# A water network through a week: tanks, level-controlled pumps and demand patterns on EPA Net3, and what sparse readings tell about the tanks no one meters

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

September 2026

**Abstract**

A water network changes through the day: tanks fill and drain, pumps switch on tank levels and on the clock, and demands follow daily patterns. This study models EPA Net3 (92 junctions, three tanks, two pumps, 18 simple controls) through a week as a sequence of steady network solves joined by tank-storage relations, and then asks what sparse readings of one day tell an operator about the tank levels no one meters. It is a forward study checked against EPANET 2.2 and an inference study with a posterior over the day's tank levels and demand. When a control switches matters more than how the tank is integrated: applying the level controls only at whole hours moves Net3's tank levels by up to 0.58 m and its heads by 27 m against EPANET, with 10 of 507 link-hours in the wrong state, while replacing forward Euler by backward Euler changes the level by centimeters. With EPANET's own step rule the model reproduces EPANET's step sequence, every status and the tank levels to within 0.5 mm over the week. From a cold start, with both pump flows and two pump-discharge heads read hourly, tank 2's level reaches 5 cm in none of the 19 draws within the day; adding tank 1's level brings tank 2 to 5 cm after a median of 12 hours and tank 3 after 12; a level gauge on tank 2 knows it from hour 0, a pressure gauge at junction 50, beside the tank, from hour 2. Against EPANET at a 60 s step, 82.4 % of the data-informed level-hours lie within two posterior standard deviations (982 of 1192; floor 90 %) and 92.3 % with the pre-registered inflation of the tank rows. In the main case the coverage floor is not met. With the inflated tank rows the coverage floor is met. The study costs 1,271 steady solves, 3,151 MAP solves and 117 EPANET runs.

## 1 Introduction

Utilities run their networks through time. Tanks absorb the difference between what the pumps deliver and what the city draws, pumps switch on the levels of those tanks and on tariff clocks, and the demands follow daily patterns. Extended-period simulation, the hydraulic solution repeated through a day or a week with the tanks updated between solutions, is how engineers study that behavior; EPANET (Rossman et al., 2020) is the reference implementation, and its algorithm, a steady solve per hydraulic step with the tank levels advanced by the net inflow and the step shortened to the moment a control or a tank limit is reached, is the standard against which any other model of the same network is read. What an operator sees of that behavior is sparse: SCADA records pump flows, a few pressures and the levels of some tanks, and the level of an unmetered tank, or the demand of the hour, has to be inferred.

State estimation for water networks has mostly been posed on steady snapshots: Han, Eguchi, Mehrotra and Venkatasubramanian (2018) estimate heads from flow meters on a factor graph of

Table 1: Net3's tanks, cylinders without volume curves, in SI units. The operating center is the level around which the study draws initial levels (Section 3.5).

| tank | bottom [m] | diameter [m] | area [m$^2$] | initial level [m] | operating center [m] | range [m] |
|---|---|---|---|---|---|---|
| 1 | 40.20 | 25.9 | 527 | 3.99 | 3.99 | 0.03–9.78 |
| 2 | 35.51 | 15.2 | 182 | 7.16 | 7.16 | 1.98–12.28 |
| 3 | 39.32 | 50.0 | 1,962 | 8.84 | 7.95 | 1.22–10.82 |

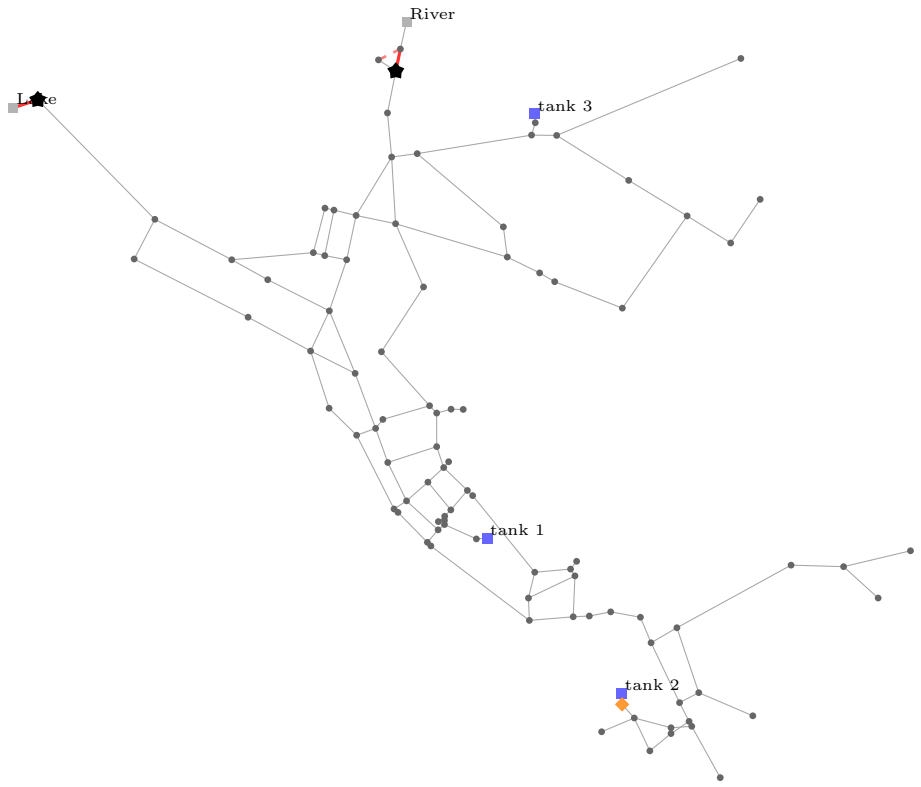
the network, and leak localization studies such as the Battle of the Leakage Detection and Isolation Methods (Vrachimis et al., 2022) read tank levels and pump flows as given inputs. Filters that carry a network's state through time exist: Romero-Ben, Irofti, Stoican and Puig (2026) fuse sensors with a dual unscented Kalman filter for leak localization. Leakage management reviews (Puust et al., 2010) stress that the value of an estimate depends on how well its uncertainty is known. Factor graphs over time windows, with the dynamics as factors between consecutive states, are standard in robotics (Dellaert and Kaess, 2017); none of the constructions used here is 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, 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 adds tank storage, time-varying demand and EPANET's simple controls to its hydraulics plugin, runs Net3 through a week against EPANET's own step log, and writes one day of it as a single factor graph over time from which the tank levels and the hourly demand are inferred with their uncertainty. Section 3 states the framework and each of these constructs precisely.

This report asks two questions. The forward one: can a model built from conservation relations carry Net3 through a week, and what decides whether it agrees with EPANET, the integrator or the timing of the switches? The inference one, put as an operator would: from a cold start, with the tank levels unknown, how many hours of readings until each tank's level is known to 5 cm; is a level gauge on tank 2 worth more or less than one pressure gauge at the best junction; and what do the switch times that SCADA logs anyway say about tank 1? Section 2 describes the network and the data, Section 3 the model and the protocol, Section 4 the answers, Section 5 their limits, and the appendices the reproduction recipe and the four problems the study leaves in its corpus.

## 2 System and data

**The network.** Figure 1 draws EPA Net3, the North Marin network shipped with EPANET 2.2 and WNTR 1.5.0: 92 junctions, two reservoirs (River and Lake), three tanks (Table 1), 117 pipes and two pumps, pump 10 drawing from Lake and pump 335 from River. A head is the water's energy per unit weight at a node, in meters; a tank's head is its bottom elevation plus its water level; a flow is the mass of water per second along a link; a demand is the outflow a junction delivers. The hydraulic time step and the demand pattern step are one hour and the horizon is 168 h. Eighteen simple controls run the pumps: fourteen switch pump 10 on at hours 1, 25, 49, ... and off at hours 15, 39, ...; four switch pump 335 on and the bypass pipe 330 closed when tank 1 falls below 5.21 m, and the reverse when it rises above 5.82 m. Nothing here is field data: the




gray squares: reservoirs; blue squares: tanks; red: pumps 10 and 335; dashed red: pipe 330 (the bypass switched with pump 335); stars: the head gauges at the pump discharge nodes 10 and 61; orange diamond: junction 50

Figure 1: EPA Net3 drawn from its file coordinates: 92 junctions (59 with a demand), two reservoirs, three tanks, 117 pipes and two pumps. Pump 10 runs on a clock; pump 335 and pipe 330 switch on the level of tank 1. The SCADA readings of the study are the two pump flows and the heads at the two starred nodes; tank 1's level is read in the main reading set; junction 50 is the single pressure gauge that says most about tank 2.

network is EPANET's public example and every truth is a simulation.

**The data of the inference.** The readings of the day are those a SCADA system would log hourly: the flows of the two pumps (standard deviation 5 kg/s, about 1 % of pump 335's flow), the heads at the pump discharge junctions 10 and 61 (10 cm), and in the main reading set the level of tank 1, the tank whose level drives the controls (1 cm). The times at which pumps and pipes switch are logged too, and the statuses are taken as known. The truths are EPANET 2.2 runs of the day with every junction demand multiplied by an hourly scale drawn around 1 and the tanks started from drawn levels; the readings are those truths plus drawn noise.

**The model.** Figure 2 shows the relations around one node at one step, Figure 3 how the steps of a day are joined into one model. The assumptions: SI units throughout, water density 1000 $\mathrm{kg/m^3}$; the Hazen–Williams law with its standard constant and a small regularizing flow; pumps on EPANET's three-point curves with an implicit check valve; tanks as cylinders; link statuses read from the SCADA log; the tank update of EPANET, forward Euler between steps; readings with

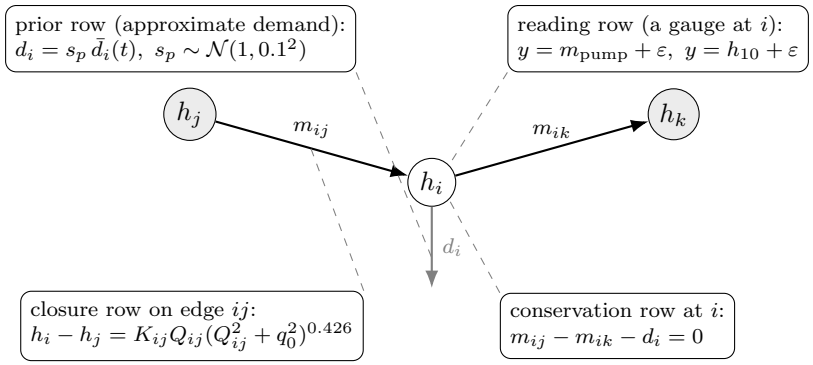



Figure 2: The model around one node at one hydraulic step. Unknowns are the heads, the link mass flows and the demands; conservation at the node, a closure law on each link, a reading row per sensor and a prior row per approximately known quantity tie them. The demands of a step are the nominal pattern demands times the demand scale $s_p$ of the pattern period. Figure 3 shows how the steps of a day are joined.

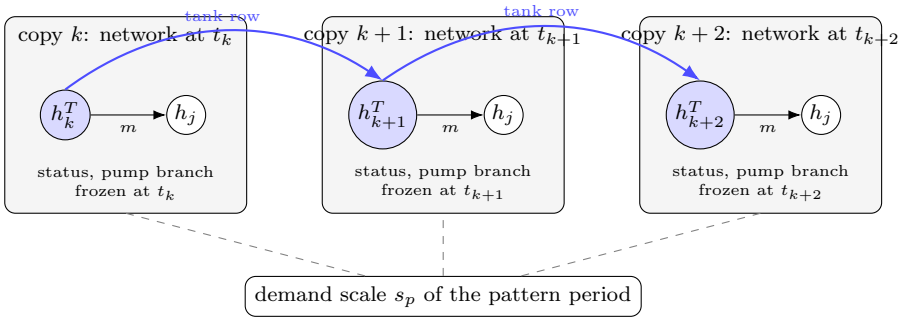



Figure 3: The unrolled window. Each step of the day is a copy of the steady network at that step's link statuses and pump branches (read from the SCADA log), with the tank heads free; a tank row ties each tank's head in one copy to the next through the net inflow of the first, and every junction demand of a copy is tied to the demand scale of its pattern period. Readings and priors attach to the variables of the copies.

independent Gaussian noise of known standard deviation; priors as stated in Section 3.5.

# 3 Methods

## 3.1 One hydraulic step

At a hydraulic time $t_k$ the tanks are fixed heads at their current levels, and the network is a steady problem. Each junction carries a head $h_i$, each link a mass flow $m_e$ (kg/s), and conservation at junction $i$, the Hazen–Williams law on a pipe $e = (i, j)$ and the pump curve on a running pump give

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

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

$$r_p = h_j - h_i - (h_0 - r\, Q_p^n), \tag{3}$$

with $Q = m/\rho$ in $\mathrm{m}^3/\mathrm{s}$, $K_e = 10.667\,L_e/(C_e^{1.852} D_e^{4.871})$ from the pipe's length, diameter and roughness, $q_0 = 10^{-3}$ m/s times the pipe's area, and $h_0, r, n$ the fitted pump curve; a closed link's residual holds its flow at zero. The demand of junction $i$ is its nominal demand times the multiplier of its pattern at the pattern period of $t_k$. The steady state is found by Newton's method with a loop over the pumps' check-valve branches.

### 3.2 Stepping in time

A tank of area $A$ changes level by its net inflow $q$, the sum of the flows of its links into it. Between two hydraulic times the tank update is forward Euler, EPANET's own,

$$A\left(h_{k+1}^T - h_k^T\right) = q_k\,\Delta t_k, \tag{4}$$

with $q_k$ from the solve at $t_k$. Before each solve the simple controls are checked: a time control fires when the clock equals its time; a level control fires when the tank's level is within one second of net inflow, $|q|/A$, of its threshold; controls are applied in file order, so a later control on the same link wins. The step $\Delta t_k$ is EPANET's: the hydraulic step, shortened to the next pattern period, the next report time, the next time control, the time a tank would fill or empty, and the time a tank would reach the level of a control that would change its link, each rounded to the second. Because the level is linear inside a step (4), the time to a threshold $h^*$ is closed form,

$$\Delta t^* = A\left(h^* - h_k^T\right)/q_k, \tag{5}$$

exact for agreement with EPANET, which computes it from the same start-of-step flows; as physics it carries an error of order $\Delta t^2(\partial q/\partial h)/A$, which lies inside the discretization error of the 1 h step itself. Two variants serve as comparisons: controls checked only at whole hours with no shortening (the counterfactual), and backward Euler (BDF1), in which each tank is a storage node with $\rho A\,h^T$ as its stored mass and the step is solved implicitly on the forward-Euler run's grid and statuses.

### 3.3 The day as one factor graph

For inference a day is written as one model (Figure 3). Every step $k$ of the day with $t_k < 24$ h is a copy of the steady network (1)–(3) at that step's link statuses and pump branches, with the tank heads free; the copies are joined by one tank row per tank and step, the forward-Euler relation (4) written in meters so that its scale does not depend on the step length; and each demand of copy $k$ is tied to the demand scale of its pattern period $p$,

$$r_k^T = h_{k+1}^T - h_k^T - \frac{\Delta t_k}{\rho A}\sum_e \pm m_{e,k}, \tag{6}$$

$$r_{i,k}^d = d_{i,k} - s_p\,\bar{d}_i(t_k), \tag{7}$$

with the sign + for a link into the tank. The step grid is the whole hours plus the switch times in the log, so the statuses are constant within each step. The unknowns are every head and flow of every copy, the tank heads at the end of the day, and the 24 demand scales; the parameters of interest are the three initial tank levels and the 24 scales, 27 in all. The readings attach to the copies at whole hours: the pump flows and discharge heads as $r = y - m$ and $r = y - h$, a tank level as $r = y - (h^T - z^T)$ with $z^T$ the tank's bottom. A logged switch of pump 335 or pipe 330 says that tank 1 stood at its control level at that time, a reading of tank 1 at the switch step. The priors are on the initial levels and the scales.

**Background: the model, the fit, and its uncertainty.** The quantities to be determined are the heads at the junctions and tanks and the mass flows on the links of every step of the day, the demand scale of every hour, and the initial levels of the tanks. Physics ties them together and the sensors constrain them. At every node a conserved quantity balances: the flows into a junction equal the flows out plus its demand, and a tank's level changes by its net inflow over the step. Along every edge a closure law relates the flow to the potential difference across it: the Hazen–Williams head loss on a pipe and the head gain of a running pump. Every sensor adds one relation between its reading and the quantity it observes (pump flows, discharge heads, a tank level, the level at a logged switch), and every quantity known approximately in advance adds a prior, a relation between the unknown and its expected value (the initial tank levels and the hourly demand scales). Each relation is written as a residual that vanishes when it holds, and together the residuals form one sparse system; this report calls them rows. No row is enforced exactly. Each carries a weight, the inverse variance of how far it may be off: a physics row is trusted to a small tolerance, a reading to its sensor's noise, a prior to its stated spread. The *maximum-a-posteriori* (MAP) state is the set of unknowns that minimizes the weighted sum of squared row misfits, $C(z) = \frac{1}{2}\sum_k w_k\, r_k(z)^2$; with only physics and readings it is the weighted least-squares fit, and it is found by Gauss–Newton iteration. The *posterior standard deviation* of an unknown is its uncertainty bar: near the minimum the misfit is approximately quadratic, so the state is approximately Gaussian with information matrix $\Lambda = J^\top W J$ (the weighted Jacobian of the rows), and the standard deviation of an unknown is the square root of the corresponding diagonal entry of $\Lambda^{-1}$. Treating the state as Gaussian around the MAP is the *Laplace approximation*. When several models compete — a leak here, a leak there, no leak — each is compared by its *evidence*, the probability of the readings under the model once its extra freedom is paid for: in the Laplace approximation

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

with $d$ the number of unknowns and the last sum over every row so that models of different size compare as densities. Differences of $\log Z$ are in *nats* (natural-log units of odds; 5 nats is odds of about 150 to 1), and the models' posterior probabilities are their evidences normalized to sum to one under equal prior odds.

### 3.4 The probabilistic model

The rows above 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 $y$,

$$p(z \mid y) \propto \prod_k \exp\!\big(-\tfrac{1}{2}\, w_k\, r_k(z)^2\big). \tag{8}$$

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 $\epsilon$:

$$C(z) = \tfrac{1}{2} \sum_k w_k \, r_k(z)^2, \tag{9}$$

$$z^* = \arg\min_z C(z), \tag{10}$$

$$z \leftarrow z - (J^\top W J)^{-1} J^\top W \, r(z) \quad \text{until } \|J^\top W \, r(z)\| < \epsilon. \tag{11}$$

Near the minimum the cost (9) is approximately quadratic, so the posterior is approximately Gaussian about $z^*$ with information matrix $\Lambda$. This is the *Laplace approximation*, and the *posterior standard deviation* $\sigma_i$ of any unknown is read from the diagonal of $\Lambda^{-1}$, from the same sparse factorization the last Gauss–Newton step used:

$$p(z \mid y) \approx \mathcal{N}\big(z^*, \Lambda^{-1}\big), \tag{12}$$

$$\Lambda = J^\top W J, \tag{13}$$

$$\sigma_i = \sqrt{(\Lambda^{-1})_{ii}}. \tag{14}$$

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,

$$\log Z = -C(z^*) + \tfrac{d}{2} \log 2\pi - \tfrac{1}{2} \log\det \Lambda + \sum_k \tfrac{1}{2} \log \frac{w_k}{2\pi}. \tag{15}$$

Differences of $\log Z$ between hypotheses, from (15), 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.

### 3.5 Protocol

Every setting was fixed before the first inference run. The truth of draw $n$ (seeds 0–19) multiplies every junction demand by hourly scales drawn from $\mathcal{N}(1, 0.1^2)$ and starts each tank at its operating center plus a draw from $\mathcal{N}(0, 0.3^2)$ m, truncated to 0.5 m inside the tank's range; the priors of the model are the same distributions. The operating center is the tank's initial level lowered, where needed, until its highest level on EPANET's nominal day stays 1 m below full: tank 3, which comes within 10.7 cm of full on that day, is centered 0.893 m below its file level; tanks 1 and 2 keep theirs. A draw whose truth fills or empties a tank is reported and left out, not replaced. The readings add Gaussian noise of the standard deviations of Section 2; a switch-time reading has 1.2 cm, a minute of tank 1's motion. The physics rows are weighted as $10^{-3}$ in their units (1 g/s of mass, 1 mm of head or level).

Five truths are compared. The *main case* is EPANET run at a 60 s hydraulic step, so that the truth's time discretization differs from the model's hourly one. Its pre-registered variant widens each tank row (6) to the discretization error it has to absorb: on five calibration draws (seeds

100–104, never reported as results) the row is evaluated on the 60 s truth at the model's grid, and the root mean square of its value in kg/s of net inflow gives each tank's standard deviation, 2.09, 1.38 and 2.93 kg/s for tanks 1, 2 and 3, converted to each row's meters by $\Delta t_k/(\rho A)$. Two harder cases change the truth's demand in ways the model does not represent: a persistent bias, every hourly scale times 1.1; and pattern 1, the default pattern of most junctions, shifted 2 h later, for which the truth's scale of hour $k$ is defined as the demand-weighted ratio $\sum_i d_i^{\mathrm{truth}}(k)/\sum_i \bar{d}_i(k)$. Their expected outcomes were written down in advance: for the bias, that the scales absorb it, their posterior means within 0.05 of the biased truth on average, and that the level coverage stays within five percentage points of the main case's; for the shifted pattern, that the level coverage falls below the main case's, because the scales must depart from 1 by more than their prior allows. The fifth truth is the model itself at a 1 h step, a check of the machinery rather than a result.

The filter at hour $k$ solves the window $[0, (k+1)\,\mathrm{h}]$ with the readings up to hour $k$; the physics of the last hour adds no information about the past, so the level at $k$ is estimated from the readings a causal estimator would have. The smoother solves the whole day. The best pressure gauge for tank 2 is chosen once, on the nominal day, by adding each junction's hourly head reading in turn to SCADA and tank 1 and keeping the one with the smallest mean variance of tank 2's level over the day.

### 3.6 Metrics

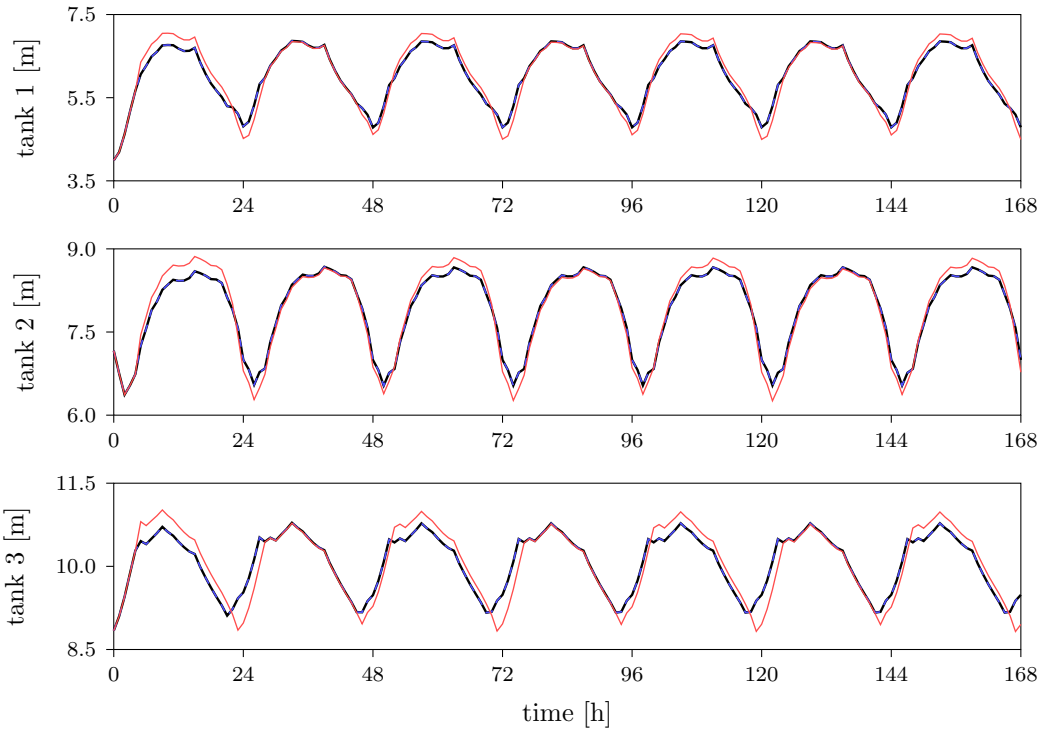
Agreement with EPANET is read as the largest difference in tank level, in head at report times and in tank net inflow, the number of steps and statuses that differ, and the largest offset of a switch time; its reference is zero, and the bounds of Table 2 are the tolerances the study sets for single-precision output and a rounded constant. The integrators are read by the largest level error against EPANET at a 60 s step, with a 30 s run as the check that the reference has converged. The operator questions are read in hours: the first hour at which the filter's posterior standard deviation of a tank's level falls below 5 cm. Calibration is read in three ways. The fraction of data-informed level-hours, those whose posterior standard deviation is below 10 cm, whose truth lies within two posterior standard deviations, against 95.45 % for a calibrated Gaussian and the pre-registered floor of 90 %; the root mean square of the $z$-scores, error over standard deviation, against 1, with a band from resampling the draws; and, for the 27 parameters of each draw, the Mahalanobis misfit

$$\chi_n^2 = e_n^\top \Sigma_n^{-1} e_n, \tag{16}$$

with $e_n$ the MAP minus the truth and $\Sigma_n$ their Laplace covariance, summed over the draws and read against the 0.5 % and 99.5 % points of the chi-square distribution with 27 degrees of freedom per draw. Level-hours of one trajectory are correlated, so the pooled counts carry no binomial interval; the per-draw counts and the bootstrap band stand beside them. Cost is counted in solves.

## 4 Results

**Switch timing against EPANET.** Figure 4 plots the three tank levels over the week for EPANET, for the time loop with EPANET's step rule and for the loop that applies the controls only at whole hours; Table 2 gives the differences. With EPANET's rule the loop takes the same 183 steps as EPANET on Net3 (27 on Net1), every status agrees at every step, the 14 level-control switches of the week happen within 2 s of EPANET's, and the tank levels agree to better than 0.5 mm and the heads to better than 1 mm; the net inflows agree to better than 20 $\mathrm{cm}^3/\mathrm{s}$.




black: EPANET 2.2; dashed blue: the time loop with EPANET's step rule (indistinguishable at this scale); red: the time loop with the level controls applied only at whole hours

Figure 4: Tank levels over Net3's week, hourly. The time loop with EPANET's step rule lies on EPANET's curve (it differs by less than 0.5 mm; Table 2). Applying the level controls only at whole hours moves pump 335's switches by up to an hour, and the levels drift apart by up to 0.58 m; tank 3 overfills by 0.20 m.

Every parity bound is met on Net1's day and on Net3's week. What remains is not round-off: the plugin's Hazen–Williams constant, 10.667 in SI units, is EPANET's 4.727 in US units rounded, and the $1.6 \times 10^{-5}$ relative difference in every pipe's loss shifts the net inflows slightly and accumulates in the levels; with the unrounded constant the inflows agree to the toolkit's precision. Applied only at whole hours, the same controls produce 10 wrong link-hours of 507, level differences of 0.35, 0.29 and 0.58 m, head differences of up to 27.2 m, and they overfill tank 3 by 0.20 m, because pump 335 keeps running up to an hour after tank 1 has reached its upper level. On Net1 the effect is 2.07 m of level. The timing of a switch is the first-order question of an extended-period model.

**The integrator.** Table 3 isolates the integrator on the 4 h before the first level switch, where timing cannot enter. Against EPANET at a 60 s step, forward Euler at 1 h errs by 3.8, 12.1 and 1.1 cm, exactly what EPANET itself errs at 1 h, and BDF1 by 3.2, 9.7 and 1.1 cm; the 60 s and 30 s references differ by less than 0.09 cm, so the reference has converged. Both methods are first order and differ by centimeters; forward Euler is the one that agrees with EPANET. A two-step method (BDF2) is second order only while the forcing is smooth, and hourly demand patterns jump every hour; the corpus problem on integrator order shows its observed order falling from 2 to about 1 across such jumps unless it is restarted at each. One-step second-order methods were not assessed.

Table 2: The time loop against EPANET 2.2's own step log. *EPANET's rule*: controls checked before every solve and the step shortened to each control or tank event; *whole hours*: controls only at the hour. Level and head differences are maxima over tanks, junctions and report times; round-off-level differences are printed as the bounds they meet. *Steps*: the time loop's against EPANET's; *wrong*: link-hours in a different state (a link-hour is one controlled link at one whole hour); *offset*: the largest difference in the time of a level-control switch.

| network | controls | steps | level diff. | head diff. | wrong | offset |
|---|---|---|---|---|---|---|
| Net1 (24 h) | EPANET's rule | 27 / 27 | $< 0.5$ mm | $< 1$ mm | 0 of 25 | 0 s |
| | whole hours | 25 / 27 | 2.07 m | 11.4 m | 1 of 25 | – |
| Net3 (168 h) | EPANET's rule | 183 / 183 | $< 0.5$ mm | $< 1$ mm | 0 of 507 | 2 s |
| | whole hours | 169 / 183 | 0.58 m | 27.2 m | 10 of 507 | – |

Table 3: Largest tank-level error in cm at the whole hours, the first three columns against EPANET at a 60 s hydraulic step, the last three against EPANET at 30 s, of Net3's first 4 h, before the first level-control switch at 4.23 h. The last row is the difference between the two references.

| method | tank 1 | tank 2 | tank 3 | tank 1 | tank 2 | tank 3 |
|---|---|---|---|---|---|---|
| forward Euler (1 h) | 3.8 | 12.1 | 1.1 | 3.9 | 12.2 | 1.1 |
| BDF1 (1 h) | 3.2 | 9.7 | 1.1 | 3.2 | 9.7 | 1.1 |
| EPANET (1 h) | 3.8 | 12.1 | 1.1 | 3.9 | 12.2 | 1.1 |
| EPANET 60 s vs 30 s | 0.03 | 0.09 | 0.01 | | | |

**What the readings pin.** Figure 5 shows the posterior standard deviation of tank 2's level through the nominal day for each reading set. The tanks forget their initial levels: with the switches held, one meter of initial error in tank 1 leaves 0.34 m after six hours and 0.13 m after a day, in tank 2 0.33 and 0.05 m, in tank 3 0.75 and 0.48 m. The linearized time constants $A/(\partial q/\partial h)$ are 3.5, 3.7 and 15.6 h; the memory decays faster than that at first and then levels off, because the three tanks exchange water through the network. With SCADA alone every tank's level stays uncertain by several centimeters at the end of the day (6.3 cm for tank 2); reading tank 1 brings tanks 2 and 3 to 1.5 and 0.8 cm by the end of the day, but tank 2's level at the start stays at 30 cm, close to its 30 cm prior: an initial level that the tank forgets within hours cannot be recovered from readings taken after it has forgotten. Of the 90 junctions, a pressure gauge at junction 50 says most about tank 2 (mean variance 0.00085 $m^2$ against a median of 0.008 $m^2$); junction 50 is the end of the short, wide pipe that joins tank 2 to the network, so its head is tank 2's head less a negligible loss, and the gauge is a level gauge with a pressure sensor's noise.

**The operator questions.** Table 4 answers the first question for each reading set over the 19 draws of the main case. With SCADA alone, tank 1 reaches 5 cm in 12 of 19 draws, at hours 20–24, tank 2 in none of the 19 draws within the day and tank 3 after a median of 20 hours (20–21). Reading tank 1 changes that: tank 1 is known at hour 0 in every draw, tank 2 after a median of 12 hours (11–12) and tank 3 after a median of 12 hours (11–12); reading tank 3 as well brings tank 2 to 5 cm after a median of 7 hours (7–9). On the second question, a level gauge on tank 2 knows its level at hour 0 in every draw; the pressure gauge at junction 50 reaches 5 cm after a median of 2 hours (2–3) — beside the tank, a pressure gauge with ten times the noise does nearly as well, because the model turns its reading into a level, and both do far better than reading the other tanks. On the third, the switch times: added to SCADA, they bring tank 1 to 5 cm after a median of 5 hours (4–21), whereas with SCADA alone tank 1 reaches it in 12 of 19 draws, at

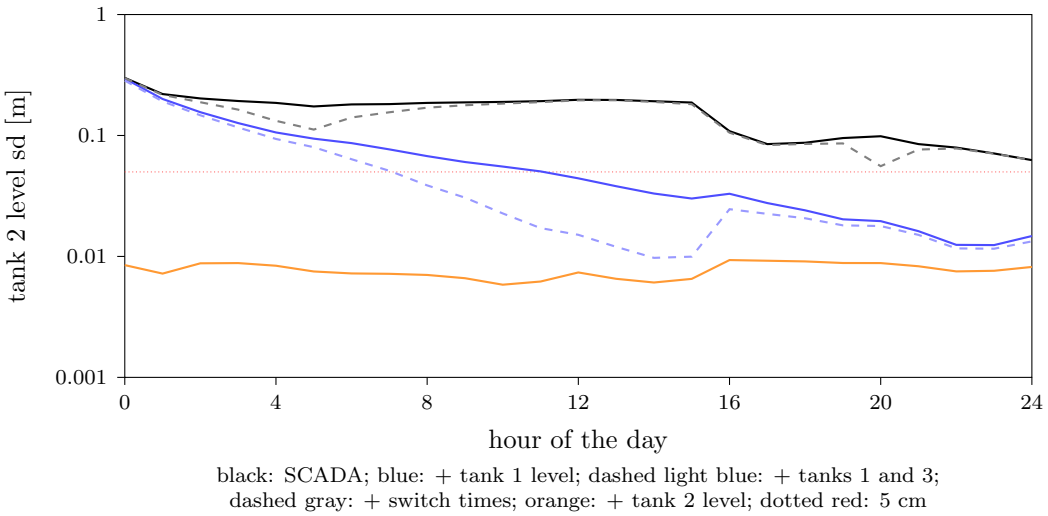
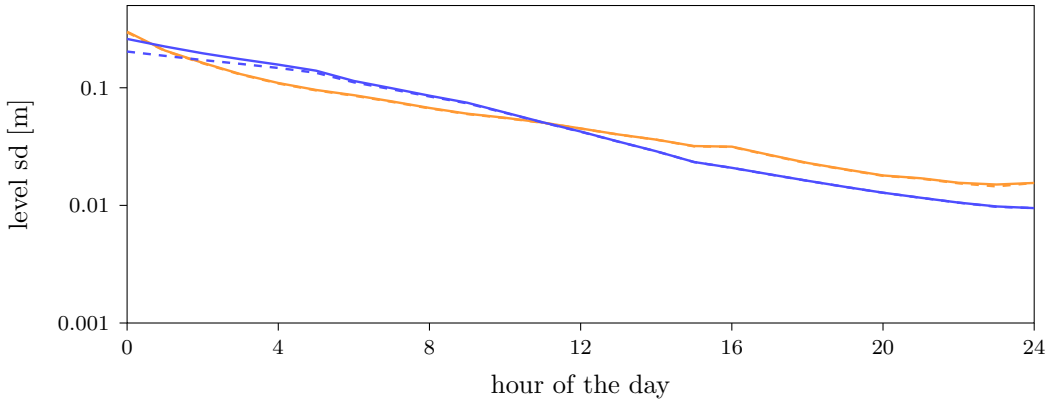



Figure 5: Posterior standard deviation of tank 2's level through the day for each reading set, smoothing over the whole day, on the nominal day (every demand scale 1, initial levels at the operating center, readings without noise). Only a gauge on tank 2 itself pins its level at the start of the day; with SCADA and tank 1 the level becomes known as the day goes on, because the flows that fill and drain the tank are inferred.

Table 4: Hours from a cold start until the filter's posterior standard deviation of each tank's level first falls below 5 cm: the median over 19 draws of the main case, with the range in parentheses. When some draws do not reach it within the day, the cell gives how many do and their range; "none" when no draw does. Hour 0 means the first reading already pins the level.

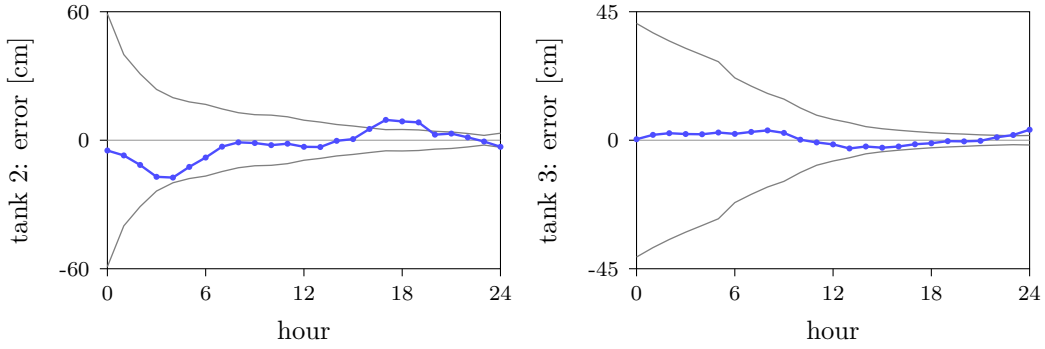
| reading set | tank 1 | tank 2 | tank 3 |
|---|---|---|---|
| SCADA | 12 of 19 draws (20–24) | none | 20 (20–21) |
| SCADA + switch times | 5 (4–21) | 13 of 19 draws (19–24) | 20 (16–20) |
| SCADA + tank 1 | 0 (0–0) | 12 (11–12) | 12 (11–12) |
| SCADA + tanks 1, 3 | 0 (0–0) | 7 (7–9) | 0 (0–0) |
| SCADA + tank 1 + tank-2 gauge | 0 (0–0) | 0 (0–0) | 5 (4–6) |
| SCADA + tank 1 + pressure gauge | 0 (0–0) | 2 (2–3) | 9 (9–10) |

hours 20–24; the gain arrives at the first switch, when the log says where tank 1 stood, and it does not carry over to tank 2 (in 13 of 19 draws, at hours 19–24). Figure 6 compares the filter with the smoother for the two unread tanks: the smoother is narrower at the start of the day, because later readings constrain the initial levels through the physics, and the two coincide once the tanks have forgotten their start. The filter's standard deviation should never be below the smoother's, which reads a superset of its readings; in a linear model it is not. Here it falls below by up to 3.9 cm: in the first draw in 15 of 450 (set, tank, hour) comparisons, in the sets SCADA, SCADA + switch times, SCADA + tank 1 + tank-2 gauge, SCADA + tanks 1, 3. Each standard deviation is a Laplace width at its own MAP point, and where a tank's level is poorly read, its width depends on the estimate through the pump curve and the pipe laws; the check the plan set, at a relative $10^{-3}$, does not hold, and the widths of weakly observed levels are to be read with that dependence in mind.




orange: tank 2; blue: tank 3; solid: filter (readings up to the hour); dashed: smoother (the whole day)

Figure 6: Filter against smoother for the two unread tanks with SCADA and tank 1's level, averaged over the draws of the main case. The smoother knows the start of the day better, because later readings constrain the initial levels through the physics; the two meet as the day goes on and the tanks forget their initial levels.




blue: estimate minus EPANET truth; gray: $\pm 2\sigma$ (clipped at the frame)

Figure 7: The first draw of the main case (seed 0), SCADA and tank 1 read: the estimate of the two unread tanks' levels minus the EPANET truth, hour by hour, with the posterior two-standard-deviation band. Early hours are wide (prior-dominated); later hours are narrow and data-informed.

**Calibration.** Figure 7 shows one draw of the main case: the errors of tanks 2 and 3 against the EPANET truth with their two-standard-deviation bands. Table 5 gives the calibration of every case. Draw 3 is left out of the main case, its variant and the check: its truth overfills tank 3 within the day. In the main case 982 of 1192 data-informed level-hours (82.4 %) lie within two posterior standard deviations of the EPANET truth, against the 90 % floor and 95.45 % for a calibrated Gaussian; the RMS $z$ is 1.60 [1.39, 1.80] and the parameter misfit 2593 on 513 degrees of freedom (above its 99 % band of 434–599). In the main case the coverage floor is not met. The source is the discretization: the truth moves the tanks continuously, the model in hourly forward-Euler steps, and the difference, several centimeters in a day (Table 3), is more than the physics rows allow. With the tank rows widened by the pre-registered rule, 1075 of 1165 (92.3 %) are covered, RMS $z$ 1.10, misfit 574 (inside its band). With the inflated tank rows the coverage floor is met. The machinery check, with the model as its own truth, covers 92.3 % with RMS $z$ 1.03 and a misfit inside its band: the construction is sound, and its coverage below the Gaussian rate comes from

Table 5: Calibration of the tank-level posterior, SCADA and tank 1 read, over the draws each case keeps. A level-hour is one tank at one whole hour; it is *data-informed* when its posterior standard deviation is below 10 cm. RMS $z$ is the root mean square of error over standard deviation (1 when calibrated), with a band from resampling the draws; $\chi^2$ is the Mahalanobis misfit of the 27 parameters summed over draws, and in parentheses where it falls against the 0.5 % and 99.5 % points of its distribution.

| truth | draws | informed, $2\sigma$ | prior-dom. | RMS $z$ [band] | $\chi^2$/dof |
|---|---|---|---|---|---|
| EPANET at 60 s | 19 | 982/1192 (82.4 %) | 207/233 | 1.60 [1.39, 1.80] | 2593/513 (above) |
| EPANET at 60 s, inflated tank rows | 19 | 1075/1165 (92.3 %) | 233/260 | 1.10 [0.92, 1.27] | 574/513 (inside) |
| EPANET at 60 s, demand +10 % | 15 | 722/933 (77.4 %) | 168/192 | 1.81 [1.55, 2.09] | 2376/405 (above) |
| EPANET at 60 s, pattern shifted 2 h | 20 | 897/1257 (71.4 %) | 213/243 | 2.66 [2.46, 2.86] | 14946/540 (above) |
| the model itself (check) | 19 | 1099/1191 (92.3 %) | 211/234 | 1.03 [0.83, 1.22] | 470/513 (inside) |

the correlation of the level-hours within a draw, which the RMS $z$ and the misfit do not show.

**Harder cases.** Under a persistent 10 % demand bias 5 of the 20 truths fill tank 3 within the day and are left out (draws 6, 8, 9, 13, 18); over the rest the posterior demand scales sit on average 0.007 below the biased truth, so the scales absorb the bias, and the data-informed coverage is 77.4 % against the main case's 82.4 %, with RMS $z$ 1.81 against 1.60. The expectation for the persistent bias held. With pattern 1 shifted two hours, the scales sit 0.004 below the truth's effective scales on average, but the coverage falls to 71.4 % and the RMS $z$ rises to 2.66: one scale per hour for the whole network cannot follow a shape that moves some junctions' demand and not others'. The expectation for the shifted pattern held.

**Cost.** The forward week is 418 steady solves and 4 EPANET runs; the inference cases take 92 MAP solves, the operator questions 2964, the observability section 95.

## 5 Discussion

The forward result is a statement about extended-period models in general: once the steady solve is right, agreement over a week is decided by when the controls act, and a model that checks its controls only at its own step boundaries is wrong by more than any integrator can repair. EPANET's rule makes the switch times exact for a forward-Euler update at no extra cost; it is also why the model inherits EPANET's own discretization error at a 1 h step, a few centimeters of level in a day, which the model then cannot distinguish from the truth. That error is what the main case measures: a posterior computed as if the hourly model were exact is overconfident against a truth that moves continuously, and the widening of the tank rows by the error they actually carry is the remedy the protocol tested. A finer hydraulic step in the model is the other remedy, at the cost of more copies in the window.

The inference results hold for the noise and the priors that produced the readings: Gaussian, independent, of known standard deviation, with demands that scale uniformly across the network. The shifted pattern shows what happens when the demand's shape is wrong rather than its size; the persistent bias is absorbed by the scales, at some cost in coverage. Link statuses are taken from the log: a model that had to infer when pumps switched would face a hybrid problem this study does not attempt. The answer to the operator's first question depends on what is read: pump flows and discharge heads alone leave the unread tanks uncertain by several centimeters all day,

because they constrain the sum of what the tanks exchange more than how it is split among them; one tank's level resolves the split, and the others then follow within half a day. An initial level that a tank has forgotten cannot be recovered, whatever the later readings. The Laplace widths of poorly read levels depend on the estimate, which is why the filter can come out narrower than the smoother there. Everything here is one public example network, simulated; the tank memory and the value of a gauge are properties of its layout.

What the study does not do: rule-based controls, pressure-dependent controls, variable-speed pumps, tanks with volume curves, and full or empty tanks are outside the time loop; transport of water and of heat along pipes, with its delay, is outside the model, so the time machinery here does not by itself carry a heating network with long travel times.

# 6 Conclusion

This report asked whether a model built from conservation relations can carry EPA Net3 through a week, and what decides its agreement with EPANET 2.2. It can: with EPANET's rule for when controls act, the model reproduces EPANET's 183 steps, every status and the tank levels to within 0.5 mm over the week; with the controls checked only at whole hours it is wrong by up to 0.58 m of level and 27 m of head, while the choice between forward and backward Euler moves the levels by centimeters. It asked, as an operator would, how long it takes to know an unmetered tank's level from a cold start: with the pump flows and two discharge heads alone, tank 2 reaches 5 cm in none of the 19 draws within the day; with tank 1's level read as well, tank 2 is known after a median of 12 hours (11–12) and tank 3 after a median of 12 hours (11–12). It asked whether a level gauge on tank 2 is worth more than a pressure gauge: the gauge knows the level at hour 0 in every draw, a pressure gauge at the junction beside the tank after a median of 2 hours (2–3). And it asked what logged switch times are worth: they pin tank 1's level at each switch, bringing it to 5 cm after a median of 5 hours (4–21), whereas with SCADA alone it reaches 5 cm in 12 of 19 draws, at hours 20–24. The uncertainties are honest to the extent the model's own time discretization allows: against a truth run at 60 s, 82.4 % of the data-informed level-hours are covered at two standard deviations, and 92.3 % once the tank rows carry their measured discretization error. The next step is a model step finer than the hour where the tanks move fastest, and the transport delay that a heating network needs.

# A Reproduction

```
cd docs/terra_graph_engine/case_studies/water_networks/net3_extended_period
PYTHONPATH=../../../../../src python run.py            # about two hours, single-threaded
PYTHONPATH=../../../../../src python make_report.py --pdf
```

`run.py` runs five stages (forward, observability, calibration, inference, operator), each writing its section of `results.json` at six significant digits; `make_report.py` reads only that file. Seeds 0–19 drive the demand scales and initial levels (a generator per seed) and, offset by 20000, the reading noise; the calibration draws use seeds 100–104; the bootstrap uses seed 7. EPANET 2.2 is the toolkit shipped with WNTR 1.5.0, run on the Net3 and Net1 files it ships. Water density 1000 kg/m$^3$.

## B Data and problems

Net3 and Net1 are the example networks distributed with EPANET 2.2 and WNTR 1.5.0 (public domain, US EPA); no field data are used. Under `problems/`, four directories: *a tank drains through a pipe* (the closed-form level against forward Euler at two steps); *a pump that cycles on a tank's level* (fill, drain and cycle times against the time loop's switch times); *second order lost across demand jumps* (the observed order of forward Euler, BDF1 and BDF2 with smooth and hourly forcing); *how fast a tank forgets its initial level* (the time constant, and the scalar Kalman filter and smoother against the unrolled model). Each is a statement, a script that solves it with the plugin, and its results file.

## References


[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, Washington, DC, EPA/600/R-20/133.

[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.

[Romero-Ben et al. 2026] Romero-Ben, L., Irofti, P., Stoican, F., Puig, V. (2026). Dual unscented Kalman filter architecture for sensor fusion in water networks leak localization. IEEE Transactions on Control Systems Technology 34(1), 343–354. doi:10.1109/TCST.2025.3610975.

[Vrachimis et al. 2022] Vrachimis, S. G., Eliades, D. G., Taormina, R., Kapelan, Z., Ostfeld, A., Liu, S., Kyriakou, M., Pavlou, P., Qiu, M., Polycarpou, M. M. (2022). Battle of the Leakage Detection and Isolation Methods. Journal of Water Resources Planning and Management 148(12), 04022068. doi:10.1061/(ASCE)WR.1943-5452.0001601.

[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.