EWE

Environment and Water Engineering

Homepage: www.jewe.ir

ISSN: 2476-3683

Carbon-aware hydraulic error correction for scope-2 emissions: A machine learning–Epanet framework for real-time pressure alignment in urban water systems

Suparatchai Vorarat[1], and Surasak Janchai[2]✉

[1,2]College of Engineering and Technology, Dhurakij Pundit University, Bangkok 10210, Thailand

ARTICLE INFO

Paper Type: Research Paper

Received: 23 October 2025
Revised: 11 November 2025
Accepted: 26 January 2026
Published: 11 February 2026

Keywords

Water distribution systems
Hydraulic modeling
Machine learning
Carbon footprint reduction
Urban infrastructure

***Corresponding author:**
Surasak Janchai
✉ Surasak.jan@dpu.ac.th

ABSTRACT

Hydraulic simulations support real-time operational decision-making in urban water distribution systems, yet pressure accuracy can deteriorate under demand variability and uncertain operating conditions. This work introduces a machine-learning (ML) pressure error-correction layer coupled with EPANET to improve short-term pressure agreement while remaining carbon-aware. Here, "carbon-aware" means that corrected outputs are mapped to indicative pumping electricity (kWh) and corresponding Scope-2 CO_2e using fixed, transparent assumptions (pump efficiency and grid emission factor); therefore, emissions estimates are intended for operational interpretation rather than full life-cycle accounting. The method is evaluated using hourly pressure records from four monitoring locations in a tourist-oriented coastal city over seven days (P1–P4; n = 672). Two hydraulic baselines Hazen-Williams and Darcy-Weisbach are developed to examine how friction formulation shapes residual errors and correction behaviour. Results indicate that Hazen-Williams provides a stronger baseline fit to observed pressures than Darcy-Weisbach under monitored conditions. Gradient-boosting correction further reduces short-term pressure errors, with larger gains when correcting the weaker baseline. Because data span only one week and four locations, findings represent proof of concept and require multi-season validation for broader generalisation.

How to ctie this paper: Vorarat, S., & Janchai, S. (2026). Experimental investigation for the effect of the presence of bridge piers on the outer bank erosion of a river's bend. *Environ. Water Eng.*, *12*(1), 37-46.

1. Introduction

Growing urbanisation and population growth are rapidly increasing the demand for safe drinking water, as well as the energy required to produce and deliver it, which cumulatively increase the pressure that urban water distribution systems (WDS) exist under to provide stable pressure, flow and water quality while controlling for costs and environmental impact (Mabrok, et al., 2022). In cities with varied topography, utilities try to maintain adequate service in areas with high-altitude or high-demand areas by increasing pressure upstream, but this can also create excessive pressure downstream, increase the potential for leakage and water main breaks, and increase the energy required for pumping (Shin et al., 2024). Coastal tourist cities are particularly impacted due to the volatility in short-term demand driven by fluctuating guest population during peak tourist seasons, which makes pressure management and energy optimisation difficult (Yin et al., 2021; Ferrari et al., 2023).

When coupled with hydraulic modelling, hydraulic modelling continues to provide operational decision-making support to WDS management (Rossman, 2000). Additionally, while the hydraulic simulation and pressure distribution modelling program EPANET is known for its use in hydraulic analysis, the operational accuracy of the model can decline when key input parameters (demands, boundary conditions, pump/valve settings, and roughness) are uncertain or not constant in time (Rossman, 2000; US EPA, 2025; Dongare et al., 2024). The literature shows that the common methods of representing friction and head loss in hydraulic modelling are by the Hazen–Williams (HW) and Darcy-Weisbach (DW) methods. Depending upon which representation is used as the baseline friction and head-loss model, the resulting pressure error structures will differ significantly between actual operating conditions (Mays, 2019; Liou, 1998; Martínez-Solano et al., 2017).

In order to reduce the discrepancies between the pressure that is simulated and the pressure that is measured, previous research has developed hybrid methods to integrate machine learning (ML) with hydraulic modelling (Zanfei et al., 2020; Minaee et al., 2019; Romero-Ben et al., 2023). In this manner, ML enhances, rather than replaces, a physics-based model (EPANET) by providing an error correction layer where systematic mismatches between what EPANET indicates and what field observations show are resolved through the use of operational conditions.

This process of integrating ML with EPANET modelling is similar to the digital twin concept, in which the hydraulic model is continuously updated with current operational data streams (Torres et al., 2024; Singh et al., 2024; Malekjani et al., 2025). Still, as is indicated in the literature on EPANET-ML calibration and correction, there remain substantial gaps in practical applications of integrating ML methods into existing hydraulic models: (i) Many studies are limited to only focusing on static calibrations or generic predictive performance and have not optimised for short-term operational corrections as would be feasible within reasonable monitoring capabilities; (ii) Comparatively few studies provide for a basic "like-for-like" comparison of HW versus DW for identical field data sets, as the manner in which friction is modelled can influence the correction behaviours; and (iii) Days of short-term pressure inaccuracies can lead to exceptionally high monthly energy consumption and Scope 2 carbon emissions incurred by coastal tourist-driven water systems that are typically characterised by extreme intra-day variation (Mays, 2019; Liou, 1998; Yin et al., 2021; Ferrari et al., 2023).

In this research, the aforementioned gaps have been addressed by providing a carbon-aware hydraulic modelling framework that is focused on deployment and carbon emissions concerns. This hydraulic modelling framework: (1) defines ML as a constrained error-correction process associated with HW and DW baselines, which enables for the application of operational data to correct pressure measurements in a usable way; (2) evaluates the HW and DW baselines concurrently to explicitly delineate how the methods for modelling friction impact upon the structure of the pressure-error relationships, as well as the effectiveness of correcting the errors under the monitored conditions of the two hydraulic modelling methods; and (3) utilises the corrected pressure data as a measure of energy and Scope 2 emissions associated with pumping in order to demonstrate the operational sustainability relevance of correcting for potential short-term inaccuracies in measured pressures (Rossman, 2000; US EPA, 2025; Makropoulos & Butler, 2010). The hydraulic modelling framework will be demonstrated through the use of measured hourly pressures collected from four locations over seven successive days (and including weekends) in Pattaya City, Thailand, a popular coastal tourist city with unusual elevation and demand patterns, where MPressure (m) will be treated as the "truth" reference for this study. Since the evaluation of hydraulic operation is dependent on the temporal nature of the study's operational period, a random 80/20 split of the seven-day data set will be reported as a within-week validation and an additional time-based 80/20 split of the same data collection will be reported as a means to validate the general correction approach in applying the ML process to the seven-day field data set.

As previously mentioned, carbon-aware hydrodynamic operation is provided through modelling by recognising the improvements in accurately predicting hydraulic pressure will result in lower electricity usage and hence lower indirect (Scope 2) CO_2 emissions incurred from pumping the water. However, the proposed carbon-aware hydraulic modelling framework does not intend to provide a complete life-cycle carbon assessment, but provides a transparent method for establishing a relationship between hydraulic modelling accuracy, electric energy consumption and Scope 2 carbon emissions. The success of the framework will ultimately help inform water utilities in their decision making processes and improve their abilities to responsibly utilise energy and carbon emissions impacts in managing their operational activities in the day-to-day manner of operating water distribution systems.

2. Materials and Methods

2.1 Study area

Aim of the study: To research and analyse hydraulic Performance (HP), evaluate the use of real-time systems (RTS) for improving HP, as well as to test the viability of using machine-learning algorithms for HP-related tasks. Design of the study: The study will be undertaken in Pattaya City, eastern Thailand, and follows the fluctuations in water demand associated with tourism-related seasonal and weekend trends. The study area will be characterised by high levels of variability in pressure due to short-term (over days or hours), seasonal, and weekend demand patterns. In addition to being influenced by these factors, Pattaya City also exhibits substantial (up to 80 m above mean sea level) changes in topography over the study area. Therefore, designing hydraulic systems for the Pattaya City distribution system is complicated by the influence of both water demand variability and differences in elevation on available head. The distribution system is centrally controlled and therefore utilises a large trunk main to convey treated water to tanks at various elevations for gravity fed distribution to consumers and additional pumps where required.

2.2 Hydraulic simulation

Hydraulic behaviours were simulated using EPANET v2.2, a widely-used modelling tool for pressure and flow behaviour within a water distribution system. The attributes of the physical network (e.g. pipe length, diameter, roughness, nodal elevation) were validated through field surveys to confirm that the structure of the hydraulic model reflected the actual conditions of the network. Two separate friction-loss formulations (Hazen–Williams and Darcy–Weisbach) were used to create the physics-based baseline pressures needed to apply error corrections throughout the analysis process. While Hazen–Williams is often preferred because of its simplicity and computational efficiency, Darcy–Weisbach provides a more physically general formulation that captures specific effects of fluid properties and different flow regimes; thus, both formulations were used as separate baseline comparison tools in order to investigate differences between the two types of error structures, rather than as a combined hydraulic model.

Outputs generated by EPANET during the simulation run were conducted at an hourly resolution to correlate with the pressure-monitoring time step used to validate the hydraulic model and for future use in machine-learning training.

The Hazen–Williams equation formula used is shown in Equation (1):

$$hf = 10.67 \cdot \frac{L}{C^{1.85} \cdot D^{4.87}} \cdot Q^{1.85} \quad (1)$$

The Darcy–Weisbach equation formula used is shown in Equation (2):

$$hf = f \cdot \frac{L}{D} \cdot \frac{v^2}{2g} \quad (2)$$

Fig.1 Study area and water distribution network in Pattaya City

At each of the four reference locations and by hour, EPANET-simulated pressures, using both HW and DW, were compared to the corresponding field-measured pressures. Hours of data were screened to remove anomalies caused by transient operational problems, sensor failures, or both prior to modeling. For identifying outliers, an interquartile range (IQR) rule was used as in the original method. The difference between the modeled data and the measured data for each location and hour was calculated to serve as training targets (error-correction variables) within a supervised ML framework (Giacomoni & Zechman, 2010). In the interest of methodological transparency, the HW and DW error channels are preserved separately, allowing independent evaluation of the response of each hydraulic formulation to the goal of modeling real-world uncertainty (Liu et al., 2025). The link between hydraulic simulations, corrected data, and carbon accounting is achieved by using a collaborative framework (CACF) incorporating (1) Physics-based EPANET Simulation, (2) Supervised ML Error Correction, and (3) Translation of Corrected Pressure/Flow Effects to Indicate Pumping Energy and Scope-2 Emission Errors. The CACF framework is compatible with SCADA systems and supports a digital twin approach to the theory and practice of hydraulic simulations.

2.3 Machine learning-based model

The machine-learning (ML) component of this work formulates the prediction of pressure from operational data as a correction of systematic error, rather than an unconstrained prediction of pressure, as typically done. The variable to be predicted is defined as the structured deviation of the EPANET-simulated pressures, either from Hazen-Williams (HW) or Darcy-Weisbach (DW) baseline, compared to the actual pressures measured in the field at each location and hour. Because the formulation is constrained by the physics-based simulation layer, this means that models can have a very high goodness-of-fit metric without being expected to be broadly generalisable beyond the observed operational conditions. In addition, to allow for clearer methodological transparency, systematic errors of the simulated pressure generated from HW and DW methods are kept as separate learning baselines and not physically combined into one baseline, so that the corrected pressure outputs learned from the model can be related back to each hydraulic formulation.

The model inputs were selected to represent various physical characteristics that are often collected in a real operating environment, with deployment-oriented correction in mind. Predictors that were included as inputs were: (i) the EPANET baseline (HW or DW), (ii) hourly hydraulic and operational variables that were measured at the same time (arising from the SCADA flow data where applicable), (iii) physical attributes that are fixed and/or change only slowly, which govern pressure levels, and (iv) descriptors of time that identify patterns in operation on a weekly basis (e.g., hour of day, day of week, and weekend flag). Elevation was retained as a key explanatory variable because, as a static head component, it dominates the pressure distribution in a network where topography is highly varied, as is the case in a coastal distribution system.

To generate a realistic decision support step during which the approximate hourly timetable is used for model training and the model inputs are consistent with the operational interpretation of field feedback, field data were aggregated to an hourly average. Continuous monitoring makes it possible to sample rapidly changing demand across a time period of one day and compare operational regime performance (weekday/weekend) and peak/off-peak demand conditions to enable short-term operational feasibility to be established rather than rely on seasonal robustness of the model. To represent deployment methods and eliminate information leakage, the data were evaluated using an 80/20 chronological split. The first 80% of the hourly sequence was the data upon which to build the training model, and the last 20% was the data that formed the hold-out test set. Doing so preserves the temporal order of the data and simulates the way that models will be deployed operationally, where once they are deployed, access to observations beyond the 80% training set will be unavailable. In addition to the chronological 80/20 split, a random 80/20 split was conducted to establish a complementary benchmark and stability check, not as an indicator of generalisation performance. The fixed seed (seed

= 42), which was applied within the random split, ensured reproducibility, and stratification by monitoring locations (four reference points) and day type (weekday/weekend) afforded comparable operating regimes between the training and test datasets. Thus, this benchmark is reported solely to identify an upper bound on model fit for the random dataset while operational conclusions shall be based on the chronological dataset evaluation. To mitigate any potential for optimistic bias and reduce the chances of information leakage within this relatively short-duration operational dataset, a number of procedural controls were established. First, the main evaluation used the strictly chronological 80:20 data split, which precluded the possibility that any future observations could be incorporated into model training; the random 80:20 split functioned solely as an upper-bound week-to-week benchmark, and the random set provided no indication of out-of-sample generalisability. Second, all predictors were restricted to times at which values were available for all variables at that hour (in addition to static attributes, such as elevation), and no features were engineered using information that was collected after that hour. Third, all pre-processing and model selection (e.g., scaling of data, hyperparameter tuning, and selecting the final model) were based solely on the training data set, without applying fitted transformations to the test set. Finally, robustness checks were performed for both weekday and weekend operational regime by examining errors across those regimes and by evaluating residual summaries (MAE/RMSE/Bias) for systematic failures arising from an operational regime or time drift. Procedures were applied uniformly for both HW and DW-based baselines to ensure that observed model corrections to HW and DW methods were indeed due to the learning behaviour of the model rather than being a result of evaluation artefacts.

As a means of disclosing operational regime-based robustness checks, results from the best-fit model (Gradient Boosting Regressor) were reported for the held-out portion of the test dataset (downloaded from the random 80/20 benchmark), stratified by monitoring location and day type. Days with a weekday/weekend split are days 01 to 05 (Monday – Friday) and days 06 to 07 (Saturday and Sunday), and peak demand hours are defined as 06:00 – 09:00 and 18:00 – 21:00, with the remaining hours defined as off-peak. When using these definitions of peak and off-peak hours and invariability, the model fit at weekdays scores MAE = 6.967m; RMSE = 8.219m (n = 86) and weekends with MAE = 6.877m; RMSE = 8.213m (n = 49). When classifying hours by peak demands, the weekday fit results yield MAE = 7.590m; RMSE = 8.950m (n = 42) whereas the off-peak score is MAE = 6.638m; RMSE = 7.863m (n = 93). The regime-level evaluation of the model is provided as a bench-mark comparison of in-week performance only (to avoid any misinterpretation of results); therefore, these results demonstrate that performance degrades substantially under conditions of high demand and volume, corresponding to the operational demand associated with the evaluated regime.

The preceding analysis provides a concurrent contrast to the previous day-long split timetable of 16-hours/8-hours and maintains the same forward movement-through-time evaluation logic. It simulates how a model would operate as users needed to make decisions for operational activity during future hours without having access to future observations. We used multiple regression algorithms for a robust assessment under operational conditions, including Random Forest Regressor, Gradient Boosting Regressor, Support Vector Regression, and Linear-based methods (Dharmarathne et al., 2025; Mays & Tung, 1992). All modelling was accomplished using Python and scikit-learn and hyperparameter tuning was performed using an optimized GridSearchCV with a five-fold cross-validation (Pedregosa et al., 2011; Hutter et al., 2019). The modelling processes included a predictive operational approach that further enhanced the predictive power of the models and was capable also to contribute towards ease of explaining the relationships between a model's inputs and outcomes (Ali et al., 2023). The best evaluating models were compared using mean absolute error (MAE), root mean square error (RMSE) and coefficient of determination (R^2). These metrics are standard means of promoting maximum accuracy in continuous pressure error evaluation for reliably assessing consistency among multiple models and baselines (Hutter et al., 2019), while examining their potential applicability within short-term demand condition changes for un-influenced operational evaluations (Beh et al., 2017). The best model selected based upon model quality (higher RMSE values for lower standard deviations) and better variance accounting (higher R^2 values) to model the data obtained during temporal phases thus contributed towards establishing operational applicability with the noted benchmarking results, while operational conclusions based upon the time-based data evaluation. The highest-performing R^2-based modelling under this fixed random benchmark served to reflect the upper limit of model-fitting capability under quasi-i.i.d. conditions. However, under operational inference, the model used for the evaluation of downstream energy/carbon conversion will function on the basis of time-based data only, while supporting practicality when choosing a final model for operational assessments (Giraldo-González & Rodríguez, 2020).

The aim of optimising corrected pressure outputs is to maximise the alignment of simulated (EPANET-generated) pressures and actual operational pressures. This enables improved situational awareness of the operational and demand-related dynamics associated with leaks and also allows for the timely adjustment of pressure levels for improved stability under challenging operational circumstances (Beh et al., 2017). The overall structure of the work involved the CRISP-DM (Cross-Industry Standard Process for Data Mining) method of using a reproducible format to document workflows for operationally accurate inferences from this project. Each model description is summarised as follows: (1) the baseline pressure data files were created using the two types of friction data from EPANET modelling (HW and DW methods); (2) the model generates corrected EPANET-based pressures by learning the systematic error of the baseline pressure simulation; and (3) the corrected pressures are associated with the energy impacts incurred while pumping (energy transports fluid) and their associated indirect scope 2 CO2 emissions will be based upon using the conversion factor defined within this project: 0.499 $kgCO_2/kWh$.

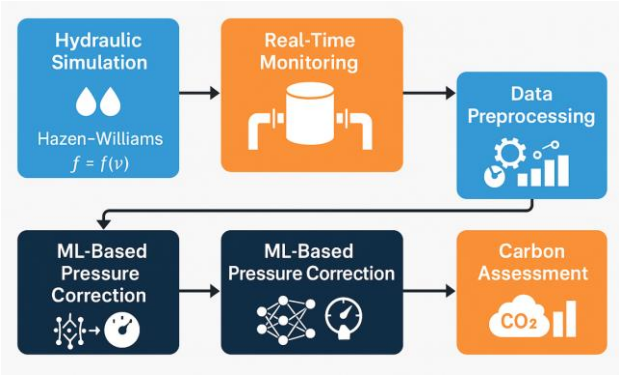

Fig.2 Framework of the proposed machine learning–assisted hydraulic modelling.

2.4 Carbon footprint and energy use

The carbon assessment in this report focuses on the indirect (scope-2) emissions associated with electricity used for pumping, using the same methods as utility-level carbon accounting. The assessment does not consider upstream infrastructure and embodied emissions. The findings of this report demonstrate that improvements in model accuracy can have a significant effect on energy savings and infrastructure-level reductions in carbon emissions (Friedrich et al. 2009; Gómez-Gardars et al. 2022). This report assessed the carbon emissions of a water distribution network in relation to the EFA of the Thai electricity grid (Ke et al., 2022). The coefficient of 0.499 kgCO2e/kWh is the official greenhouse gas emission coefficient set by the Thailand Greenhouse Gas Management Organisation, and it serves as the basis for converting simulated scenarios of electricity use to their corresponding carbon emissions. The calculations for this report include four total steps as follows:

1. Excess head:

$$H_{excess}(t) = P_{model}(t) - P_{measured}(t) \quad (3)$$

Where: P_{model} and $P_{measured}$ are simulated and measured pressures (m). Positive H_excess indicates overestimation.

2. Pump Power:

$$P_{pump}(t) = \frac{\rho g Q(t) H_{excess}(t)}{\eta} \quad (4)$$

Where: ρ = 1000 kg/m3, g =9.81 m/s^2, Q(t) from SCADA (m^3/s), and pump efficiency η=0.70

3. Energy use:

$$E = \sum_t P_{pump}(t) \cdot \Delta t \quad (5)$$

Where: Δt is the 1-hour interval (1.0 h), consistent with the hourly pressure observations; SCADA flow is aggregated to the same hourly resolution.

4. CO_2 emissions:

$$CO_2 = E \; x \; EF \quad (6)$$

The baseline power generation efficiency is defined as EF = 0.499 kg CO 2 e/kWh (Ke et al., 2022), for which there is a limit to the level of uncertainty that can be accounted for in Section 3.3.1. The campaign measured energy use and Scope 2 emissions during 7 days of operation, and the monthly values are derived by scaling from the average daily value across a 30-day period based upon the assumption that all 30 days would be similar regarding short-term operating conditions.To establish a consistent methodology for deriving the operational carbon footprint, the efficiency of the pump and the EF have been fixed for the purposes of estimate reproduction (Efficiency (η) = 0.70; EF = 0.499 kg CO2 e/kWh) in the base case scenario; however, both η and EF are subject to variability that impacts their absolute values on the measured energy / CO2 outputs (i.e., η = 0.60-0.80; EF = 0.40-0.60 kg CO 2 e/kWh). A single-way sensitivity analysis has been performed (Section 3.3.1) to examine the likely ranges of η and EF and to provide estimates of uncertainty in policy interpretation, while the relative comparisons between the two hydraulic and machine-learning assessed scenarios remain consistent based upon the assumptions of the respective scenario attributes.

3. Results

3.1 Model selection

This analysis will use different criteria in addition to including predictive accuracy in the evaluation of machine learning algorithms. The purpose of developing machine learning algorithms in this work is different from simply achieving high predictive accuracy, instead, the long-term goal of this analysis is to provide stable short-term corrections into SCADA workflows with low processing latency and deployable practicality. The hashed-random 80/20 benchmark method will characterise matching week-to-week corrections against multiple operating situations and provide reproducible results that ensure all four monitoring locations and weekday/weekend timeframes are probably represented in both training and testing sets. Given that the dataset covers only one week of operation, the hashed-random benchmark is not considered to be a measure of long-term accuracy; rather, it is an upper boundary marker of the ability to correct inaccuracies over shorter periods of time. A chronological timed benchmark split analysis was also performed and included in Supplementary Materials as an additional method of assessing the robustness of the dataset.

Based on the hashed-random 80/20 benchmark (Table 1), a large number of machine learning algorithms had good predictive accuracy. The machine learning model that appeared to have the highest accuracy was the Gradient Boosting Regressor (GBR) with RMSE = 8.217 m and R^2 = 0.850, and the second-best predictive accuracy was obtained by randomly selecting the Random Forest (RF) models with RMSE = 8.276 m and R^2 = 0.848. Other models that had competitive predictive accuracy were Support Vector Regressor (SVR) with RMSE = 8.335 m and R^2 = 0.846 and Multillear Perceptron (MLP) with RMSE = 8.334 m and R^2 = 0.846. The predictive accuracy of the Extreme Gradient Boosting was somewhat lower, with RMSE = 8.623 m and R^2= 0.835, while the predictive accuracy of the KNN model was the weakest, with RMSE = 9.249 m and R^2 = 0.810. The HW

baseline achieved an R^2 score of 0.842 with an RMSE of 8.426 m already exhibiting strong predictive accuracy, so any improvements made through the use of machine learning should be seen as incremental to this existing physics-based mapping of accuracy rather than as transformations of the machine learning model's accuracy. The choice of the machine learning model was made based on more than just numerical differences in R^2 and RMSE value; selection of the machine learning model was considered a methodological and deployment decision characterised by operationally realistic considerations of model stability, simplicity, and deployability; that is, predictive behaviour could be predicted and that ease of applying and operationally using this technology would exist. There was no formal statistical hypothesis testing or statistical superiority proposed. Rather, the analyses and results were reported as comparative data for machine learning algorithms based on the same evaluation assumption.

3.2 Model performance evaluation

Using both the HydraulicA series of steps were taken to simulate the evaluation of the hydraulic model. These included the development of a comparison for each of the predictive models, creating a timestamp for the data, establishing a method for learning the systematic error, and creating corrective models that were fit to the measured data. The method of evaluating each was the systematic error that the model learned; therefore, it was necessary to find out the difference between the measurements of hydraulic pressure (HW - Darcy-Weisbach) and the measured values. The learning of the error between the two hydraulic baseline values, HW and DW, results in a corrected pressure that can be measured using the same metrics that were developed for the original baseline.

Table 1 summarises test-set performance under the random split.

Model	MAE (m)	RMSE (m)	R^2	MAPE (%)	Bias (m)
HW baseline	6.787	8.426	0.842	9.298	-0.840
DW baseline	10.219	11.949	0.683	12.547	-7.751
DT corrected	6.983	8.443	0.842	9.948	2.204
RF corrected	6.972	8.276	0.848	9.786	1.719
SVR corrected	6.758	8.335	0.846	9.467	0.883
KNN corrected	7.735	9.249	0.810	10.485	1.136
GBR corrected	6.935	8.217	0.850	9.691	1.616
XGB corrected	7.343	8.623	0.835	10.053	1.158
MLP corrected	6.767	8.334	0.846	9.621	1.376

Results of tests conducted using the Principled Random Benchmark are summarised in Table 1. Results of the HW Baseline show good correlation with measurements (MAE=6.787m, RMSE=8.426m, R2=0.842, MAPE=9.298%) and have a small negative bias (Bias=-0.840m); this means on average it was slightly underestimated. The DW Baseline was found to perform significantly worse than the HW Baseline (MAE=10.219m, RMSE=11.949m, R2=0.683, MAPE=12.547%); furthermore, the DW Baseline has a large negative bias (Bias=-7.751m), indicating a pattern of systematic underestimation when compared to the HW Baseline in the observed experiment. Minor improvements in performance were observed across the Corrected Models in relation to the HW Baseline, with the top performing Corrected Models being GBR (MAE=6.935m, RMSE=8.217m, R2=0.850, MAPE=9.691%, Bias=1.616m) and RF (MAE=6.972m, RMSE=8.276m, R2=0.848, MAPE=9.786%, Bias=1.719m). SVR and MLP show similarly competitive performance to GBR and RF Corrected Models (SVR: MAE=6.758m, RMSE=8.335m, R2=0.846, MAPE=9.467%, Bias=0.883m; MLP: MAE=6.767m, RMSE=8.334m, R2=0.846, MAPE=9.621%, Bias=1.376m). A number of Corrected Models also exhibited a positive bias; this raises operational relevance for workers in terms of interpreting benefits of Pressure Control and downstream Energy/CO_2 Reduction.

Overall, the findings indicate that there is potential for Machine Learning-based Corrections to improve within-week Fitting Estimations when mixed operational states were present, particularly when compared to the less valid DW Baseline. Although additional improvements were not witnessed on HW Baselines beyond minor levels, consideration should be given to the effects of bias directionality and limitations of deployment in making interpretations based on this study's data.

3.3 Energy and CO_2 reduction

To assess the energy related to pumps as a result of excess pressures, the excess-head framework developed in Section 2.4 was utilised. Only positive overestimation contributes to excess energy from pumping. Although the monitoring period comprised one operational week, monthly energy and Scope-2 CO_2 figures are given as an indicative estimate for comparing scenarios based upon fixed operational assumptions ($\eta = 0.70$, EF = 0.499 $kgCO_2e/kWh$, and representative daily usage $\approx$ 15,000 m^3/day) without creating a complete emissions inventory. Using the HW baseline for the operational week surveyed, the average positive excess head experienced by this facility is approximately 2.55 m, translating to an indicative electric energy consumption of approximately 4.46 MWh/month; this equates to ~2.23 tCO_2e/month when using the emission factor referenced above. The above figures represent the baseline level of operational inefficiency associated with pressure under the operational assumptions, thus providing a uniform basis for comparing the corrected scenarios.

Although some ML correction models are statistically better (lower RMSE/R^2) than the random benchmark, many of these models also demonstrate positive bias as shown in Table 1. This information is important for operational purposes as the benefits from an electrical energy perspective as described with the excess head framework depend not only upon the magnitude of the total error (RMSE/MAE) but also upon whether or not corrected pressures systematically overestimate

measured values. If excess energy is calculated purely as a result of positive overestimation, then the purpose of that positive bias is to reduce the anticipated energy benefits unless additional operational constraints are implemented (i.e., bias calibration, asymmetric penalty for overestimating loss, or corrective procedure based upon ratio-quantile). It should therefore be recognised that the energy/CO_2 figures shown here are subject to interpretation depending on the application of controls and would be best suited to bias-dependent post-processing for real-world applications.

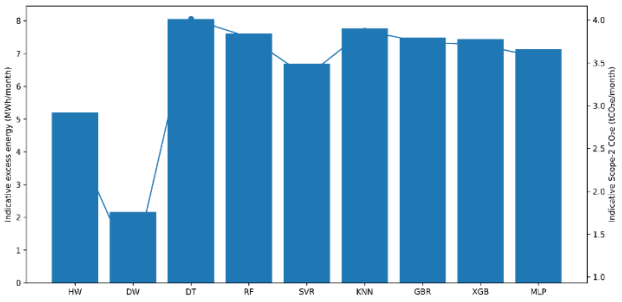

Fig.3 Indicative excess energy & scope-2 CO_2e across models

The summary figure 3 provides indicative excess energy and Scope-2 CO_2 outcomes for each of the model types summarised in Figure 2 under the same assumptions. The improvement of prediction-fit or other metrics of performance do not always equate to proportional improvement in energy efficiency or reduced CO2 emissions, particularly when taking into account bias direction.

3.3.1 Sensitivity of indicative excess energy and scope-2 CO_2

The analysis of operational energy usage and CO_2 emission levels under Scope 2 in Section 4.3 utilizes pump efficiency (η) and grid emissions factor (EF) as parameters to conduct a one-way sensitivity analysis that establishes a base case and provides context for evaluating the data on a policy level and operationally. The excess head formula (Equations 4-6) indicates that for electrical energy estimates, η is inversely correlated with the estimated value, and therefore when estimating the CO_2 emissions at scope 2 scale, EF will have a direct linear correlation with this value. For that reason, uncertainty regarding η will directly affect the overall magnitude of energy used while uncertainty regarding EF will only affect the calculated CO_2 emissions at scope 2. Using the HW baseline indicative excess energy estimate of 4.46 MWh/month as discussed in Section 4.2 as the base case (i.e., η=0.70, EF=0.499 kgCO_2e/kWh), Table 2 presents plausible upper and lower limits for η & EF as well as the resultant value ranges for each scenario. When considered together these factors will modify how much absolute energy is used; however, the conclusion that reducing positive excess head (i.e., inaccurate pressure estimation) will continue to be important operationally as it translates directly into energy savings and associated CO_2 emissions.

Table 2. One-way sensitivity bounds of indicative excess energy and Scope-2 CO_2 under plausible ranges of pump efficiency and grid emission factor (based on the HW baseline).

Case	Pump efficiency, η	Grid EF (kgCO_2e/kWh)	Indicative excess energy (MWh/month)	Scope-2 CO_2 (tCO_2e/month)
Lower-bound (best)	0.80	0.40	3.90	1.56
Base-case	0.70	0.499	4.46	2.23
Upper-bound (worst)	0.60	0.60	5.20	3.12

In examining Table 2, an increase in η by twenty percentage points leads to an increase in the amount of electricity consumed each month (MWh/month) from about 5.20 MWh to about 3.90 MWh per month or approximately 16.7% relative increase to a 12.5 % relative decrease from the base case. Also, changing EF from a lower value of 0.40 to a higher value of 0.60 changes the amount of carbon dioxide (CO_2) generated over the same time period from approximately 1.56 tCO_2 to about 3.12 tCO_2, which also demonstrates the impact of changes in parameter assumptions on CO_2 emissions produced by each baseline. Parameter changes to both hydraulic baseline models and machine learning adjusted scenarios produced consistent measured values across all conditions (including noise), thus allowing for direct comparison.

3.4. Spatial–Temporal Error Distribution

With four distinct monitoring points with heterogenous pressure regimes throughout an entire operational week, the new data set contains considerable levels of variance associated with spatial and temporal aggregation. The findings of the spatial-temporal error analysis suggest: (i) variation exists among locations based upon elevation/zone, (ii) temporal error variation exists, and weekend operation impacts, and (iii) the hydraulic baseline calculated from DW data experiences larger amounts of negative bias across all four hydraulic baselines than does the HW hydraulic baseline under monitored conditions. The primary use of the spatial-temporal Error Plots presented in Figure 4 is for identification of those hydraulic baseline differential patterns that exist structurally (deviation) and for assessment of where ML Correction provides the greatest amount of value, not as universalisation of performance among the four hydraulic baselines. In Figure 4(a), the absolute error for the HW baseline is represented as |Sim_HW − Measured|, and in Figure 4(b), the absolute error for the GBR-corrected case is represented as |GBR − Measured| (best-performing correction applied to the DW baseline). The comparison of these two Figures demonstrates that Correction provides changes in the magnitude and distribution of errors by locations and hours; thereby allowing the evaluation of the model performance at operationally relevant temporal ranges.

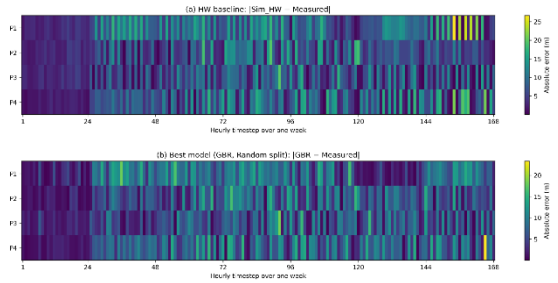

Fig.4 Heatmaps of absolute error for Hazen–Williams (HW) and Gradient Boosting (GBR) models

4. Discussion

Using a coastal urban area as an example, this study shows that by combining machine-learning corrections with hydraulic simulations done in EPANET software, you can significantly improve the accuracy of the short-term operational pressure of a water distribution system. The study used one week of hourly pressure data collected at four different monitoring sites to evaluate the accuracy of using Hazen–Williams (HW) baseline in that area with how closely they align with field pressure readings, yielding strong correlation using baseline HW values (R^2 = 0.842; RMSE = 8.426 m), while applying machine learning-based corrections on HW values resulted in minor incremental improvement when both are assessed using the same evaluation criteria. When using a principle-based 80/20 random-split benchmark (seed = 42), the model that provided the best fit is Gradient Boosted Regression (GBR) models (R^2 = 0.850; RMSE = 8.217), closely followed by Random Forests (RF) (R^2 = 0.848; RMSE = 8.276 m), and SVR and MLP models also showing similar high quality of fit ($R^2 \approx 0.846$). The random-split benchmark is reported only as an upper bound of within-week fit under quasi-i.i.d. assumptions, whereas operational interpretation should primarily rely on chronologically structured evaluation. Results support the theory that by linking a well-defined hydrologic model (in this case, HW data) to a statistical-based approach using machine learning, continual improvement can be made in addressing the systematic short-horizon (i.e., weekly) differences between measured values and modeled hydraulic with reduced variability. However, reliance on a random-split performance metric (using an 80/20 random split of the total number of measurements) may not always represent a long-term or seasonally applicable conversion of water pressure as maximum attainable correction capability—therefore a chronologically-structured evaluation of model output data must also be considered and shown (in added Supplementary Data) for enhancing the robustness of these results. As additional support for this approach and for providing an insight into specific dates/times where excess error occurred, Refer to Figure 4-South Florida Water Management District North Palm Beach use case: North Palm Beach: 50th percentile HW overall mean positive excess head = 2.55 m, or approximately 4.46 mega W/h; 2.23 tons CO_2e/month of excess electricity use under baseline scenarios (efficiency = 70%, emissions factor = 0.499 kgCO_2e/kWh, and representative usage). The stated ranges of energy and emissions values are to be seen as scenario-comparative estimates, and thus their potential for real-world application should be considered evolutionary; improved RMSE/R^2; as such, positive overestimated models are not always going to yield equal returns in terms of energy or carbon reduction, since only positive over-valuation contributes to excess pumping requirements in this methodology. The proposed methodology/model should be viewed as more operationally carbon-informed establishing the linkages from modelled hydraulic to potential energy and CO_2 impact validation across longer durations (and including peak demand conditions, equipment performance faults, etc.) and comparative assessments with historically accepted calibration methods must all be pursued for making general application of the methodology.

5. Conclusion

The results of this research confirm that treating operational correction modelling for measuring hydraulic system performance as a task to estimate pressure from existing (i.e., baselines) hydraulic data will provide a greater short-term level of reliability for pressure estimates when operational data is plentiful. The HW baseline pressure estimates agree closely with field-collected measurement data (R^2=0.842) over one week of hourly monitoring at four reference locations. When compared to the HW baseline, the correction estimates provided by machine learning show continued improvement for each evaluation metric under the same evaluation framework; the GBR model provides the best fit of all 10 models tested (R^2=0.850), while the Random Forest (RF) model provides comparable accuracy (R^2=0.848). In addition to evaluating accuracy metrics, the framework also provides a way for utilities to effectively integrate hydraulic modelling of their networks with a data-driven approach to pressure management.

From the perspective of long-term sustainability, the results demonstrate that there is a direct relationship between continuous overpressure in a hydraulic system and the volume of pumping energy required beyond the amount needed to supply demand. The HW baseline estimate of average excess head for this analysis is approximately 2.55 m. When the estimated excess head is taken together with the fixed assumptions of 70% pump efficiency and 0.499 kg CO_2e/kWh for electricity emissions, the average monthly excess electricity associated with overpressure would be approximately 4.46 MWh/month, or 2.23 tCO_2e/month. The above numbers should be used to provide a basic scenario for comparing available emissions when estimating emissions under the current operational hydraulic modelling environment. Thus, implementing an excess-head-based pressure management model with appropriate operational constraints could potentially support utilities in reducing energy and Scope-2 emissions.

Before implementing an operational correction model in practice, utilities should consider several factors. First, reliable data pipelines and sensing technology (i.e., SCADA or pressure monitoring) should be used to validate both model inputs and outputs. Second, improvements in RMSE and R^2 do not automatically equate to proportional increases in energy or carbon benefits; therefore, the direction of bias from positive overestimation of excess energy should be taken into account

when determining the degree to which the utility could benefit by implementing the model in practice. Third, the ability to generalise the model results from the 1-week validation period to other periods of peak demand, during seasons of peak demand variability, when sensors may have been operating incorrectly, and to other network types will require additional validation. In practice, an ideal implementation approach would start with pilot projects located in the areas of the highest demand, as well as linking the corrected estimates of pressure in real-time operational control strategies that use PRVs and adjusting the operating practices of the associated pumps. Lastly, utilising the framework outlined above provides a standardised, reproducible approach to integrating a combination of hydraulic modelling, machine learning, and operational carbon-informed decision support for the benefit of smart, efficient water systems, in order to achieve efficiencies that are consistent with the United Nations' Sustainable Development Goal (SDG) No. 6 (clean water and sanitation) and SDG No. 13 (climate action).

References

Ali, Y. A., Awwad, E. M., Al-Razgan, M., & Maarouf, A. (2023). Hyperparameter search for machine learning algorithms for optimizing the computational complexity. Processes, 11(2), 349. https://doi.org/10.3390/pr11020349

Beh, E. H. Y., Zheng, F., Dandy, G. C., Maier, H. R., & Kapelan, Z. (2017). Robust optimization of water infrastructure planning under deep uncertainty using metamodels. Environmental Modelling & Software, 93, 92–105. https://doi.org/10.1016/j.envsoft.2017.03.013

Dharmarathne, G., Abekoon, A. M. S. R., Bogahawattha, M., Alawatugoda, J., & Meddage, D. P. P. (2025). A review of machine learning and internet-of-things on the water quality assessment: Methods, applications and future trends. Results in Engineering, 26, 105182. https://doi.org/10.1016/j.rineng.2025.105182

Dongare, P., Sharma, K. V., Kumar, V., & Mathew, A. (2024). Water distribution system modelling of GIS–remote sensing and EPANET for the integrated efficient design. Journal of Hydroinformatics, 26(3), 567–588. https://doi.org/10.2166/hydro.2023.281

Ferrari, E., Verda, D., Pinna, N., & Muselli, M. (2023). Optimizing water distribution through explainable AI and rule-based control. Computers, 12(6), 123. https://doi.org/10.3390/computers12060123

Friedrich, E., Pillay, S., & Buckley, C. A. (2009). Carbon footprint analysis for increasing water supply and sanitation in South Africa: A case study. Journal of Cleaner Production, 17(1), 1–12. https://doi.org/10.1016/j.jclepro.2008.03.004

Giacomoni, M. H., & Zechman, E. M. (2010). A complex adaptive systems approach to simulate urban water resources sustainability. In World Environmental and Water Resources Congress 2010: Challenges of Change (pp. 262–271). ASCE. https://doi.org/10.1061/41114(371)262

Giraldo-González, M. M., & Rodríguez, J. P. (2020). Comparison of statistical and machine learning models for pipe failure modeling in water distribution networks. Water, 12(4), 1153. https://doi.org/10.3390/w12041153

Gómez-Gardars, E. B., Rodríguez-Macias, A., Tena-García, J. L., & Fuentes-Cortés, L. F. (2022). Assessment of the water–energy–carbon nexus in energy systems: A multi-objective approach. Applied Energy, 305, 117872. https://doi.org/10.1016/j.apenergy.2021.117872

Hutter, F., Kotthoff, L., & Vanschoren, J. (Eds.). (2019). Automated machine learning: Methods, systems, challenges. Springer. https://doi.org/10.1007/978-3-030-05318-5

Ke, J., Khanna, N., & Zhou, N. (2022). Analysis of water–energy nexus and trends in support of the sustainable development goals: A study using longitudinal water–energy use data. Journal of Cleaner Production, 371, 133448. https://doi.org/10.1016/j.jclepro.2022.133448

Liou, C. P. (1998). Limitations and proper use of the Hazen–Williams equation. Journal of Hydraulic Engineering, 124(9), 951–954. https://doi.org/10.1061/(ASCE)0733-9429(1998)124:9(951)

Liu, F., Ding, N., Zheng, G., & Xu, J. (2025). Carbon emission prediction and reduction analysis of wastewater treatment plants based on hybrid machine learning models. Environmental Engineering Research, 30(2), 240403. https://doi.org/10.4491/eer.2024.403

Mabrok, M. A., Saad, A., Ahmed, T., & Alsayab, H. (2022). Modeling and simulations of water network distribution to assess water quality: Kuwait as a case study. Alexandria Engineering Journal, 61(12), 11859–11877. https://doi.org/10.1016/j.aej.2022.05.038

Makropoulos, C. K., & Butler, D. (2010). Distributed water infrastructure for sustainable communities. Water Resources Management, 24, 2795–2816. https://doi.org/10.1007/s11269-010-9580-5

Malekjani, N., Kharaghani, A., & Tsotsas, E. (2025). Physics-informed and data-driven neural networks with dimensional and non-dimensional inputs for single-droplet evaporation: Investigating the role of increasing physical complexity in predictive ability. Chemical Engineering Science, 306, 121911. https://doi.org/10.1016/j.ces.2025.121911

Martínez-Solano, F. J., Iglesias-Rey, P. L., Mora-Meliá, D., & Fuertes-Miquel, V. S. (2017). Exact skeletonization method in water distribution systems for hydraulic and quality models. Procedia Engineering, 186, 286–293. https://doi.org/10.1016/j.proeng.2017.03.246

Mays, L. W. (2019). Water resources engineering (3rd ed.). John Wiley & Sons.

Mays, L. W., & Tung, Y. K. (1992). Hydrosystems engineering and management. McGraw-Hill.

Minaee, R. P., Afsharnia, M., Moghaddam, A., Ebrahimi, A. A., Askarishahi, M., & Mokhtari, M. (2019). Calibration of water quality model for distribution networks using genetic algorithm, particle swarm optimization, and hybrid methods. MethodsX, 6, 540–548. https://doi.org/10.1016/j.mex.2019.03.008

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html

Romero-Ben, L., Alves, D., Blesa, J., Cembrano, G., Puig, V., & Duviella, E. (2023). Leak detection and localization in water distribution networks: Review and perspective. Annual Reviews in Control, 55, 392–419. https://doi.org/10.1016/j.arcontrol.2023.03.012

Rossman, L. A. (2000). EPANET 2.0 user manual (EPA/600/R-00/057). U.S. Environmental Protection Agency, Water Supply and Water Resources Division.

Shin, G., Kwon, S. H., Lim, S., & Lee, S. (2024). Utilizing calibration model for water distribution network leakage detection. Engineering Proceedings, 69(1), 105. https://doi.org/10.3390/engproc2024069105

Singh, A., Maheshwari, A., & Singh, S. (2024). Digital twin framework for leakages detection in large scale water distribution systems: A case study of IIT Jodhpur campus. IFAC PapersOnLine, 57(1), 280–285. https://doi.org/10.1016/j.ifacol.2024.05.048

Torres, L., Jiménez-Cabas, J., Ponsart, J.-C., Theilliol, D., Jiménez-Magaña, M. R., & Guzmán, J. E. V. (2024). Pipe roughness calibration approach for water distribution network models using a nonlinear state observer. Results in Engineering, 24, 102713. https://doi.org/10.1016/j.rineng.2024.102713

United States Environmental Protection Agency. (2025). EPANET: Water research. https://www.epa.gov/water-research/epanet

Yin, Y., Lin, G., Jiang, D., Fu, J., & Dong, D. (2021). Multi-scenario simulation of a water–energy coupling system based on system dynamics: A case study of Ningbo City. Energies, 14(18), 5854. https://doi.org/10.3390/en14185854

Zanfei, A., Menapace, A., Santopietro, S., & Righetti, M. (2020). Calibration procedure for water distribution systems: Comparison among hydraulic models. Water, 12(5), 1421. https://doi.org/10.3390/w12051421