit (15) is adequate at $\alpha = 0.01$ in 20 of 20 draws, so the residuals are consistent with the stated noise. The estimator's head mean-square error (0.0029 m^2) is better than the 0.02 m^2 Han et al. report. Their estimate used a different sensor set and a damaged network, so this is a statement on their metric, not a replication of their number.

Weighted least squares. With the demands of the first draw known, the MAP heads and the Levenberg–Marquardt heads differ by less than 10^{-9} m and the flows by less than 10^{-9} kg/s — the same minimizer to round-off; the exact figures, which depend on the machine's BLAS kernel, are in the results file. Against the EPANET truth of that draw the RMSE is 9.5×10^{-5} m, the residual of the meter noise on a fully determined state. This is the sense in which the estimator is standard: the physics residuals are soft constraints with a very large weight, and everything else is weighted least squares.

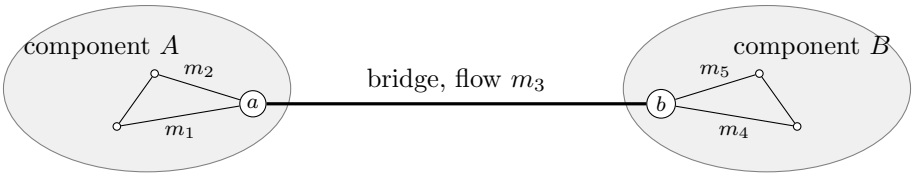
Decomposition. Net3's open-link graph has 31 articulation points (all junctions), 35 biconnected components of which 32 are single links, and one large component of 56 nodes; the next largest have 6 and 4 nodes. This is the structure Han et al. decompose on. Table 2 asks what it separates in the information matrix of the known-demand model. Removing the 31 articulation heads leaves 1 interior block: every component still talks to every other. Figure 4 shows why — the conservation residual (1) at an articulation node contains the flows of the links meeting there, so the bridge

Table 1: Per draw: heads within 2σ and 3σ (of 92), head RMSE and largest error against EPANET, mean posterior standard deviation, and the chi-square misfit of the fitted model on 59 degrees of freedom with its p-value ($\alpha = 0.01$; every draw is adequate). Totals: 1789 and 1839 of 1840; RMSE 5.4 cm over all pairs.

seed	2σ	3σ	RMSE [cm]	max [cm]	mean σ [cm]	χ^2	p
0	92	92	3.7	26.7	2.9	45.3	0.905
1	92	92	2.2	12.9	2.3	42.3	0.951
2	91	92	1.6	6.0	2.7	52.1	0.725
3	92	92	3.1	15.2	3.6	55.1	0.619
4	92	92	4.2	15.0	2.3	67.6	0.207
5	92	92	5.0	42.2	2.3	39.4	0.977
6	92	92	4.2	27.2	2.7	69.5	0.165
7	90	91	14.8	124.9	2.8	59.9	0.443
8	92	92	2.9	19.9	2.4	73.5	0.097
9	92	92	2.8	11.3	2.6	57.9	0.515
10	92	92	4.6	31.7	2.2	68.4	0.188
11	92	92	3.7	25.7	2.7	49.3	0.812
12	82	92	2.8	8.1	3.0	83.5	0.020
13	74	92	6.1	24.9	3.2	55.5	0.606
14	92	92	2.0	9.3	2.8	54.0	0.661
15	90	92	3.6	27.0	2.1	72.9	0.106
16	90	92	3.1	20.0	2.3	47.4	0.862
17	86	92	6.9	37.5	2.9	48.0	0.845
18	86	92	6.5	47.2	2.6	65.0	0.276
19	88	92	7.2	51.9	2.3	51.7	0.739

flow ties the two sides together after the head is gone. Adding the 32 bridge flows to the interface splits the interior into 18 blocks; adding all 64 flows incident to articulation junctions gives 19. In each case the block-Schur marginal (17) equals the full inverse to better than 10^{-10} m^2 (round-off; the exact figures are in the results file), so the interface marginal is exact whichever interface is chosen; what the interface decides is only whether the interior factorizes. The predictor (18) puts the decomposed cost at 0.24 of the monolithic cost for the head interface and 0.26 for the one that actually separates; the difference is the price of a separator of 63 rather than 31 unknowns. A node separator of the graph is a variable separator of the factor graph only when the factors at the separating nodes are split with it.

Pipe breaks. In the break scenario, 12 of the 116 open pipes, chosen at random, are broken by adding a 5 kg/s outflow at a junction end of each, 60 kg/s in all (8.8 % of the network's demand), with the demands of that draw unknown to the estimator as before and 58 flow meters. Figure 5 places each true leak node by its rank under the two routes, Table 3 lists the nodes with their demand priors and metering, and Table 4 gives the two top-k lists. The continuous route (19) finds 5 of 12 in its top 12 (and no more in its top 24); the hybrid route (20) finds 5 in its top 12 and 7 in its top 24. The leak-free model ranks 50th of the 93 hypotheses and the top hypothesis carries posterior probability 0.98, so the presence of leaks is not in doubt; their placement is. The estimator's broken-pipe recall (42 % at $k = 12$) is worse than the 80 % Han et al. report. Table 3 shows where the misses come from. A leak competes with the demand prior (6) at its node: of the 11 leak nodes whose demand prior is narrower than the leak, 8 are found by at least one route; of the 1 whose prior is as wide as the leak or wider, none is. A 5 kg/s outflow at a node whose demand



mass row at a: $m_1 + m_2 - m_3 - d_a = 0$ couples m_1, m_2 (in A) with m_3
closure of the bridge: $h_a - h_b - K_3 m_3 |m_3|^{0.852}/\rho^{1.852} = 0$ couples h_a, h_b, m_3
mass row at b: $m_3 - m_4 - m_5 - d_b = 0$ couples m_3 with m_4, m_5 (in B)

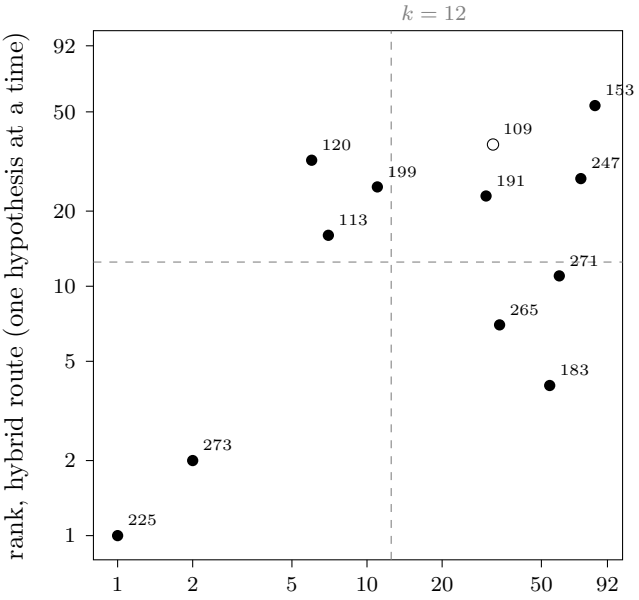
Figure 4: Why the articulation heads alone do not separate the information matrix. Two biconnected components A and B meet the bridge at the articulation nodes a and b. Removing a and b disconnects the graph, but the information matrix $\Lambda = J^\top W J$ couples two unknowns whenever a row contains both: the mass row at a couples the flows of A with the bridge flow m_3, and the mass row at b couples m_3 with the flows of B. Eliminating only h_a and h_b leaves m_3 in the interior, and through it one connected block; eliminating the bridge flows as well splits the interior. On Net3 the 31 articulation heads leave 1 block; adding the 32 bridge flows gives 18 (Table 2).

Table 2: Three interfaces on the known-demand model of Net3 (211 unknowns). Columns: interface size s; connected blocks of the interior information matrix after the interface is removed; and the dense elimination-cost predictor of equation (18) with the components' unknown counts from the graph decomposition. In every row the block-Schur marginal variance equals the same entry of the full inverse to better than 10^{-10} m^2 (round-off; exact figures in the results file).

interface	s	interior blocks	cost ratio
articulation heads	31	1	0.237
heads + bridge flows	63	18	0.261
heads + all incident flows	95	19	0.326

is uncertain to 5.9 kg/s is a plausible demand, and the estimator treats it as one. A metered link nearby matters as well — found by at least one route: 1 of the 3 leak nodes with no metered incident link, 7 of the 9 with one. The false positives are not scattered: every hybrid false positive is within 3 hops of a true leak (Table 4), the continuous route's within 5, so the evidence lands in the right neighborhood and the unmetered links inside it cannot say which node. The clearest case is the tie: junctions 163, 164, 166 have the same evidence gain to six digits — they lie on one unmetered path and are a single hypothesis, which no ranking can split. The continuous route buys its 1 solve against the hybrid route's 93 with a weaker ranking statistic: a z-score on a jointly estimated outflow, where 12 true outflows share their signal with 80 spurious ones.

Convergence of two hypotheses. The hybrid route's single-hypothesis solves take a median of 12 Gauss–Newton iterations (10), but the hypotheses at junctions 241 and 243 take more than 100 each (the exact counts, which move by a few iterations between BLAS kernels at the round-off floor of the gradient, are in the results file). The two lie on a dead-end pipe whose flow is essentially zero, where the regularized Hazen–Williams law (2) has its largest curvature and a hypothesized outflow is nearly collinear with the dead-end demand; Gauss–Newton's neglected second-order term then slows convergence from quadratic to linear, the gradient shrinking by about five percent per iteration. The solves converge to the same tolerance as the others; the iteration cap was raised to admit them.



filled: demand prior narrower than the 5 kg/s leak; open: wider (or the node carries no demand variable)

Figure 5: Where the 12 true leak nodes rank under the two localization routes (log axes; the dashed lines are the top-k cutoff). A node in the lower-left square is found by both routes. Filled markers are nodes whose demand prior is narrower than the leak. Table 3 lists the same nodes with their demand priors and metering.

Cost. 1 Newton solve, 118 MAP solves and 22 EPANET runs produce every number in this report (119 engine solves in all).

Table 3: The 12 leak nodes of the break scenario (5 kg/s each): nominal demand, the standard deviation of its prior (30 % of the nominal, floor 0.5 kg/s; zero where the node carries no demand and so no demand variable), metered incident links over degree, and the node's rank under each route with whether it is in the top $k = 12$.

node	demand [kg/s]	prior σ [kg/s]	metered	rank (cont.)	top-k	rank (hybrid)	top-k
225	1.9	0.58	1/1	1	yes	1	yes
273	0.0	0.00	2/3	2	yes	2	yes
183	0.0	0.00	0/3	54	no	4	yes
265	0.0	0.00	1/3	34	no	7	yes
271	0.0	0.00	1/3	59	no	11	yes
113	1.7	0.51	1/3	7	yes	16	no
191	6.9	2.08	1/3	30	no	23	no
199	10.1	3.03	2/3	11	yes	25	no
247	5.9	1.78	0/3	72	no	27	no
120	0.0	0.00	2/4	6	yes	32	no
109	19.6	5.87	1/2	32	no	37	no
153	3.7	1.12	0/2	82	no	53	no

Table 4: The top $k = 12$ of each route; true leak nodes are starred. Continuous route: $|z|$ of the latent outflow (19) and its MAP value. Hybrid route: log-evidence gain (20) over the leak-free model in nats and the MAP outflow of that hypothesis. "hops" is the graph distance from a false positive to the nearest true leak node.

rank	node	$\lvert z\rvert$	outflow [kg/s]	hops	node	gain [nats]	outflow [kg/s]	hops
1	225*	6.07	6.9		225*	20.1	7.3	
2	273*	3.99	7.9		273*	15.9	8.9	
3	123	2.17	2.4	2	179	13.5	18.3	1
4	40	1.90	2.6	2	183*	13.1	18.9	
5	231	1.50	1.7	5	177	12.9	17.3	2
6	120*	1.47	8.3		181	11.9	15.8	1
7	113*	1.34	6.1		265*	11.4	8.8	
8	103	1.21	8.3	1	163	11.2	9.1	1
9	117	1.16	5.1	1	164	11.2	9.1	2
10	237	1.10	-1.9	3	166	11.2	9.1	3
11	199*	1.10	6.2		271*	11.1	14.7	
12	131	1.08	1.6	4	161	11.0	9.1	2

5 Discussion

The calibration result holds for the noise model that produced the readings: Gaussian, independent, of known standard deviation. Field meters drift and fail, and the misfit test (15) is the report's only guard against that; it flagged nothing here because nothing was wrong. The demand prior of 30 % is both the source of the calibration's honesty and the reason the break scenario recovers only 5 of 12 leaks: a leak is a demand increase to the physics, and a prior that allows a 5 kg/s demand deviation cannot forbid a 5 kg/s leak. Tighter demand knowledge, more snapshots sharing one leak, or metered links around the candidate node are the three ways out, and the second is the construction a leak-localization study over a day of readings takes up. The comparison with Han et al. is on their metrics, not their data: their sensor set, damage model and network state differ, and the report claims no more than the direction of the two verdict sentences. The decomposition result is exact for the interface marginal and coarse for the cost predictor, which counts dense operations on a sparse problem; it says which interface separates, not how fast a real solver would run. The slow convergence of two hypotheses at a zero-flow dead end is a property of Gauss–Newton on the regularized Hazen–Williams law, not of the hypotheses; a Levenberg–Marquardt damping or a second-order correction would remove it. Everything here is a single steady snapshot of a public example network; extended-period behavior, tanks as storage and real telemetry are outside the study.

6 Conclusion

This report asked four questions of one set of conservation and Hazen–Williams residuals on EPA Net3, read from flow meters on half of its pipes and one head gauge with every demand uncertain to 30 %. Whether the posterior uncertainties are honest: over 1840 (draw, junction) pairs, 97.2 % of the heads lie within two posterior standard deviations of an independent hydraulic solver's truth and 99.9 % within three, against 95.45 % and 99.73 % for a calibrated Gaussian, with an adequate misfit in every draw; the uncertainty is honest to slightly conservative. Whether the estimator is the familiar one: with the demands known it coincides with weighted least squares to round-off. What the articulation-point decomposition separates: the graph, but not the information matrix, until the bridge flows join the interface; the interface marginal is exact either way, the factorization is not. Which junctions leak when 12 pipes are broken: 5 and 5 of 12 leak nodes in the top 12 under the two routes, 7 in the top 24, every false positive within 3 hops of a true leak, with the misses explained by demand priors wider than the leak and by unmetered neighborhoods, and one tie that no ranking can break. The recall is below the 80 % Han et al. report on their own scenario, and the reason is information rather than the solver: a single snapshot with 30 % demand uncertainty cannot tell a 5 kg/s leak from a plausible demand. A leak that is constant across a day while demands vary can be told apart, and that is the next study's construction.

A Reproduction

```
cd docs/terra_graph_engine/case_studies/water_networks/state_estimation_net3
PYTHONPATH=../../../../../src python run.py
PYTHONPATH=../../../../../src python make_report.py --pdf
```

`run.py` writes `results.json` (every number at six significant digits, plus the network's coordinates for Figure 1) and `timings.json`; `make_report.py` reads only `results.json`. Seeds 0–19

drive the demand draws, the meter sets and the reading noise; the break scenario uses seed 100. EPANET 2.2 is run through WNTR 1.5.0 on the Net3 file it ships; water density 1000 kg/m^3.

B Data and problems

Net3 is the North Marin example network distributed with EPANET 2.2 and WNTR 1.5.0 (public domain, US EPA); no field data are used. Under `problems/`, four directories: *articulation points of Net3* (the decomposition, its bridges and the elimination-cost predictor); *weighted least squares equals MAP* (SciPy against the engine on a three-pipe network); *posterior width against meter sets* (which single flow meter narrows a head the most); *the Schur marginal equals the full inverse* (the interface marginal over two heads two ways). Each is a statement, a script that solves it with the plugin, and its results file.

References

[Han et al. 2018] Han, Q., Eguchi, R. T., Mehrotra, S., Venkatasubramanian, N. (2018). Enabling state estimation for fault identification in water distribution systems under large disasters. In: 2018 IEEE 37th Symposium on Reliable Distributed Systems (SRDS), pp. 161–170. doi:10.1109/SRDS.2018.00027.

[Irofti et al. 2025] Irofti, P., Romero-Ben, L., Stoican, F., Puig, V. (2025). Factor graph optimization for leak localization in water distribution networks. arXiv:2509.10982. doi:10.48550/arXiv.2509.10982.

[Rossman et al. 2020] Rossman, L. A., Woo, H., Tryby, M., Shang, F., Janke, R., Haxton, T. (2020). EPANET 2.2 User Manual. U.S. Environmental Protection Agency, EPA/600/R-20/133.

[Puust et al. 2010] Puust, R., Kapelan, Z., Savic, D. A., Koppel, T. (2010). A review of methods for leakage management in pipe networks. Urban Water Journal 7(1), 25–45. doi:10.1080/15730621003610878.

[Dellaert and Kaess 2017] Dellaert, F., Kaess, M. (2017). Factor graphs for robot perception. Foundations and Trends in Robotics 6(1–2), 1–139. doi:10.1561/2300000043.