

CIDE: Calibration identifiability-governed with structural diagnosis of water distribution networks

Alexandre da Silva Gonçalves

Universidade Federal de Goiás: Universidade Federal de Goiás

Guilherme da Cruz dos Reis

Universidade Federal de Goiás: Universidade Federal de Goiás

Kamila Almeida dos Santos

Universidade Federal de Goiás: Universidade Federal de Goiás

Tatiane Souza Rodrigues Pereira

Universidade Federal de Goiás: Universidade Federal de Goiás

Klebber Teodomiro Martins Formiga

klebberformiga@ufg.br

Universidade Federal de Goiás: Universidade Federal de Goiás <https://orcid.org/0000-0003-1094-362X>

Research Article

Keywords: hydraulic calibration, water distribution networks, identifiability, pressure-dependent demand, leakage, pressure-regulating turbine

Posted Date: October 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10515184/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

CIDE: Calibration identifiability-governed with structural diagnosis of water distribution networks

Alexandre Silva Gonçalves¹, Guilherme da Cruz dos Reis¹,
Kamila Almeida dos Santos¹, Tatiane Souza Rodrigues Pereira¹,
Klebber Teodomiro Martins Formiga^{1*}

^{1*}Escola de Engenharia Civil e Ambiental, Universidade Federal de
Goiás, Goiânia, Goiás, Brazil.

*Corresponding author(s). E-mail(s): klebberformiga@ufg.br;

Abstract

The calibration of water distribution networks is generally based on the difference between the observed and simulated pressures and flows, a criterion which is inadequate whenever demand, leakage and roughness all cause similar responses or when a single flow record both balances the system and evaluates the fit. In this study an identifiability-governed calibration and structural diagnosis (CIDE) is applied to the Guanabara sector, which is controlled by a pressure-regulating turbine and which EPANET represents through a pressure-reducing valve that enforces the hourly downstream pressure profile. The measured set-point is introduced as a boundary condition, the inflow is included once as an observation with error, the parameters are shared over two estimation days and six days are set aside for validation. The model incorporates pressure-dependent demand, distributed leakage, a four-block demand shape, a robust loss function, multiple starts and a singular-value analysis of the dimensionless Jacobian. The calibration root-mean-square error was 0.798 m for pressure and 1.040 L/s for flow, the validation flow error was 1.229 L/s and the Nash–Sutcliffe efficiency was 0.830. The Hazen–Williams multiplier was 0.808 and fell within an admissible range of 0.45 to 1.55, while leakage was below 10^{-3} L/s and could not be distinguished from its zero value, showing a correlation of about ± 0.99 with the demand terms. The local rank was 7/7 and the consumption–loss partition remained unidentifiable. A decomposition of the telemetry into minimum-night-flow, independent of the inversion, indicates that the physical leakage lies between 6% and 32% of the inflow, the exact figure depending on the assumed fraction of night demand, an amount which is absorbed by the demand terms. The model is useful for prediction within the range of observed data and does not allow for a causal assignment of the inflow.

Keywords: hydraulic calibration, water distribution networks, identifiability, pressure-dependent demand, leakage, pressure-regulating turbine

1 Introduction

Hydraulic models of water distribution networks support decisions that range from daily operation and pressure control to reinforcement planning, service-level assessment and loss management. The usefulness of these decisions depends on the model representing both the available energy and the spatial and temporal distribution of withdrawals. The parameters that most govern this response — effective roughness, nodal demands, leakage and operational states — are incompletely known, and calibration adjusts inputs and parameters until simulated pressures and flows approach the measurements. Numerical closeness and physical adequacy do not coincide when a model reproduces the fitting data while attributing the result to the wrong mechanisms.

Classical calibration methods relied on field tests, balances and iterative adjustment of demands and resistance coefficients (Walski 1983; Ormsbee and Lingireddy 1997). The inverse-problem formulation allowed parameters to be estimated by least squares, nonlinear programming, genetic algorithms and probabilistic methods (Lansey and Basnet 1991; Kapelan et al. 2007; Do et al. 2016). Automation widened the accessible parameter space without removing the informational limitation. Lansey and Basnet (1991) showed that a nearly zero objective function does not guarantee recovery of the true parameters, and Kang and Lansey (2011) demonstrated that joint estimation of demand and roughness can be inferior to a sequential strategy. The problem is not only to find a better minimum, but to establish which parameter combinations the observations distinguish.

In real networks, non-uniqueness takes recurrent forms. Increasing demand, reducing the Hazen–Williams coefficient or introducing leakage raises head loss and produces similar changes in the measured pressures, and under sparse instrumentation these effects become collinear. A flexible algorithm compensates errors of topology, elevation, diameter or valve state through continuous parameters, and the fit alone accepts this compensation as calibration. Estimating the demand pattern and the resistance coefficients in a single search treats the interference between them as a difficulty of the optimiser, to be met by a wider exploration of the parameter space (Dini and Tabesh 2014), which returns a solution without establishing that the observations distinguish its components. Observation uncertainty, parameter uncertainty and structural discrepancy must be separated, and calibration must be compatible with the intended use of the model (Savić et al. 2009; Hutton et al. 2014a). Practical identifiability provides the language for this separation and asks whether the data, the excitation and the inverse formulation distinguish the parameters within the domain of interest (Brun et al. 2001; Guillaume et al. 2019).

There is also a problem of information accounting. Inflow is often used to reconstruct demand or to impose closure of the same day’s balance, and if the demand multipliers remain free while that same flow reappears in the objective function, one

piece of information is used twice under incompatible roles. A downstream pressure or an observed set-point can be imposed on the model and, at the same time, treated as an independent target. The calibration then appears more informed than it is, and the error matrix does not represent the dependence introduced. Reconstructing a validation day’s demand from that day’s own flow measures a conditional prediction, not out-of-sample predictive capacity (Chew et al. 2022).

The operation of pressure-regulating turbines illustrates the distinction between an operational decision and a system response. The downstream set-point is known and varies through the day, whereas inflow and internal pressures are responses. In this study the real device was represented in EPANET by a pressure-reducing valve only to impose the downstream pressure, without describing rotation, efficiency or energy recovery. Re-estimating the set-point to reduce the error would transfer to the turbine control the role of compensating roughness, demand or cadastral error. Demand-driven analyses assume full delivery even under insufficient pressure, an assumption that distorts delivered demand, inflow and loss estimation when part of the network operates at low pressure. Pressure-dependent formulations and the distinction between consumption and leakage are necessary in these states (Giustolisi et al. 2008; Tabesh et al. 2011; Rossman et al. 2020), but they do not, by themselves, resolve the demand–leakage separation without sufficient independent excitation or measurement.

The literature offers the components to address parts of the problem: calibration under multiple loading conditions, prior information and uncertainty quantification; PDD with pressure-dependent leakage; SVD of sensitivity matrices; robust loss functions; and measurement-campaign design (Lansey et al. 2001; Sanz and Pérez 2015; Zanfei et al. 2020). Multi-period formulations take the first of these components furthest, estimating parameters and states together over several loading conditions and testing beforehand whether the measurement set determines them (Díaz et al. 2017). What these contributions do not supply is a single protocol, applied to a real network, in which measured operation is fixed before inversion, each observation has a single role, dimensionality is conditioned by information, validation stays outside the fit, anomalies are not erased from the primary metric, hydraulic admissibility is separated from numerical convergence, and the legitimate output can be “incompatible structure” rather than a forced parameter set.

This work proposes the CIDE framework and applies it to the Guanabara sector, instrumented with four pressure loggers and the turbine telemetry. Three questions guide the analysis. Do shared physical parameters, estimated with two days and evaluated on the remaining six, improve pressure and flow without reusing validation information? Do hydraulic guards and the SVD of a dimensionless Jacobian distinguish predictive usefulness from parameter interpretability? Does a low aggregate error authorise interpreting roughness, demand and leakage as physical properties? The contribution lies in the protocol, not the optimiser: the downstream pressure is classified as a boundary and removed from the calibration vector; inflow enters once as an observation with error; complexity is reduced to a global roughness multiplier, a sector leakage term and few demand terms; a Jacobian scaled by the uncertainties and

the parameter ranges governs interpretation; and validation, data-quality sensitivity and physical guards are reported separately.

2 Materials and methods

The workflow has six linked steps (Fig. 1). First, the physical link between the telemetry, the turbine and the component holding the sensors is confirmed. Second, two days are reserved for estimation and the remainder for validation. Third, a parsimonious model is built with PDD, pressure-dependent leakage and shared demand patterns. Fourth, the robust inversion is run with multiple starts. Fifth, each simulation must satisfy the guards on downstream-pressure control, balance, warm-up and finiteness. Sixth, the dimensionless Jacobian is decomposed by SVD and the parameters at the bounds are inspected. The calibration is interpreted physically only when performance, guards, rank and position within the parameter domain are jointly coherent.

The final classification admits three outcomes: a model rejected for structural incompatibility; a model that is predictively useful with partially interpretable parameters; or a model accepted within the tested domain. Rank deficiency does not imply structural error, since it may indicate only a parametrisation more complex than the data support, in which case groups are merged or parameters are fixed. A structural diagnosis is issued when the parametric reduction has already been applied and a spatially coherent systematic residual, parameters at compensating bounds and stability of the failure across alternative designs still coexist. Guanabara fell into the second category.

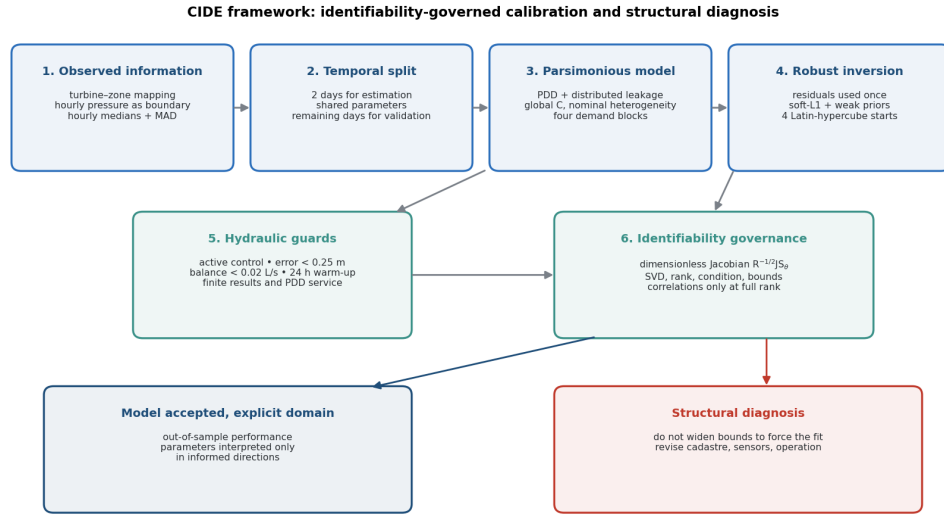


Fig. 1 CIDE framework. The observed operational information is separated from the parameters before inversion; hydraulic guards and identifiability govern the decision between accepting the model within an explicit domain and issuing a structural diagnosis.

2.1 Study network

The model represents a real network in the Guanabara sector, in Goiânia, with a population of about 10,000 inhabitants. The inlet is controlled by a pressure-regulating turbine, operated remotely under fixed rules, and EPANET represents this control by an equivalent pressure-reducing valve (PRV). The full model contains 198 junctions, one reservoir, 249 pipes and one control element. The calibrated component, immediately downstream of the turbine, contains 168 nodes and 219 pipes, totals 14,892.00 m and has a base demand of 6.522 L/s. Four pressure loggers were installed at nodes 469798, 472497, 475495 and 497160 (Fig. 2, Table 1).

The turbine telemetry includes flow, upstream pressure, downstream pressure and set-point, at a modal resolution of about six minutes. The single turbine identified in the model feeds the component holding the four sensors. This topological check preceded the calibration, because an incorrect mapping would change the meaning of the mass balance and yield a different inverse problem. The loggers covered about seven days (29 May to 5 June 2026) and the telemetry covered about four and a half months, which provides the seasonal context of the flow. The upstream reservoir head does not enter the zone calibration, since the turbine regulates the downstream pressure independently of the source level. The measured upstream pressure was used only to confirm that the device operated with a positive differential in every hour, and it entered neither the boundary conditions nor the objective function. The utility states an elevation uncertainty of 1 to 2 m; the calibrated component is bounded by the regulated downstream head rather than by the reservoir level, so this uncertainty affects the nodal pressures and not the energy imposed at the inlet.

Two days inform the estimation, Saturday 30 May and Wednesday 3 June, which exposes the shared parametrisation to two consumption categories without allowing a free multiplier per date. Validation covers 29 and 31 May and 1, 2, 4 and 5 June, six days predicted without re-estimation. The observed set-point was identical in every interval, at 7 m between 00:00 and 06:00 and 15 m for the rest of the day, and the hourly profile applied to the simulations was the median of the two calibration days. The 4th of June was classified as a weekday because it is a Thursday; its coincidence with the Corpus Christi holiday, with no local operational metadata, is acknowledged as a limitation.

2.2 Code

The calibration was implemented in a modular, open Python package running on EPANET 2.2 through the WNTR library (Rossman et al. 2020; Klise et al. 2017). Separate modules handle the reading and aggregation of the series, the imposition of the boundary conditions, the parsimonious hydraulic model, the robust inversion and the identifiability diagnosis, and a test suite checks the topology, the hydraulic guards and the core routines. The modular separation keeps each series with a single role and makes the result reproducible: the estimated parameters, the per-sensor and per-day metrics, the dimensionless Jacobian and the calibrated models are written to files, and the same configuration reproduces the reported numbers. This organisation is part of the contribution, because traceability — which data define the boundary,

Guanabara — full model: 198 nodes, 249 pipes; domain: 168 nodes, 219 pipes

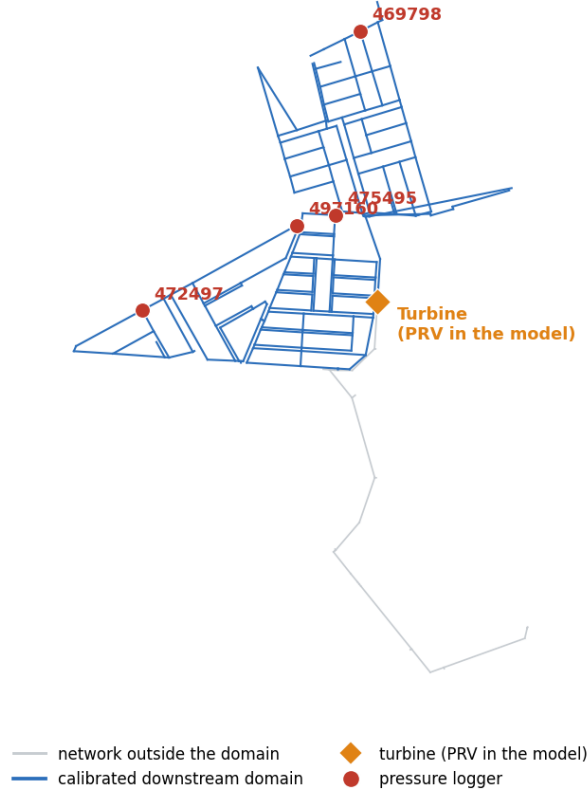


Fig. 2 Full Guanabara model and downstream domain used in the calibration. The four loggers belong to the component fed by the telemetered turbine, represented in EPANET by a PRV element.

which enter the fit, which remain for validation — is as important as the algorithm when the model informs pressure reduction or works prioritisation.

2.3 Data and observational error

The sub-hourly series were aggregated to hourly values by the median, less sensitive to peaks than the mean, and the dispersion was estimated by the scaled median absolute deviation, $1.4826 \text{ median}|x - \text{median}(x)|$. The total error combines a floor and the dispersion in quadrature,

$$\sigma_{s,t} = \sqrt{\sigma_{\text{floor}}^2 + \sigma_{\text{MAD},s,t}^2}, \quad (1)$$

Table 1 Characteristics of the Guanabara sector, instrumentation and temporal split.

Characteristic	Guanabara
Full model	199 nodes, 250 links; 198 junctions, 249 pipes, 1 equivalent PRV
Calibrated domain	168 nodes, 219 pipes, 14,892.00 m
Base demand in the domain	6.522 L/s
Length-weighted nominal C	139.831
Pressure loggers	469798; 472497; 475495; 497160
Turbine (PRV ID) / downstream 00–06 h; 06–24 h	454031_n2a / 7; 15 m
Calibration days	30/05 and 03/06/2026
Validation days	29 and 31/05; 01, 02, 04 and 05/06
P/Q observations in calibration	192 / 48
P/Q observations in primary validation	474 / 111

with a floor of 0.30 m for pressure and, for flow, the larger of 0.15 L/s and 3% of the absolute hourly median. These terms represent instrument precision, within-hour variability and representativeness error, not pure independent noise. Consecutive observations remain autocorrelated, a limitation considered when interpreting the sample size and the SVD.

Range, finiteness and consistency filters were applied before aggregation, following the distinction between useful and misleading calibration data (Walski 2000). No validation point was excluded by a rule derived from the training set. Each sensor’s alert envelope was built only from the two calibration days, taking the hourly median and range, and a validation observation is flagged when it departs from the centre beyond a limit proportional to the range,

$$|o_{s,t} - m_{s,t}| > \kappa \Delta_{s,t}, \quad (2)$$

with Δ equal to the training range. The alerts remain in the primary metric, and a secondary metric, labelled QA sensitivity, measures only the effect of removing them. This separation prevents a rule learned on the training set from silently improving the validation. Missing and anomalous records are commonly detected and compensated before the parameters are estimated, so that the demands do not absorb sensor failure (Chu et al. 2021). Here the detection is kept outside the estimation and outside the primary metric, because the envelope is fitted to the calibration days and applying it to the validation days would let the training set decide which predictions are judged.

2.4 Hydraulic formulation

For each node i and hour t , the continuity balance in the component was written as

$$Q_t^{\text{in}} = \sum_i (q_{i,t}^d + q_{i,t}^\ell), \quad (3)$$

where q^d is the delivered consumption and q^ℓ the leakage. Pressure-dependent demand used minimum pressure $p_{\min} = 0$ m, required pressure $p_{\text{req}} = 15$ m and exponent

$\gamma = 0.5$,

$$q_{i,t}^d = \begin{cases} 0, & p_{i,t} \leq p_{\min}, \\ q_{i,t}^{\text{req}} \left(\frac{p_{i,t} - p_{\min}}{p_{\text{req}} - p_{\min}} \right)^\gamma, & p_{\min} < p_{i,t} < p_{\text{req}}, \\ q_{i,t}^{\text{req}}, & p_{i,t} \geq p_{\text{req}}. \end{cases} \quad (4)$$

The sector leakage was parametrised as a total reference flow at 10 m, $Q_{\ell,10}$, for the whole calibrated component — not per unit length — distributed among the nodes in proportion to half the length of each incident pipe,

$$q_{i,t}^\ell = Q_{\ell,10} \frac{L_i}{\sum_j L_j} \left(\frac{p_{i,t}}{10} \right)^{N_\ell}, \quad L_i = \frac{1}{2} \sum_{k \in \mathcal{E}(i)} \ell_k, \quad (5)$$

with $N_\ell = 0.5$ in the principal scenario. This exponent corresponds to flow through an opening of fixed area and is the lower end of the range reported for real networks, where values between 0.5 and 1.5 arise from the deformation of the leak area with pressure; the conservative end was adopted because the available excitation has two set-point levels and does not resolve the exponent, and $N_\ell = 1$ was carried as a robustness scenario (Gupta et al. 2016; AL-Washali et al. 2016). The length-based distribution converts the sector quantity into emitters in a reproducible way, without a free coefficient per node (Gupta et al. 2016). A preliminary analysis with two pipe groups produced a correlation above 0.95 between the roughness multipliers, so the final parametrisation adopted a single global multiplier m_C over each nominal value,

$$C_j = m_C C_j^{\text{nom}}, \quad (6)$$

which preserves the cadastral heterogeneity and estimates only the global resistance level. The temporal demand shape was defined by four six-hour blocks, with the first fixed at 1 and the other three estimated, and the four values were normalised to unit mean, separating shape from scale. Two category scales, weekday and weekend, were estimated because the two calibration days covered both. The nodes retained their base demands and relative nominal patterns.

The hourly downstream-pressure profile was imposed as the PRV setting through daily controls, with the valve active in every hour, and the flow was not imposed, remaining a response of the balance among demand, leakage, resistance and PDD. Imposing pressure and flow at the same inlet would over-specify the problem and remove from the flow the independent content needed to estimate demand and losses. The model was simulated for 48 h with an hourly step, and only the second cycle, after warm-up, was compared with the observations. An evaluation was considered valid only when four guards held: finite results; control active in 100% of the hours; a maximum downstream-pressure error below 0.25 m; and a maximum difference between inflow and component outflow below 0.02 L/s.

2.5 Inverse problem and validation

The flow was not used to build demand closure. Pressures and flow entered the objective function once, each observation standardised by its total error,

$$F(\theta) = \sum_{s,t} \rho \left[\left(w_P \frac{P_{s,t}^{\text{sim}}(\theta) - P_{s,t}^{\text{obs}}}{\sigma_{s,t}^P} \right)^2 \right] + \sum_t \rho \left[\left(\frac{Q_t^{\text{sim}}(\theta) - Q_t^{\text{obs}}}{\sigma_t^Q} \right)^2 \right], \quad (7)$$

with unit pressure weight in the principal scenario. The vector held seven parameters: m_C , $Q_{\ell,10}$, three shape terms and two category scales. Their admissible ranges were 0.45 to 1.55 for m_C , 0 to 5.029 L/s for $Q_{\ell,10}$, 0.10 to 8.0 for each shape term and 0.25 to 4.0 for each category scale. The leakage limit corresponds to 80% of the mean inflow on the calibration days, and the ranges are wide relative to the estimates, so a parameter that reaches a bound reports a formulation that does not require it rather than a truncated search. Weak priors were incorporated as residuals, lognormal centred at 1 for m_C with log deviation 0.35, a leakage penalty centred at zero, and lognormal priors for the shapes and scales, which stabilises weak directions without turning the solution into a Bayesian posterior. The log deviation was 0.75 for the shapes and the scales, and the leakage residual was $Q_{\ell,10}$ divided by 2.515 L/s, that is 40% of the mean calibration inflow, so a leakage of the order of the sector consumption is penalised by about one standard residual. The fit used nonlinear least squares with a reflective trust region, minimising the soft-L1 transform of the squared standardised residuals in the form implemented by SciPy (Virtanen et al. 2020), from four starting points — the reference state and three from a Latin hypercube (McKay et al. 1979) — with the best solution chosen by total robust cost. Each start was limited to 90 evaluations, the variables were scaled by the Jacobian, the robust loss made the transition at unit standardised residual, and the tolerances on the cost, on the step and on the gradient were 2×10^{-5} . That limit acts on the counter kept by the optimiser, which registers the calls that advance the iteration and not the extra calls that build each forward-difference Jacobian; the evaluation totals reported in Section 3.1 come from a counter inside the package that increments on every call to the residual function, so the two quantities measure different things and the second is not bounded by 90. The numerical derivatives used a relative step of 0.015, so the absolute increment is 1.5% of the current value of each parameter and shrinks with it, which matters for the leakage term and is examined in Section 3.3. The hypercube samples were clipped to 2% inside each bound, which avoids initialising the search on a hydraulically infeasible state. The estimation therefore uses a direct optimiser; the Jacobian described next is a diagnosis after convergence, not the search method.

After estimation, the Jacobian of the 240 data residuals with respect to the parameters was computed by finite differences and made dimensionless by weighting with the observational uncertainties and scaling by half the admissible range of each parameter,

$$\tilde{J} = R^{-1/2} J S_\theta, \quad (8)$$

where R is diagonal with the observational variances and S_θ holds the parameter half-ranges. The singular-value decomposition

$$\tilde{J} = U \Sigma V^\top \quad (9)$$

determines the effective rank at a relative threshold of 1% of the largest singular value. At full rank, the condition number is $\sigma_{\max}/\sigma_{\min}$ and the local parameter correlation is obtained from $(\tilde{J}^\top \tilde{J})^{-1}$; at deficient rank, the pseudoinverse correlations are not reported, since they would assign zero variance to the unobservable space. Parameters with $|r| > 0.95$, close to the bounds or associated with singular values near the cut are treated as weakly separable.

The two-day parameters were frozen and applied to the remaining days, without recomputing closure, shape or scale from the validated day's flow. For each sensor, day and variable, the number of observations, the root-mean-square error, the mean absolute error, the bias and the Nash–Sutcliffe efficiency (Nash and Sutcliffe 1970) were reported,

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{k=1}^n (s_k - o_k)^2}, & \text{MAE} &= \frac{1}{n} \sum_{k=1}^n |s_k - o_k|, \\ \text{Bias} &= \frac{1}{n} \sum_{k=1}^n (s_k - o_k), & \text{NSE} &= 1 - \frac{\sum_{k=1}^n (s_k - o_k)^2}{\sum_{k=1}^n (o_k - \bar{o})^2}, \end{aligned} \quad (10)$$

where s and o are the simulated and observed values and $\bar{o}$ the observed mean. The unadjusted comparator used the same hydraulic protocol — active hourly control, PDD, warm-up and guards — with $m_C = 1$, zero leakage, flat shape and unit scale, so that the error reduction isolates the effect of the estimated parameters. The primary validation preserved every observation approved by the basic filters, and the QA sensitivity measures the effect of removing the flagged points. All robustness scenarios — leakage exponent $N_\ell = 1$, required pressure $p_{\text{req}} = 10$ and 20 m, and pressure weights 0.5 and 2 — were re-inverted with the same split of two estimation days and six validation days.

3 Results

3.1 Hydraulic consistency and aggregate performance

The selected solutions satisfied the execution guards: control stayed active in every hour, the maximum downstream-pressure error stayed below 3.1×10^{-5} m and the balance error below 2.5×10^{-4} L/s. The difference between cycles after warm-up was below 4.8×10^{-4} in the monitored quantities. The discrepancies described below do not arise from convergence failure, inactive control, balance violation or an undissipated initial state. The four starts terminated at robust costs between 139.816 and 139.823, a spread of 0.005%, and at root-mean-square standardised residuals between 1.4275 and 1.4277, with 63, 90, 121 and 82 residual evaluations, 356 in all. The estimates of

the seven parameters agreed to the fourth decimal place across the starts, which is consistent with a single basin of attraction over the region those four starts explored and does not exclude other minima elsewhere in the admissible box.

Table 2 compares the unadjusted model and the estimated model under the same hydraulic protocol. The unadjusted model already includes PDD, warm-up, active control and the observed turbine schedule, so the error reductions isolate the effect of the parameters. Pressure and flow improved by about 40% on the two fitting days. Out of sample, flow RMSE fell 31.9%; for pressure, the reduction was 2.4% in the primary metric, which keeps every observation approved by the basic filters, and 28.7% in the QA sensitivity, which is a secondary analysis; the difference between the two comes from ten abrupt records preserved at logger 469798 and discussed in Section 3.2.

Table 2 Performance of the unadjusted and estimated models under the same protocol. Bias is simulated minus observed; P in m and Q in L/s. Validation is the primary metric; the QA row is a secondary analysis that excludes the ten flagged records at logger 469798 and does not replace the primary validation.

Set	Var.	n	Unadjusted RMSE	Estimated RMSE	MAE	Bias
Calibration	P	192	1.334	0.798	0.652	+0.188
Calibration	Q	48	1.754	1.040	0.820	−0.081
Validation	P	474	4.772	4.656	1.376	+0.976
Validation (QA)	P	464	1.455	1.038	0.761	+0.352
Validation	Q	111	1.805	1.229	0.970	−0.094

3.2 Predictive performance and parameter caveats

The Hazen–Williams multiplier was $m_C = 0.808$, corresponding to a weighted effective C of 112.926, away from the bounds and plausible as an aggregate indicator of higher resistance than the cadastre. The reference leakage at 10 m, defined for the whole calibrated component, was below 10^{-3} L/s and is not distinguishable from the zero bound. The relative step of the optimiser makes the finite-difference increment 1.5% of the current value, and at estimates of that order the increment falls two orders of magnitude below the 10^{-4} L/s at which the noise of the hydraulic solver begins to dominate the derivative, so the term stops being resolved once it drops below about 7×10^{-3} L/s and the last part of the descent to zero carries no usable gradient. The scales were 0.862 for weekdays and 0.962 for weekends, which corresponds to mean demands of 5.62 and 6.27 L/s. The estimated shape terms were 2.520 for 06–12 h, 2.652 for 12–18 h and 2.191 for 18–24 h against the value fixed at 1 for 00–06 h, which after normalisation to unit mean gives 0.478, 1.205, 1.268 and 1.048 for the four blocks. The night block therefore carries less than half the daily mean withdrawal, and the two daytime blocks differ by 5%, an amplitude compatible with a residential sector under night pressure modulation.

Figure 3 presents the observed and simulated series at the four loggers. Three sensors are reproduced across the whole window, with validation RMSE between 0.62

and 1.34 m and bias from -0.50 to $+0.70$ m, and nodes 472497 and 497160 keep a small positive daytime bias. Logger 469798 follows the cycle for most of the period, except on 5 June, when ten observations between 03–06 h and 09–14 h ranged from 0.1 to 28.8 m while the simulation ranged from 31.65 to 39.55 m. The QA rule, learned on the calibration days, flagged these ten values, which remained in the primary validation and raised the sensor RMSE to 9.23 m; removing them only in the QA sensitivity lowers the aggregate pressure RMSE from 4.656 to 1.038 m. The distance between the two figures is what a data-cleaning decision is worth on this sector, and procedures that detect and compensate such records before the fit report the second (Chu et al. 2021). The episode is compatible with sensor removal, air intake, local depressurisation or a recording fault, and the files do not allow choosing among these hypotheses. It is an event confined to one sensor and one day, and not the focus of this work.

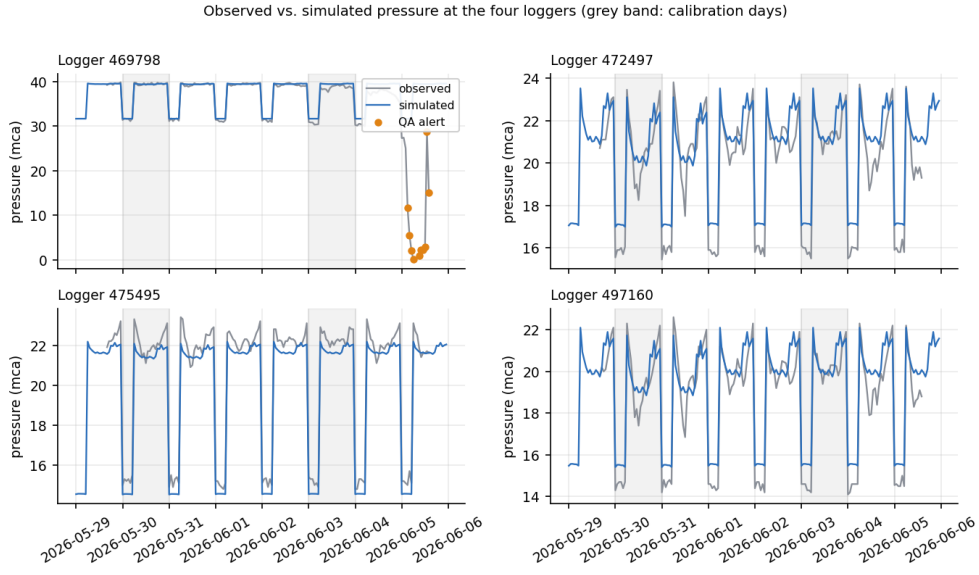


Fig. 3 Observed (grey) and simulated (blue) pressure at the four loggers; the light-grey band marks the two calibration days and the orange points mark the ten QA alerts at logger 469798 on 5 June. Three sensors are reproduced across the window; the localised episode at logger 469798 is the only marked departure.

Table 3 reports the validation metrics of each series. Nash–Sutcliffe efficiency separates the loggers more sharply than RMSE, since it refers the error to the observed variance of each sensor: 0.961 at 475495, 0.866 at 497160 and 0.811 at 472497, against 0.031 at 469798 with the ten alerts retained and 0.876 once they are removed. Pooling the four series gives 0.678 in the primary metric and 0.982 in the QA sensitivity, and the inflow reaches 0.830 out of sample against 0.876 on the calibration days. The pooled figure refers the error to the variance of the four series taken together, which includes the differences of level between sensors, so it sits above what each sensor

obtains on its own and belongs beside the per-sensor values rather than in their place. The small loss between fitting and prediction indicates that the demand shape and the category scales transfer to days outside the estimation.

Table 3 Validation metrics by series, with parameters frozen after the two estimation days. Bias is simulated minus observed. Pressure in m and inflow in L/s. The QA columns exclude the ten flagged records at logger 469798 and are a secondary analysis.

Series	n	RMSE	MAE	Bias	NSE	n (QA)	RMSE (QA)	NSE (QA)
P 469798	117	9.230	3.254	+3.201	0.031	107	1.338	0.876
P 472497	119	1.125	0.936	+0.699	0.811	119	1.125	0.811
P 475495	119	0.619	0.541	−0.499	0.961	119	0.619	0.961
P 497160	119	0.970	0.803	+0.541	0.866	119	0.970	0.866
P pooled	474	4.656	1.376	+0.976	0.678	464	1.038	0.982
Inflow	111	1.229	0.970	−0.094	0.830	111	1.229	0.830

The inflow residual has a structure that the aggregate metrics do not expose. Averaged over the six validation days, the simulation exceeds the observation by 0.17 L/s between 00 and 06 h and falls short by 0.17 L/s over the rest of the day, and the largest departures occur at 07 h, with +1.61 L/s, at 18 h, with −2.45 L/s, and at 23 h, with +1.89 L/s. The four-block shape holds the withdrawal constant inside each six-hour interval and cannot follow the morning and evening transitions, and the residual changes sign exactly at the block boundaries. Refining the shape would follow the transitions, at the cost of adding terms to a set already showing correlations above 0.96 among the demand directions.

3.3 Identifiability

Figure 4 brings together the out-of-sample performance, the singular spectrum of the dimensionless Jacobian and the local parameter correlation. Panel (a) places the validation pressures on the identity line, apart from the ten flagged records, and the inflow inset shows the same agreement for the balance; panels (b) and (c) refer to the same solution. The Jacobian had singular values 208.5; 190.6; 121.5; 107.1; 49.63; 36.13 and 4.412. The rank was 7/7 at the 1% threshold, the ratio of the smallest to the largest singular value was 2.1% and the condition number was approximately 47. All local directions exceeded the cut, and the smallest singular value sits just above it. Repeating the decomposition with finite-difference steps between 5×10^{-2} and 3×10^{-4} L/s in the leakage direction holds the rank at 7/7 and moves the condition number between 47.3 and 43.0, so neither depends on the particular step; below 10^{-4} L/s the step reaches the noise floor of the hydraulic solver and the spectrum stops describing the model.

The correlation matrix shows that leakage and the demand terms are nearly collinear. Shape 06–12 and shape 12–18 correlate at +0.99, leakage and the weekday scale at −0.99, leakage and the weekend scale at −0.99, and every combination among

leakage, shapes and scales shows $|r| > 0.96$. The roughness multiplier is the only well-separated direction. Full rank means that the seven directions remain distinguishable at the chosen threshold, which describes the local geometry of the residual surface and not the predictive capacity; the prediction is supported by the temporal validation of Section 3.2, and neither of the two supports the causal separation of the balance components. Zero-bound leakage does not demonstrate the absence of physical loss; it indicates that, given the instrumentation, the pressure excitation and the flexibility of the demands, no positive leakage contribution was needed to minimise the objective. The consumption–leakage sum is informed for the aggregate prediction, whereas its partition remains practically unidentifiable.

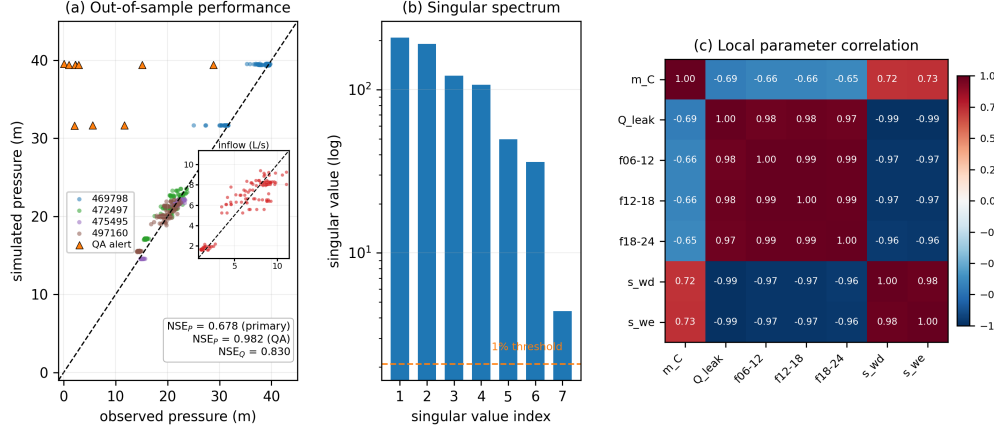


Fig. 4 Identifiability diagnosis. (a) Simulated against observed pressure on the six validation days, by logger, with the ten QA alerts as triangles and the inflow in the inset; (b) the singular spectrum of the dimensionless Jacobian, with the 1% threshold; (c) the local correlation of the seven parameters. Out-of-sample agreement and full local rank coexist with correlations near ± 0.99 among leakage, shapes and scales, which indicate the unidentifiability of the consumption–leakage partition.

Table 4 collects the parameters and the identifiability and service diagnostics of the principal solution.

3.4 Pressure-dependent service and robustness

Figure 5 describes the pressure deficit in the component. Of the 168 junctions downstream of the turbine, 143 carry base demand, and the service indicators are computed over those, since the remaining 25 represent topology and serve no consumer. The minimum pressure at a junction with demand was 4.65 m on the weekday model and 4.63 m on the weekend one. Eighty of the 143 junctions, 55.9%, fell below the required pressure of 15 m in at least one hour, and 18.7% of the node–hour pairs were in deficit. The deficit concentrates in the night modulation, with 80 junctions below 15 m between 00 and 06 h, when the set-point is 7 m, against 9 during the day, when it is 15 m. Mean delivered demand was 98.9–99.0% of the required demand,

Table 4 Parameters, identifiability and hydraulic diagnostics of the principal solution. The service indicators are computed over the 143 junctions that carry base demand, out of the 168 in the calibrated component.

Indicator	Guanabara
m_C / weighted effective C	0.808 / 112.926
Reference leakage at 10 m (component)	$< 10^{-3}$ L/s
Demand scale	0.862 weekday; 0.962 weekend
Effective rank	7/7
$\sigma_{\min}/\sigma_{\max}$	2.1%
Local condition number	≈ 47
Parameter at a bound	leakage
Minimum pressure at a node with demand	4.65 m
Nodes with demand below 15 m at some instant	80/143
Node-hours below 15 m, nodes with demand	18.7%
Mean delivered demand	98.9–99.0%

because most of the deficit occurs near the threshold and not at all nodes at once. The worst delivery occurs at 0 h, at the start of the night modulation, with 92.1% of the required demand delivered. The minimum over all 168 junctions, including those without demand, is 3.58 m, at node 478604, a dead end at the highest elevation of the component, 748.05 m against a range of 720.00 to 748.05 m, fed by a single pipe of 410.09 m and 100 mm. That node serves no consumer and its pressure responds to the leakage coefficient within the numerical noise of the estimate, so it is reported here and kept out of the service indicators. An acceptable fit at the instrumented points can therefore coexist with distributed service loss across the network.

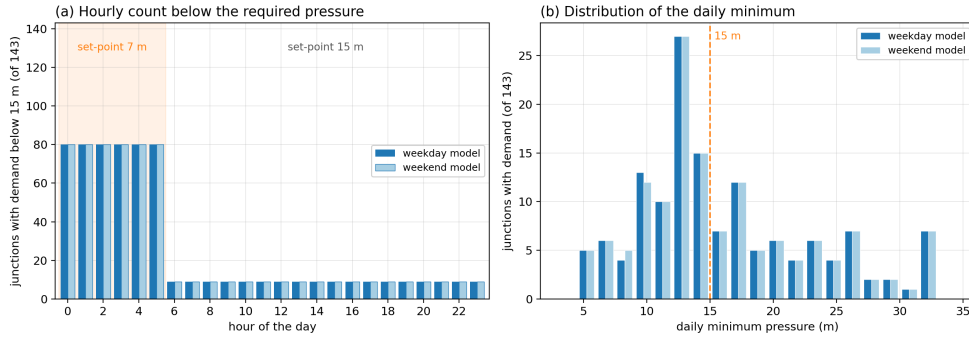


Fig. 5 Number of junctions with demand below 15 m in each hour of the day, over the 143 junctions that carry demand, and distribution of the daily minimum pressure over the same junctions, with the 15 m threshold. The set-point is 7 m between 00 and 06 h and 15 m for the rest of the day, and the weekday and weekend models give the same hourly counts.

Figure 6 and Table 5 summarise the robustness scenarios. With $p_{\text{req}} = 20$ m, the errors stayed nearly equal, the rank remained 7/7 and leakage stayed below 10^{-3} L/s, but 106 of the 143 junctions with demand fell below the required pressure and mean

delivery dropped to 95.3–95.7%, so the requirement changes the service assessment without affecting the sensor fit. With $p_{\text{req}} = 10$ m the count falls to 24 junctions and the node-hours in deficit to 4.2%, while the errors, the rank and the estimated parameters stay inside the spread of the other scenarios. Changing the leakage exponent from 0.5 to 1 did not change the results, since the coefficient stayed at its lower bound, and pressure weights 0.5 and 2 preserved rank 7/7, m_C near 0.805 and leakage below 10^{-3} L/s. The diagnosis is stable across the six scenarios, and the service assessment is conditional on p_{req} .

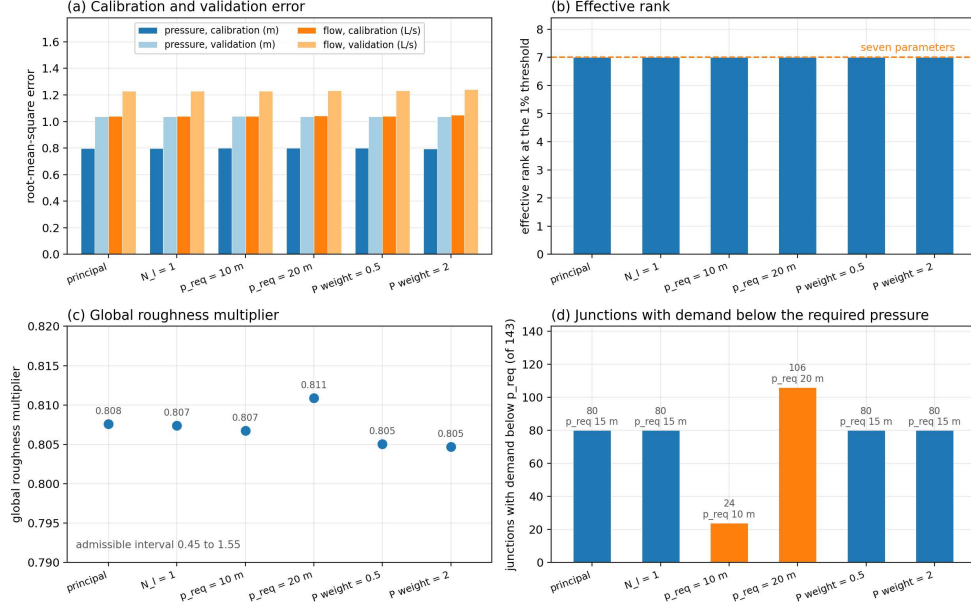


Fig. 6 Calibration and validation errors, effective rank and global roughness multiplier across the six scenarios, and the number of junctions with demand below the required pressure in each. The required pressure changes the service count without changing the estimates.

4 Discussion

4.1 Adequacy and identifiability

For the Guanabara network, the objective function is necessary to estimate parameters but insufficient to decide what a calibrated model means. The fall of the out-of-sample error and the reproduction of the hourly cycles give predictive usefulness within the observed period, the turbine control policy and the observed regime. Zero-bound leakage, the extreme correlations and the qualitative sensitivity to the required pressure prevent interpreting all parameters as causally recovered properties. The same set can be adequate for predicting pressures and flows under similar conditions and inadequate for attributing the inflow between consumption and loss.

Table 5 Robustness scenarios. Calibration (cal.) and validation (val.) RMSE in m for pressure and in L/s for flow; leakage in L/s. The validation columns are the primary metric and the QA column is the secondary analysis. The estimated leakage lies below 10^{-3} L/s in every scenario, which is below the noise floor of the finite differences, so the six values are not distinguishable from one another and are reported as a bound.

Scenario	P cal.	Q cal.	P val.	P val. QA	Q val.	Rank	m_C	Leak.
principal	0.798	1.040	4.656	1.038	1.229	7/7	0.808	$< 10^{-3}$
$N_\ell = 1$	0.798	1.040	4.656	1.038	1.229	7/7	0.807	$< 10^{-3}$
$p_{\text{req}} = 10$ m	0.801	1.040	4.657	1.041	1.229	7/7	0.807	$< 10^{-3}$
$p_{\text{req}} = 20$ m	0.801	1.044	4.655	1.037	1.234	7/7	0.811	$< 10^{-3}$
P weight = 0.5	0.801	1.041	4.656	1.038	1.232	7/7	0.805	$< 10^{-3}$
P weight = 2	0.795	1.049	4.656	1.036	1.242	7/7	0.805	$< 10^{-3}$

This distinction recalls an old warning of network calibration, that estimated parameters are not automatically properties of the system, because different combinations can generate equivalent responses and a structural error can be transferred to continuous variables (Walski 1983; Lansey and Basnet 1991). The practical contribution of CIDE makes that warning operational, since it submits the solution to explicit criteria of information use, temporal validation, hydraulic consistency, rank, correlation and position at the bounds, and admits the refusal to interpret the parameters as a legitimate outcome. This refusal does not require a true model. Environmental models are not verified in the strong logical sense, and validation remains conditional on the tests performed (Oreskes et al. 1994); structural incompatibility means only that, within the observed conditions and uncertainties, the cadastral structure and the parametrisation do not reproduce pressure and flow without extreme compensations.

The scaling of the Jacobian is what gives the diagnosis its meaning. Singular values of a dimensional Jacobian change when leakage goes from L/s to m^3/s or when a multiplier’s range is reparametrised, and weighting by the uncertainties and the half-ranges compares parameter variations of similar meaning (Sanz and Pérez 2015). The 1% threshold is not a physical law, and the smallest singular value is 2.1% of the largest, so a slightly larger cut would remove one direction. For this reason the spectrum is presented together with the correlations, not only the 7/7 label. Full rank and practical identifiability are not synonyms, since the local inversion exists while shapes, scales and leakage remain nearly collinear (Brun et al. 2001; Guillaume et al. 2019).

Field tests, balances and iterative adjustment established the practice of reconciling a model with the measured operation before estimating anything (Walski 1983; Ormsbee and Lingireddy 1997). The inverse formulation replaced that reconciliation by explicit minimisation, and its evolutionary and Bayesian descendants widened the search and attached distributions to the estimates (Lansey and Basnet 1991; Kapelan et al. 2007; Do et al. 2016; Hutton et al. 2014a), brought the demand pattern and the resistance coefficients into a single estimation (Dini and Tabesh 2014) and gathered several loading conditions into one state-estimation problem preceded by an observability test (Díaz et al. 2017), while real-time frameworks carried the same question into continuous monitoring, where model error and measurement error have to be

separated as the records arrive (Hutton et al. 2014b). Singular-value analysis of sensitivity matrices was proposed to design measurement campaigns and to rank the demand components that the sensors distinguish (Sanz and Pérez 2015), and pressure-dependent formulations with distributed leakage supplied the hydraulics required when part of the network operates below the service threshold (Giustolisi et al. 2008; Tabesh et al. 2011). CIDE retains these components and changes the order in which they act and the authority each one holds. The roles of the series are fixed before the inversion, so the set-point does not become a free variable and the inflow is not counted twice; the parameter vector is reduced to what two days of four pressures and one flow support; the scaled Jacobian is read after convergence, as a diagnosis of the accepted solution and not as a criterion for designing the campaign; and the admissible outcomes include declining to interpret part of the vector, which the objective function alone does not report.

4.2 Roles of the information

Much of the innovation lies not in adding parameters but in limiting what each series can do. The turbine’s downstream-pressure profile is a recorded operational decision, and treating it as a boundary condition prevents the calibration from using the controller to absorb unexplained head loss. The inflow enters as an uncertain target and is not used beforehand to force the demand multipliers to close each day’s balance. Imposing pressure and flow at the same inlet would over-specify the problem in this isolated sector, and the flow would cease to respond to demand and leakage, losing its main balance information.

The non-redundant allocation makes the validation more demanding. If a withheld day’s flow reconstructed its own demand, the flow agreement would be partly tautological and the pressure would be predicted conditionally on the information one intends to assess. With shapes, scales, roughness and leakage frozen after two days, the flow of the remaining days is a genuine observation of the temporal transfer, and the validation flow RMSE falls 31.9% without daily closure. The unadjusted comparator shares the turbine control, the observed schedule and the PDD, and differs only in the parameters, so the estimated gain is conservative. Reconstructing the measured operation first can produce larger gains than any estimation algorithm, and mixing the steps prevents knowing which correction produced the result (Walski 2000).

A second arrangement is available for sectors with inlet metering, in which the measured flow is imposed as a boundary and the pressures alone form the objective. It suits problems in which the total withdrawal is not in dispute and the question is how that total distributes over the network, and it gains information when several inlets deliver flow signals that respond to different parts of the system. In Guanabara it would consume the only series that fixes the level of demand, because a single inlet with imposed flow turns the sum of consumption and leakage into a datum and leaves the demand scales and the leakage term to be separated by four pressures, and pressure monitoring inside a district has been shown not to determine how the loss distributes, the flow measured at the district boundary being what carries that information (Berardi and Giustolisi 2021). A scan that fixes the leakage at successive values and re-estimates the remaining six parameters indicates how little the pressures

constrain that separation, since the calibration pressure RMSE moves by 0.02 m between zero leakage and 2.0 L/s, a third of the mean inflow, while the total cost rises by 42.6. That rise comes from the flow residuals, the leakage prior contributing 0.28 of it, which places the constraint on the separation in the inlet flow and not in the pressures. Keeping the flow as an observation with error preserves its balance content and leaves the disagreement visible in the residual, whereas imposing it would move the same disagreement into the estimated parameters.

4.3 Leakage and service

Leakage estimated as zero should be read as a parameter not required by the formulation, not as a network without losses. Consumption was represented by four blocks and category scales, leakage was distributed by length and varies with the square root of the pressure, and the schedule repeated two levels during the window; the correlations near -0.99 between leakage and the scales show that increasing one share and reducing the other produces similar responses at the sensors. Separating the two shares requires distinct signatures that the available excitation does not provide (Sanz et al. 2016).

The minimum-night-flow decomposition of the turbine telemetry supplies a physical estimate that does not pass through the inversion, obtained from the water balance and the pressure dependence of the loss rather than from the fit (Tabesh et al. 2009). Under a legitimate night demand of 20% of the mean and exponent 0.5, the seven days shared with the loggers give a mean leakage of 0.390 L/s, 6.3% of the inflow, and the whole four-and-a-half-month record gives 0.466 L/s, 7.6%, against a mean inflow of about 6.2 L/s. The night fraction moves the estimate more than the exponent does. Over the shared week, fractions of 0.10 and 0.15 raise the leakage to 1.282 and 0.873 L/s, 20.9% and 14.2% of the inflow, and fractions of 0.24 and above are incompatible with the record, because the night flow is 23.5% of the mean and leaves no room for a larger legitimate share; the full record gives a ratio of 24.1% and the same incompatibility just above 0.24. Doubling the exponent to 1.0 at a fraction of 0.20 raises the full-record estimate to 0.744 L/s, 12.1%, and at a fraction of 0.10 it raises the shared-week estimate to 1.869 L/s, 30.4%. Across the assumptions compatible with the record the physical loss ranges from 6.3% to 30.4% of the inflow on the shared week and from 7.6% to 32.1% on the full record, an amplitude the demand scales can absorb, which is why the inversion required no positive leakage term. The width of that range is set by the assumed fraction of legitimate night demand, which the available records do not measure. Distributed leakage mimics any unexplained head loss, so a residual reduction obtained by freeing it against four pressures does not identify a loss. The flow measured at the inlet, and not the pressure fit, is what arbitrates the loss estimate (AL-Washali et al. 2016; Kang and Lansey 2011; Berardi and Giustolisi 2021).

Pressure-dependent analysis is necessary because part of the domain falls below the 15 m threshold in the night hours. A demand-driven model would impose full consumption and overestimate flows and head losses in the critical hours, whereas PDD makes the feedback between pressure and consumption explicit (Giustolisi et al. 2008; Tabesh et al. 2011). The Brazilian standard ABNT NBR 12218 fixes a minimum

dynamic pressure of 10 m in the distribution network (Araújo et al. 2022), and the 15 m adopted here is a service target above that regulatory floor rather than the floor itself. The required pressure is an assumption about service, not a parameter of the simulator, and the 15 and 20 m scenarios produce a similar sensor RMSE while the fraction of node-hours in deficit rises from 18.7% to more than 52%. A conclusion about delivery must state the adopted threshold and link it to standards and local plumbing conditions, and the 10 m scenario lowers the same fraction to 4.2% while leaving the errors, the rank and the estimated parameters unchanged.

Node-hours below the threshold count a condition and not a volume, and the two indicators separate in this sector. Mean delivered demand stays at 98.9–99.0% of the required demand and falls to 92.1% in the worst hour, so the deficit is shallow and concentrated in the night modulation, when withdrawal is at its lowest, rather than spread through the day. The count is also the indicator most sensitive to the inputs, because the elevations carry the uncertainty of 1 to 2 m stated by the utility and each node aggregates connections at different floor levels, so nodes lying within that margin of the threshold change side without any change in the hydraulics. A service assessment of the sector should fix p_{req} from the 10 m floor of ABNT NBR 12218 and from the plumbing of the buildings served, and should report the delivered fraction next to the node-hour count.

4.4 Limitations

The study considers one sector, a short window and four sensors. Two days inform the fit, as the design proposes, and the 240 hourly observations do not amount to 240 independent realisations because of autocorrelation, so the singular values are computed on a residual set whose effective size is smaller than its nominal size, and the rank should be read as an ordering of the directions rather than as a count of independent constraints. The two days mix a Saturday and a Wednesday, which improves the excitation, and the six validation days follow them inside the same week, so the transfer that was tested is between days and not between seasons or between demand regimes. The set-point policy was constant across the available days, which made it legitimate to use the learned profile as the validation boundary, but did not test transfer to a new control policy.

Roughness was represented by a global multiplier, adequate to the available information and insufficient to locate degraded pipes; leakage was distributed by length and should not be interpreted spatially; the PDD and leakage exponents were fixed; the elevations carry an uncertainty of 1 to 2 m; and the turbine was represented by a PRV only to reproduce the downstream control, without computing head across the runner, efficiency or energy. The information that would narrow the result is known and verifiable. Pressure tests at more than one level, independent consumption measurement and the minimum night flow together with authorised consumption would break the collinearity between demand scales and leakage, and temporary sensors at nodes that respond differently to scale and to leakage would raise the smallest singular value. The correlation matrix provides the agenda: new observations should separate

the weekday and weekend scales, the daytime shapes and the distributed loss, maximising the independence of the Jacobian columns (Lansey et al. 2001; Sanz and Pérez 2015).

5 Conclusions

The work calibrated a real sector controlled by a pressure-regulating turbine, aiming to establish how the inflow measured at the inlet should be used and how far the pressure fit alone supports conclusions on the same data. The measured downstream pressure was imposed as a time-varying head profile through the equivalent PRV, the flow governed the demand level as an observation with error, and the pressures entered the objective function on the two estimation days and were then tested on the six days outside the fit. The model driven in this way reproduced the four pressures with a calibration RMSE of 0.798 m and reduced the validation flow RMSE by 31.9%, and activating the turbine as a reducing valve was the structural condition for that reproduction.

The roughness multiplier settled at 0.808, inside an admissible interval of 0.45 to 1.55, while leakage converged below 10^{-3} L/s, at its lower bound within the resolution of the estimate, with correlations near ± 0.99 with the demand terms. Local rank was 7/7, but consumption and loss were not causally separated. The minimum night flow, extracted from the existing telemetry, places the physical leakage between 6% and 32% of the inflow, the width of that interval coming from the assumed fraction of legitimate night demand and not from the record, an amplitude the demand scales absorb. The conclusion, transferable to other sectors with inlet flow measurement, is that the flow should enter once, as an observation with error that fixes the level of demand and as the basis of a minimum-night-flow estimate of the loss, while the pressures validate the model and locate structural errors, without calibrating leakage against a few sensors. The quality of the pressure fit is not, by itself, evidence of a correct loss estimate, and the information that would resolve the remaining local residual and the formalisation of the identifiability analysis are the paths to narrow the result.

Acknowledgements.

Declarations

- Funding:
- Conflict of interest: the authors declare no competing interests.
- Ethics approval and consent to participate: not applicable.
- Consent for publication: not applicable.
- Data availability:
- Materials availability: not applicable.
- Code availability:
- Author contribution:

References

- AL-Washali T, Sharma S, Kennedy M (2016) Methods of assessment of water losses in water supply systems: a review. *Water Resources Management* 30:4985–5001. <https://doi.org/10.1007/s11269-016-1503-7>
- Araújo JVSd, Villanueva JMM, Cordula MM, et al (2022) Fuzzy control of pressure in a water supply network based on neural network system modeling and IoT measurements. *Sensors* 22:9130. <https://doi.org/10.3390/s22239130>
- Berardi L, Giustolisi O (2021) Calibration of design models for leakage management of water distribution networks. *Water Resources Management* 35:2537–2551. <https://doi.org/10.1007/s11269-021-02847-x>
- Brun R, Reichert P, Künsch HR (2001) Practical identifiability analysis of large environmental simulation models. *Water Resources Research* 37:1015–1030. <https://doi.org/10.1029/2000WR900350>
- Chew AWZ, Wu ZY, Walski T, et al (2022) Daily model calibration with water loss estimation and localization using continuous monitoring data in water distribution networks. *Journal of Water Resources Planning and Management* 148:04022019. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001546](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001546)
- Chu S, Zhang T, Xu C, et al (2021) Dealing with data missing and outlier to calibrate nodal water demands in water distribution systems. *Water Resources Management* 35:2863–2878. <https://doi.org/10.1007/s11269-021-02873-9>
- Díaz S, Mínguez R, González J (2017) Calibration via multi-period state estimation in water distribution systems. *Water Resources Management* 31:4801–4819. <https://doi.org/10.1007/s11269-017-1779-2>
- Dini M, Tabesh M (2014) A new method for simultaneous calibration of demand pattern and Hazen–Williams coefficients in water distribution systems. *Water Resources Management* 28:2021–2034. <https://doi.org/10.1007/s11269-014-0592-4>
- Do NC, Simpson AR, Deuerlein JW, et al (2016) Calibration of water demand multipliers in water distribution systems using genetic algorithms. *Journal of Water Resources Planning and Management* 142:04016044. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000691](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000691)
- Giustolisi O, Savić D, Kapelan Z (2008) Pressure-driven demand and leakage simulation for water distribution networks. *Journal of Hydraulic Engineering* 134:626–635. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2008\)134:5\(626\)](https://doi.org/10.1061/(ASCE)0733-9429(2008)134:5(626))
- Guillaume JHA, Jakeman JD, Marsili-Libelli S, et al (2019) Introductory overview of identifiability analysis: a guide to evaluating whether you have the right type of data for your modeling purpose. *Environmental Modelling & Software* 119:418–432.

<https://doi.org/10.1016/j.envsoft.2019.07.007>

- Gupta R, Nair AGR, Ormsbee LE (2016) Leakage as pressure-driven demand in design of water distribution networks. *Journal of Water Resources Planning and Management* 142:04016005. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000629](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000629)
- Hutton CJ, Kapelan Z, Vamvakeridou-Lyroudia L, et al (2014a) Application of formal and informal Bayesian methods for water distribution hydraulic model calibration. *Journal of Water Resources Planning and Management* 140:04014030. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000412](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000412)
- Hutton CJ, Kapelan Z, Vamvakeridou-Lyroudia L, et al (2014b) Dealing with uncertainty in water distribution system models: a framework for real-time modeling and data assimilation. *Journal of Water Resources Planning and Management* 140:169–183. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000325](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000325)
- Kang D, Lansey K (2011) Demand and roughness estimation in water distribution systems. *Journal of Water Resources Planning and Management* 137:20–30. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000086](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000086)
- Kapelan ZS, Savić DA, Walters GA (2007) Calibration of water distribution hydraulic models using a Bayesian-type procedure. *Journal of Hydraulic Engineering* 133:927–936. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2007\)133:8\(927\)](https://doi.org/10.1061/(ASCE)0733-9429(2007)133:8(927))
- Klise KA, Bynum M, Moriarty D, et al (2017) A software framework for assessing the resilience of drinking water systems to disasters with an example earthquake case study. *Environmental Modelling & Software* 95:420–431. <https://doi.org/10.1016/j.envsoft.2017.06.022>
- Lansey KE, Basnet C (1991) Parameter estimation for water distribution networks. *Journal of Water Resources Planning and Management* 117:126–144. [https://doi.org/10.1061/\(ASCE\)0733-9496\(1991\)117:1\(126\)](https://doi.org/10.1061/(ASCE)0733-9496(1991)117:1(126))
- Lansey KE, El-Shorbagy W, Ahmed I, et al (2001) Calibration assessment and data collection for water distribution networks. *Journal of Hydraulic Engineering* 127:270–279. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2001\)127:4\(270\)](https://doi.org/10.1061/(ASCE)0733-9429(2001)127:4(270))
- McKay MD, Beckman RJ, Conover WJ (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21:239–245. <https://doi.org/10.1080/00401706.1979.10489755>
- Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I — a discussion of principles. *Journal of Hydrology* 10:282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Oreskes N, Shrader-Frechette K, Belitz K (1994) Verification, validation, and confirmation of numerical models in the Earth sciences. *Science* 263:641–646. <https://doi.org/10.1126/science.263.5159.641>

[//doi.org/10.1126/science.263.5147.641](https://doi.org/10.1126/science.263.5147.641)

- Ormsbee LE, Lingireddy S (1997) Calibrating hydraulic network models. *Journal of the American Water Works Association* 89:42–50. <https://doi.org/10.1002/j.1551-8833.1997.tb08177.x>
- Rossman LA, Woo H, Tryby M, et al (2020) EPANET 2.2 user manual. Tech. Rep. EPA/600/R-20/133, US Environmental Protection Agency, Cincinnati, <https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P1010BBK.txt>
- Sanz G, Pérez R (2015) Sensitivity analysis for sampling design and demand calibration in water distribution networks using the singular value decomposition. *Journal of Water Resources Planning and Management* 141:04015020. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000535](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000535)
- Sanz G, Pérez R, Kapelan Z, et al (2016) Leak detection and localization through demand components calibration. *Journal of Water Resources Planning and Management* 142:04015057. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000592](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000592)
- Savić DA, Kapelan ZS, Jonkergouw PMR (2009) Quo vadis water distribution model calibration? *Urban Water Journal* 6:3–22. <https://doi.org/10.1080/15730620802613380>
- Tabesh M, Asadiyani Yekta AH, Burrows R (2009) An integrated model to evaluate losses in water distribution systems. *Water Resources Management* 23:477–492. <https://doi.org/10.1007/s11269-008-9284-2>
- Tabesh M, Jamasb M, Moeini R (2011) Calibration of water distribution hydraulic models: a comparison between pressure dependent and demand driven analyses. *Urban Water Journal* 8:93–102. <https://doi.org/10.1080/1573062X.2010.548525>
- Virtanen P, Gommers R, Oliphant TE, et al (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* 17:261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Walski TM (1983) Technique for calibrating network models. *Journal of Water Resources Planning and Management* 109:360–372. [https://doi.org/10.1061/\(ASCE\)0733-9496\(1983\)109:4\(360\)](https://doi.org/10.1061/(ASCE)0733-9496(1983)109:4(360))
- Walski TM (2000) Model calibration data: the good, the bad, and the useless. *Journal of the American Water Works Association* 92:94–99. <https://doi.org/10.1002/j.1551-8833.2000.tb08791.x>
- Zanfei A, Menapace A, Santopietro S, et al (2020) Calibration procedure for water distribution systems: comparison among hydraulic models. *Water* 12:1421. <https://doi.org/10.3390/w12051421>