Uncertainty-Aware Graph Neural Network Surrogates for Wastewater Digital Twins Using Conformal Prediction

Revin Naufal Alief[1], Imane El-Ghabi[2], Guy Henckens[3], Victoria Degeler[2], Alexander Lazovik[1], and Dilek Düştegör[1]

[1] Bernoulli Institute, University of Groningen, The Netherlands
{r.n.alief, a.lazovik, d.dustegor}@rug.nl
[2] Informatics Institute, University of Amsterdam, The Netherlands
{i.elghabi, v.o.degeler}@uva.nl
[3] Aveco de Bondt, The Netherlands
ghenckens@avecodebondt.nl

Abstract. Sludge accumulation, root intrusion, pump failures, and highly variable inflows continuously alter the hydraulic behavior of urban wastewater systems. Rapid detection and response would require dense real-time monitoring, which is often economically and operationally infeasible. As a result, reliable Digital Twins (DTs) are needed to infer unobserved system states and support risk-aware operational decisions. Artificial Intelligence-based models are deployed inside DTs to support fast state estimation; however, these models typically provide deterministic estimations without quantifying their reliability. This paper presents an uncertainty-aware modeling framework for wastewater Digital Twins by integrating Conformal Prediction (CP) with a Graph Neural Network (GNN). The GNN captures the topological structure of wastewater networks to estimate water levels under varying rainfall conditions with limited sensor availability, while CP produces statistically valid prediction intervals for each estimated node. Results show that the conformal GNN framework achieves target coverage levels of 80%, 90%, and 95% with deviations within about 0.3 percentage points, while producing spatially varying uncertainty intervals. For overflow nodes, empirical coverage ranged between 80% and 93%, supporting risk-aware decision-making at critical locations.

Keywords: Wastewater Networks · Graph Neural Networks · Conformal Prediction · Uncertainty Quantification.

1 Introduction

Urban wastewater systems play a critical role in public health and environmental protection. As urban areas expand and rainfall patterns become more variable due to climate change, wastewater networks operate under increasingly dynamic conditions [6]. At the same time, dense sensor deployment across large-scale wastewater networks is economically and practically infeasible, leaving substantial portions of the system unobserved and making network-wide state estimation challenging [15]. Digital Twin (DT) have therefore emerged as a promising framework for wastewater monitoring and decision support.

Modern DTs integrate physical infrastructure, sensor data, and computational models to enable continuous system awareness and operational decision making. In the water application, such architectures have been applied to both real-time monitoring and predictive control of wastewater systems [4]. Because DTs operate continuously and interact with real-time data streams, their analytics layer requires computationally efficient surrogate models capable of estimating network states faster than repeated hydraulic simulations.

To meet this need, Artificial Intelligence (AI) models have been increasingly applied in wastewater systems for state estimation and prediction. However, most approaches focus on isolated components or temporal forecasting and do not explicitly incorporate the spatial structure of sewer networks [12]. Wastewater systems naturally exhibit graph-structured connectivity governed by pipes and manholes, making Graph Neural Networks (GNNs) well suited for topology-aware state estimation [16,9]. However, existing GNN-based models remain largely deterministic, producing point estimates without quantifying predictive reliability.

For operational DTs, reliability is as important as accuracy. Overconfident predictions may lead to unnecessary interventions [10]. Conformal Prediction (CP) provides a model-agnostic approach for constructing prediction intervals [1] that quantify the reliability of model. Despite its advantages, the integration of CP with GNN-based models in wastewater networks remains largely unexplored. Thus, in this paper, we address this gap by proposing an uncertainty-aware surrogate modeling framework for wastewater DTs that integrates Conformal Prediction with a Graph Neural Network. The main contributions of this work are as follows:

1. We extend graph-based surrogate modeling for wastewater DTs by integrating Conformalized Quantile Regression into a GATRes architecture, enabling calibrated water level estimation intervals in networked hydraulic systems.
2. Through simulation-based evaluation, we show that the proposed conformal GNN framework preserves target coverage levels (80%, 90%, and 95%) across both wastewater networks, supporting the effectiveness of conformal calibration in a transductive graph setting.
3. We analyze the resulting uncertainty behavior from a water engineering perspective, examining its relationship with water level indicator (W_{flood}), especially flood depth, and highlighting the trade-off between coverage, interval width, and spatial variability in predictive reliability.

2 Related Works

Data-driven and AI-based models have been increasingly adopted in wastewater systems to support state estimation and prediction, particularly as computationally efficient alternatives to repeated hydraulic simulations. Many approaches rely on temporal deep learning architectures, including convolutional and recurrent neural networks. For example, hybrid CNN–RNN soft sensors have been proposed for wastewater quality prediction [2], and sequence-based models such as LSTM, GRU, and Transformer architectures have been evaluated for effluent forecasting under varying rainfall conditions [14]. While these methods achieve strong predictive performance, they generally treat wastewater systems as collections of independent time series and do not explicitly incorporate network topology.

To address this limitation, recent studies have leveraged GNNs to integrate spatial connectivity into learning. Graph-based models have been applied to stormwater drainage forecasting [8] and network-wide state estimation in water distribution systems under sparse sensing conditions [13], demonstrating the benefits of topology-aware learning. However, existing GNN-based models for water and wastewater systems remain largely deterministic and do not provide calibrated uncertainty estimates.

More recently, uncertainty quantification has gained attention in water-related machine learning applications. CP provides a model-agnostic framework for constructing statistically valid prediction intervals and has been applied to urban water demand forecasting [7]. Extensions of CP to graph neural networks have recently been investigated in transductive settings, primarily on benchmark citation networks [5]. Despite these advances, the integration of conformal uncertainty quantification with GNN-based models in wastewater networks remains largely unexplored.

3 Methodology

3.1 Pipeline Overview

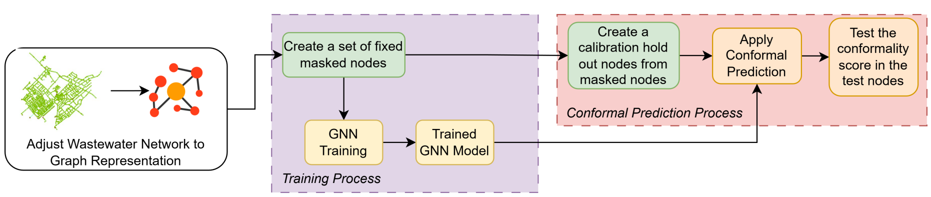


Fig. 1. Integration pipeline of CP with the GNN surrogate.

Figure 1 illustrates the overall pipeline of the proposed framework, which is evaluated using hydraulic simulation outputs from two wastewater networks representing parts of the Dutch Municipality Heusden. We refer to these cases as the smaller network ($N = 247$, $E = 230$) and the medium network ($N = 894$, $E = 968$) [3], where N and E denote the number of nodes and pipes, respectively.

Table 1. Graph representation and GNN input-output setting.

Component	Description
Graph	Nodes represent hydraulic components; edges represent physical connections, mainly pipes.
Masking	30% of nodes are observed sensor locations; 70% are unobserved nodes estimated by the model.
Input	Water-level values at observed nodes; unobserved nodes are masked.
Target	Water level at unobserved nodes, relative to ground level; positive values indicate overflow.
Output	Estimated water level, lower quantile, and upper quantile for each unobserved node.

Each simulation snapshot represents the water-level state of the network at a given time step under a rainfall scenario. The network is represented as a graph, and sparse monitoring is emulated by treating 30% of the nodes as observed sensor locations, while the remaining nodes are treated as unobserved locations where water levels must be estimated.

The GNN surrogate estimates water levels at unobserved nodes from the partially observed network state. Water levels are defined relative to ground level: negative values indicate below-ground water levels, while positive values indicate overflow. After training, conformal calibration is applied to the GNN outputs to construct calibrated prediction intervals, resulting in both point estimates and uncertainty bounds.

3.2 Graph Neural Network Surrogate for Wastewater Systems

The wastewater network is modeled as an undirected graph $G = (V, E)$, where nodes represent hydraulic components and edges represent physical connections. Table 1 summarizes the graph representation, masking setup, input features, target variable, and model outputs used by the GNN surrogate. Let V_{obs} denote the observed sensor nodes and $V_{\text{mask}} = V \setminus V_{\text{obs}}$ denote the unobserved nodes. For conformal calibration, V_{mask} is further partitioned into calibration nodes V_{calib} and test nodes V_{test} using a 10%/90% split.

As the base surrogate, we adopt the Graph Attention Network with Residual connections (GATRes) proposed by Truong et al. [13], which has shown strong performance for network-wide hydraulic state estimation under sparse sensor coverage. The model operates in a snapshot-based manner, where each snapshot represents the network state at a given time step under a rainfall condition. Information from observed water-level nodes is propagated through the graph topology to estimate water levels at unobserved nodes.

To support uncertainty quantification, we extend GATRes following [5]. Instead of producing only a single point estimate per node, the model generates three outputs: a point estimate, a lower quantile estimate, and an upper quantile estimate. The training objective combines the mean squared error of the point estimate with pinball loss terms for the lower and upper quantile estimates:

$$\mathcal{L}_\tau(y, \hat{y}) = \begin{cases} \tau(y - \hat{y}), & \text{if } y \geq \hat{y}, \\ (1-\tau)(\hat{y} - y), & \text{otherwise,} \end{cases} \tag{1}$$

where y denotes the true water level at a given node and time step obtained from the simulator, $\hat{y}$ denotes the predicted water level or quantile estimate, and $\tau \in (0, 1)$ specifies the quantile level.

3.3 Conformal Prediction for Graph-Based Models

Conformal Prediction constructs calibrated prediction intervals through post-hoc statistical calibration. In this paper, calibration refers to adjusting prediction intervals using held-out calibration nodes, not to hydraulic model parameter calibration. Given a trained model and a calibration set, CP aims to achieve a predefined coverage level, meaning that a target proportion of true water-level values should fall inside the predicted intervals. Since graph-based predictions are dependent through network connectivity, we follow Conformalized GNN (CF-GNN) framework [5] and use a transductive setting in which the graph topology remains fixed across training, calibration, and testing, with exchangeability assumed over node-level prediction residuals.

To construct prediction intervals, we use Conformalized Quantile Regression (CQR) [11]. The GNN estimates lower and upper conditional quantiles, denoted by $\hat{q}_{\alpha/2}(x)$ and $\hat{q}_{1-\alpha/2}(x)$. On a separate calibration set, conformity scores are computed as

$$s_i = \max\left\{\hat{q}_{\alpha/2}(x_i) - y_i,\ y_i - \hat{q}_{1-\alpha/2}(x_i)\right\}, \tag{2}$$

where y_i denotes the true water level at node i. The conformal correction (η) is then obtained as the empirical $(1-\alpha)$-quantile of the calibration scores $\{s_i\}$. The calibrated prediction interval for a test input x is defined as

$$C_\alpha(x_i) = \left[\hat{q}_{\alpha/2}(x_i) - \eta,\ \hat{q}_{1-\alpha/2}(x_i) + \eta\right]. \tag{3}$$

These prediction intervals are interpreted as measures of predictive reliability rather than physical uncertainty, reflecting the statistical uncertainty induced by model error and limited observations.

4 Results and Discussions

4.1 Coverage Results

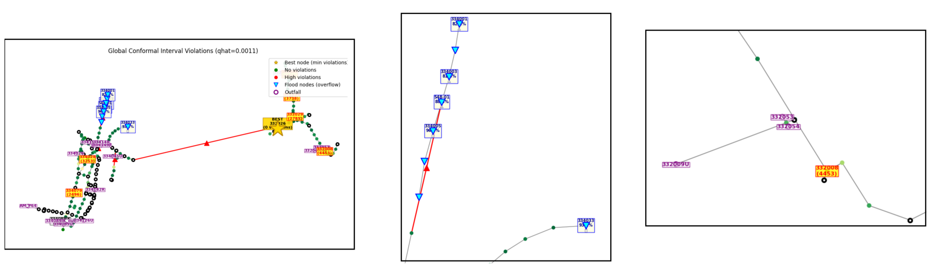
Table 2 summarizes the conformal prediction results for both wastewater networks. Target coverage denotes the desired proportion of true simulated water-level values expected to fall inside the prediction interval, while achieved coverage is the empirical proportion observed on the test nodes. The parameter η represents the water-level reliability margin added to, or subtracted from, the initial prediction interval after evaluating errors on the calibration nodes. $\overline{W}_{\text{flood}}$ denotes the mean interval width over all test snapshots and nodes in V_{test}. Across different target coverage levels of 80%, 90%, and 95% for all simulations, the achieved coverage closely matches the target coverage, with deviations within about 0.3 percentage points in nearly all cases. These results confirm that the proposed CF-GNN framework preserves the desired coverage properties under the transductive graph setting.

Table 2. Conformality score and coverage results on both smaller and medium network.

Network	**Target Coverage (%)**	**Achieved Coverage (%)**	η	$\overline{W}_{\text{flood}}$ **(m)**
Smaller Network	80	79.9	0.0035	0.13
	90	90.2	0.0003	0.34
	95	95.0	-0.0012	0.52
Medium Network	80	79.9	-0.0009	0.21
	90	89.7	0.0200	0.60
	95	95.0	-0.0326	0.77

The values of η remain close to zero for the smaller network, indicating that the initial quantile intervals are already well calibrated. In the medium network, the larger positive or negative η values indicate that CQR applies stronger interval adjustment, suggesting greater calibration difficulty under increased network complexity. The $\overline{W}_{\text{flood}}$ increases with higher target coverage, reflecting the expected trade-off between reliability and precision. The medium network consistently requires larger $\overline{W}_{\text{flood}}$ values than the smaller network, suggesting that increased network size and hydraulic complexity lead to higher uncertainty.

4.2 Network Closer Look



(a) (b) (c)

Fig. 2. Estimation reliability at the 90% target coverage level in the smaller network area. (a) spatial reliability overview, (b) overflow nodes, and (c) lowest-reliability node.

Figure 2(a) shows the spatial distribution of prediction reliability at the 90% target coverage level. Reliability varies across node locations rather than being uniformly distributed, highlighting the value of network-aware uncertainty quantification. Figure 2(b) focuses on the five identified overflow nodes, where empirical coverage ranges between 80% and 93%, indicating that the reliability of surrogate predictions differs across critical locations. Figure 2(c) shows the lowest-reliability node, which is located near the outfall and connected to a velocity-changing pump, suggesting that local hydraulic complexity may reduce model reliability. Such localized diagnostics can support targeted sensor placement or model refinement.

5 Conclusion and Future Works

This paper presented an uncertainty-aware modeling framework for wastewater DTs by integrating CQR with a GNN under sparse sensing conditions. Simulation-based evaluations on two wastewater networks demonstrate that the proposed approach consistently achieves target coverage levels (80%, 90%, and 95%) with minimal deviation, validating the effectiveness of conformal calibration in a transductive graph setting. The flood-specific interval width metric highlights the trade-off between prediction reliability and interval size, and shows that larger and more complex networks require wider intervals to maintain the desired coverage. The lowest prediction reliability is observed at the manhole near the outfall connected to a pump, suggesting that more complex hydraulic conditions may reduce model reliability in this area. Overall, these findings enhance the reliability and interpretability of graph-based models within wastewater DTs, supporting risk-aware decision-making under limited observability. The current surrogate is snapshot-based and does not explicitly model temporal dynamics or flow variables. Future work will extend the framework to temporal graph models for joint estimation of water levels and flows, validate the approach on larger and more hydraulically complex urban drainage networks, and investigate adaptive sensor placement strategies informed by spatial uncertainty patterns.

Acknowledgments. Research funded by the Dutch Research Council NWO Merian Fund - Cooperation China (CAS) program, through the project DDTClean: Intelligent Wastewater Treatment - Distributed Digital Twin for Clean Water (482.22.007).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Angelopoulos, A.N., Bates, S.: A gentle introduction to conformal prediction and distribution-free uncertainty quantification (2022), arXiv:2107.07511

2. Cao, J., Xue, A., Yang, Y., Cao, W., Hu, X., Cao, G., Gu, J., Zhang, L., Geng, X.: Deep learning-based soft sensor for microbial wastewater treatment efficiency prediction. Journal of Water Process Engineering **56**, 104259 (2023). https://doi.org/10.1016/j.jwpe.2023.104259
3. Ghabi, I.E., Alief, R.N., Düstegör, D., Henckens, G., Lazovik, A., Degeler, V.: Gnn-based state estimation for urban wastewater digital twins. In: Proceedings of the ACM/SIGAPP Symposium on Applied Computing (2026), to appear
4. Ghorbani Bam, P., Rezaei, N., Roubanis, A., Austin, D., Austin, E., Tarroja, B., Takacs, I., Villez, K., Rosso, D.: Digital twin applications in the water sector: A review. Water **17**(20), 2957 (2025). https://doi.org/10.3390/w17202957
5. Huang, K., Jin, Y., Candès, E., Leskovec, J.: Uncertainty quantification over graphs with conformalized graph neural networks. In: Advances in Neural Information Processing Systems. vol. 36 (2023)
6. Inbar, O., Shahar, M., Avisar, D.: Predictive modeling of bod throughout wastewater treatment: A generalizable machine learning approach for improved effluent quality. Environmental Science: Water Research & Technology **10**(10), 2577–2588 (2024). https://doi.org/10.1039/d4ew00111g
7. Iwakin, O.M., Moazeni, F.: Improving urban water demand forecasting using conformal prediction-based hybrid machine learning models. Journal of Water Process Engineering **58**, 104721 (2024)
8. Li, M., Shi, X., Lu, Z., Kapelan, Z.: Predicting urban stormwater drainage system states using graph-wavenet. Sustainable Cities and Society **115**, 105877 (2024). https://doi.org/10.1016/j.scs.2024.105877
9. Li, Z., Liu, H., Zhang, C., Fu, G.: Real-time water quality prediction in water distribution networks using graph neural networks. Water Research **250**, 121018 (2024)
10. Liu, S., Lu, D., Painter, S.L., Griffiths, N.A., Pierce, E.M.: Uncertainty quantification of machine learning models to improve streamflow prediction. Frontiers in Water **5**, 1150126 (2023). https://doi.org/10.3389/frwa.2023.1150126
11. Romano, Y., Patterson, E., Candès, E.J.: Conformalized quantile regression. Advances in Neural Information Processing Systems **32** (2019)
12. Sun, W., Gao, Y., Zhou, J., Shah, K.J., Sun, Y.: Artificial intelligence in wastewater treatment systems: Recent developments and future directions. Water **17**(16), 2432 (2025). https://doi.org/10.3390/w17162432
13. Truong, H., Tello, A., Lazovik, A., Degeler, V.: Graph neural networks for pressure estimation in water distribution systems. Water Resources Research **60**(7) (2024). https://doi.org/10.1029/2023WR036741
14. Voipan, D., Voipan, A.E., Barbu, M.: Evaluating machine learning-based soft sensors for effluent quality prediction under variable weather conditions. Sensors **25**(6), 1692 (2025). https://doi.org/10.3390/s25061692
15. Wang, S., Xu, K., Zhou, Y.: Cost-effective sensor placement optimization for large-scale urban sewage surveillance. Sustainable Cities and Society **103**, 105250 (2024). https://doi.org/10.1016/j.scs.2024.105250
16. Xing, L., Sela, L.: Graph neural networks for state estimation in water distribution systems. Journal of Water Resources Planning and Management (2022)