GNN-Based State Estimation for Urban Wastewater Digital Twins

Imane El Ghabi*
Informatics Institute, University of Amsterdam
The Netherlands
i.elghabi@uva.nl

Revin Naufal Alief*
Bernoulli Institute, University of Groningen
The Netherlands
r.n.alief@rug.nl

Dilek Düstegör
Bernoulli Institute, University of Groningen
The Netherlands
d.dustegor@rug.nl

Guy Henckens
Aveco de Bondt
The Netherlands
ghenckens@avecodebondt.nl

Alexander Lazovik
Bernoulli Institute, University of Groningen
The Netherlands
a.lazovik@rug.nl

Victoria Degeler
Informatics Institute, University of Amsterdam
The Netherlands
v.o.degeler@uva.nl

Abstract

Urban wastewater systems are critical infrastructures whose monitoring is complicated by sparse sensing, stochastic inflows, and complex hydraulic behavior. Physics-based hydraulic models provide accurate simulations but are computationally demanding and unsuitable for real-time operation. We adapt a graph neural network surrogate model GATRes to wastewater networks and train it on synthetic data generated from calibrated hydraulic simulations. The approach is evaluated on two wastewater systems: a municipal network in the Netherlands and the open Shunqing stormwater dataset from China. In the Dutch network, the model achieves estimation errors within 20–50 cm across multiple rainfall scenarios, which aligns with the precision expected by water domain experts. Results demonstrate strong overall accuracy under sparse sensing and across unseen rainfall scenarios, while highlighting challenges near hydraulic structures such as pumps and outfalls, where errors are higher but remain within acceptable ranges. These findings show that GNN surrogates can provide an accurate and scalable approach for state estimation in wastewater networks, contributing to monitoring and operational decision-making in smart cities.

CCS Concepts

• **Computing methodologies** → **Neural networks**; *Modeling and simulation*; • **Applied computing** → *Environmental sciences*; • **Information systems** → Decision support systems.

Keywords

Graph Neural Networks, Surrogate Modeling, Wastewater Networks, State Estimation, Smart Cities, Digital Twin

ACM Reference Format:

Imane El Ghabi, Revin Naufal Alief, Dilek Düstegör, Guy Henckens, Alexander Lazovik, and Victoria Degeler. 2026. GNN-Based State Estimation for Urban Wastewater Digital Twins. In *The 41st ACM/SIGAPP Symposium on Applied Computing (SAC '26), March 23–27, 2026, Thessaloniki, Greece.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3748522.3779823

*Imane El Ghabi and Revin Naufal Alief contributed equally to this work.

SAC '26, Thessaloniki, Greece

ACM ISBN 979-8-4007-2294-3/2026/03
https://doi.org/10.1145/3748522.3779823

1 Introduction

Access to clean water and sanitation is critical for ensuring health, food security, livelihoods, industry, gender equality, education, and environmental sustainability [1]. Yet, the ongoing global water and sanitation crisis threatens these objectives, as poor water management amplifies risks across all aspects of life. According to the UN, over 80% of wastewater (and up to 95% in some of the least developed countries) is discharged into the environment without treatment or reuse [4]. Recognizing the severity of this issue, access to clean water and sanitation was established as one of the 17 Sustainable Development Goals (SDGs) in September 2015, targeting fair access to safe, high-quality potable water, good sanitation, and proper hygiene for everyone by 2030 [1]. One key pathway to achieving this target is the development of efficient urban wastewater systems.

Urban wastewater systems are essential components of smart cities and critical infrastructures, where reliable monitoring is necessary to reduce risks of flooding, combined sewer overflows, and infrastructure failure. In smart city contexts, these systems must operate efficiently under various conditions while maintaining service reliability. Yet, despite their importance, water systems are often inefficient, with 10–40% lost through leaks and only 1.7% of water being reused, making monitoring and management even harder [10]. Achieving reliable monitoring remains difficult due to three main challenges: (i) sensor coverage is sparse because of cost and maintenance constraints, (ii) inflows are highly stochastic and vary with weather, and (iii) flow dynamics are shaped by complex hydraulic interactions involving pumps, weirs, and outfalls. These challenges make it difficult for operators to obtain a complete and timely picture of system states such as flood depths and flows, limiting their ability to make informed decisions for infrastructure management and emergency response.

Physics-based hydraulic modeling platforms such as InfoWorks Integrated Catchment Modeling (*Infoworks ICM*) or Storm Water Management Model (*SWMM*) are widely used by municipalities to simulate and manage urban drainage. While they provide accurate estimations when carefully calibrated, they are computationally demanding and unsuitable for real-time decision support [7]. This creates a gap between the need for scalable monitoring of wastewater networks and the practical limitations of physics-based tools.

Digital Twins (DTs) have been proposed as a promising paradigm to address such challenges. One way to enable DTs is by coupling physical systems with data-driven models that support continuous monitoring, estimation, and control [18]. However, developing DTs for urban wastewater systems requires reliable and efficient surrogate models that can estimate complete network states from sparse sensor observations in real time.

Data-driven surrogate models have recently emerged as an efficient alternative. By learning from synthetic or historical data, surrogates can provide rapid estimates of network-wide states from limited sensor inputs [7]. Graph neural networks (GNNs) are particularly suited for this task because wastewater networks are inherently graph-structured, with nodes (e.g., manholes, pumps, outfalls) connected by links (e.g., pipes, conduits). Prior studies have demonstrated the effectiveness of GNNs in water distribution systems [11, 16]. In particular, *GATRes*, a physics-aware graph attention residual network, has shown strong pressure-estimation performance under sparse sensing via robust training with random sensor placement [16]. However, wastewater networks present unique challenges: they are gravity-driven, subject to stochastic inflows, and prone to overflow events. These properties make wastewater modeling a distinct problem that requires dedicated solutions.

In this paper, we adapt the GATRes model to urban wastewater networks and evaluate it on data generated from calibrated hydraulic simulations. This represents the first application of GATRes to wastewater systems, extending its proven capabilities from water distribution networks to the distinct challenges of gravity-driven drainage systems. Our main case study is the Heusden network in the Netherlands, developed in collaboration with the local municipality. The model is further tested on the open Shunqing stormwater dataset from Sichuan, China [19], enabling validation across different infrastructures and contexts.The contributions of this paper are as follows:

(1) **GNN-based surrogate for wastewater state estimation:** We develop a deployable learning framework for *gravity-driven* wastewater systems using the GATRes architecture. The pipeline introduces (i) a graph-snapshot representation derived from hydraulic simulations, (ii) an outfall-value completion step to ensure full graph connectivity, and (iii) masking-based training to emulate sparse sensing and enable virtual sensing at non-instrumented nodes.
(2) **Expert-validated estimation accuracy:** The proposed surrogate achieves estimation errors within 20–50 cm across multiple rainfall scenarios in the Heusden network. This range was validated by domain experts as operationally sufficient for decision-making in smart wastewater management.
(3) **Generalization across heterogeneous wastewater systems:** We evaluate the approach on two distinct networks, the Heusden wastewater system (modeled in InfoWorks ICM) and the Shunqing stormwater system (modeled in SWMM). Consistent results across differing topologies, flow regimes, and simulation platforms confirm the model's robustness and external validity.

This work establishes a validated surrogate modeling framework for state estimation in wastewater systems. By combining physics-based simulations with graph-based learning, it advances data-driven digital twins for urban water management.

2 Related Work

Research on digital technologies for smart cities spans transportation, energy grids, and water distribution networks, with data-driven models enabling resilient and adaptive monitoring. For wastewater systems as critical infrastructures, recent approaches cluster into four directions: (1) simplifying physics-based simulations, (2) leveraging machine learning for predictive analytics, (3) applying graph-based deep learning to capture network-wide dynamics, and (4) integrating Digital Twins (DTs) for decision making.

(1) Physics-Based Surrogates and Hybrid Models: Reducing Calibration Burden. To address the computational cost of detailed urban drainage simulations, Meirlaen et al. [12] introduced mechanistic surrogate models, simplified yet physically consistent representations calibrated on virtual data from high-fidelity simulations. Likewise, Vaes et al. [17] used non-linear reservoir models, calibrated against hydrodynamic references, to assess combined sewer overflow (CSO) impacts on receiving waters, showing that simpler, well-calibrated models can outperform highly complex ones when drivers (e.g., rainfall) are uncertain.

(2) Machine Learning for Rapid Process Estimation: Bridging Data Gaps. Zhao et al. [20] estimate readily and slowly biodegradable COD in municipal wastewater using ORP-based features with ensemble learning, demonstrating accurate surrogate sensing for biological processes. Inbar et al. [9] develop a generalizable BOD estimation model across 30 plants using routinely measured water-quality variables, illustrating how ML can enable real-time control without explicit mechanistic modeling, though interpretability and robustness under distribution shifts remain challenges.

(3) Graph-Based Deep Learning: Capturing Network Topology and Nonlinear Dynamics. Recent work employs Graph Neural Networks (GNNs) to exploit infrastructure topology. Chen et al. [3] propose a dynamic attention GCN for early sludge bulking diagnosis, and Baskar et al. [2] design a node-level capsule GNN for effluent COD in papermaking wastewater treatment, both focusing on plant-level subsystems. At the network scale, Li et al. [11] use a gated GNN for real-time water-quality estimation in water distribution networks (WDNs) under sparse sensing. Truong et al. [16] address WDN pressure estimation with a physics-aware GNN and robust random sensor-placement training; their approach was tested within a DT context for a Dutch WDN [6]. In contrast, urban wastewater (gravity-driven, overflow-prone) remains comparatively underexplored at full-network scale.

(4) Digital Twins for Water Infrastructure: Toward Graph-Based Integration. DT platforms increasingly combine IoT data streams with AI for forecasting and decision support. Homaei et al. [8] outline a DT architecture for water systems that integrates time-series learners (e.g., LSTM, gradient boosting) to support operational use cases. Complementarily, Ngo et al. [13] survey how GNNs can empower DTs by modeling structural dependencies across assets. Beyond urban drainage, Roudbari et al. [15] couple GNN-based

flood forecasting with interactive visualization to inform policy and emergency planning.

Collectively, physics-based surrogates reduce simulation cost while preserving interpretability; ML bridges sensing gaps for fast estimation; and GNNs capture spatial dependencies essential for system-wide modeling. Yet most studies target treatment subsystems or potable WDNs, leaving an opportunity to extend graph-based surrogates to *wastewater* networks at city scale.

3 Proposed Pipeline

The overall pipeline architecture and methodology used in this study, from dataset generation to model design and training, are shown in Figure 1. The topology of the system (e.g., pipes, manholes) alongside static and dynamic parameters such as roughness, water levels, and infiltration rates, serve as inputs to the pipeline. These elements are fed into physics-based modeling platforms (e.g., *InfoWorks ICM*, *SWMM*), comprehensive tools for urban drainage and water-network modeling [21, 22], which generate a synthetic dataset designed to span a wide range of the system's hydraulic behaviors and operational conditions. By covering both typical and extreme scenarios, this dataset captures the relationships between inputs and outputs, and the resulting system states in each scenario are used as independent snapshots for spatial state estimation.

Following the simulation, the dataset is preprocessed to clean, normalize, and format the data into graph snapshots that represent the state of the network at different timesteps ($T_1, T_2, \ldots, T_n$). These snapshots are used as inputs to a GNN training module, where a masking strategy is applied to hide certain node values (e.g., water levels) during training. The GNN learns to infer or reconstruct the masked values based on the known parts of the graph. The model's estimations are then compared to the ground truth (clear target), and the loss is minimized iteratively. The performance is finally evaluated through model validation, completing the pipeline for robust and generalizable spatial state estimation in urban water networks.

4 Implementation on a Use Case

We implement the pipeline on two case studies: (i) the Heusden wastewater network (The Netherlands), simulated with InfoWorks ICM; and (ii) the Shunqing stormwater network (Sichuan, China), modeled in SWMM using the dataset from [19]. For the Heusden dataset, we generate synthetic hydraulic states under multiple rainfall scenarios and parameter variations. For the Shunqing case, we use the dataset provided in [19], which is based on observed rainfall events. In both settings, the data are preprocessed into graph snapshots for GNN training, with separate train/validation/test splits and model instances. We then describe the state-estimation GNN, including inputs, architecture, and training procedure.

4.1 Dataset

4.1.1 Heusden. For the first dataset, this paper creates a synthetic yet realistic dataset that captures the dynamic behavior of a wastewater system under varying conditions, enabling robust training and evaluation of GNN models for water-level estimation using *InfoWorks ICM*. The wastewater network model is based on the infrastructure from the Heusden municipality in the Netherlands,

Table 1: Structural details of the network within Heusden, the Netherlands.

Nodes		***Links***	
Entities	***Amount***	***Entities***	***Amount***
Manholes	873	Pipes	928
		Channels	14
Storages	9	Orifices	1
		Pumps	13
Outfalls	12	Weirs	11
		Flap Valves	1
Total Nodes	894	***Total Links***	968

including key elements such as outfalls, manholes, pipes, and pumping stations. The model captures the structural and spatial complexity of a real-world urban drainage system, making it a suitable basis for testing data-driven estimation methods. The structural details of the network are summarized in Table 1.

To ensure the dataset captures a wide range of system responses, we simulate multiple rainfall events following classifications from the Dutch technical guidelines, specifically Rainfall Events #01 to #10 [14]. These events vary in intensity and duration, allowing us to assess model performance under different hydraulic loading conditions. Additionally, we vary several parameters commonly used in hydraulic model calibration tasks, as listed in Table 2, following guidance from [5]. Other parameters, such as pipe diameters, slopes, and network connectivity, are configured in consultation with domain experts in water engineering.

The simulated outputs consist of water-level values at various nodes, which serve as the target variables for our state-estimation model. Data are recorded at 5-minute intervals to capture short-term variations in water levels. Through the variety of rainfall events, the dataset includes both normal and extreme system conditions, providing a comprehensive foundation for training and evaluation.

Although the simulations produce time-series data, each timestep is treated as an independent system state for model training. Using different rainfall events and the parameter variations listed in Table 2, we generate a total of 255 unique *InfoWorks ICM* simulation scenarios, resulting in 19,320 different snapshots.

To evaluate generalization across hydraulic scenarios, we split the dataset by rainfall event ("bui") index, ensuring each event is assigned exclusively to either training (9,660 states), validation (3,864 states), or testing (5,796 states) without any snapshot overlap.

Each rainfall scenario (Bui#01–Bui#10) represents an actual rain event in The Netherlands characterized by a specific *return period*, which indicates how often an event of that intensity is expected to occur. Based on standard hydrological classifications [14], we assign the following intensity categories:

- **Low intensity:** Bui#01–Bui#04 (0.25–0.5 year return period)
- **Medium intensity:** Bui#05–Bui#06 (1 year)
- **High intensity:** Bui#07–Bui#08 (2 years)
- **Extreme intensity:** Bui#09–Bui#10 (5 and 10 years)

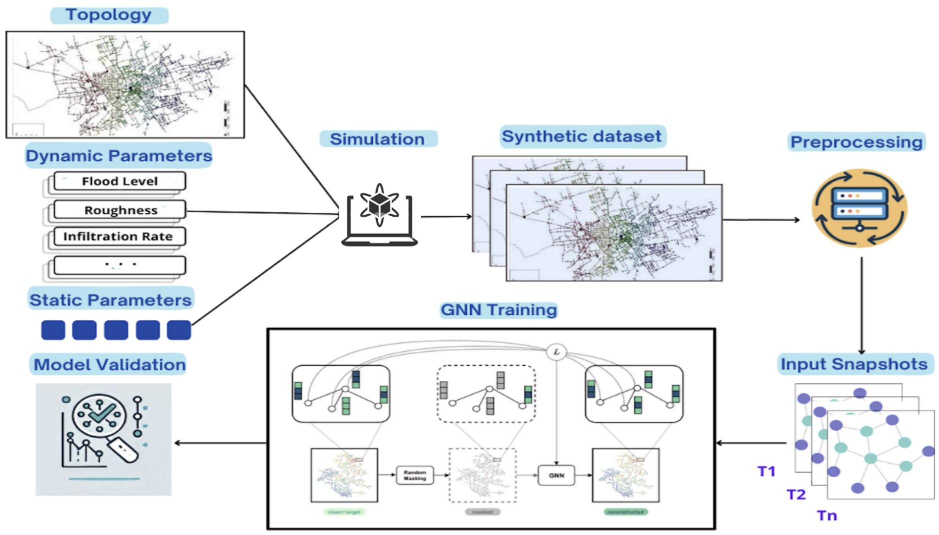

Figure 1: Proposed pipeline of GNN-based state estimation for Urban Wastewater Digital Twin.

Table 2: Scenarios parameter values.

Parameters	Values
Bottom Roughness	Ranged between 2.0 mm and 8.0 mm
Top Roughness	Ranged between 2.0 mm and 8.0 mm
Discharge Coefficient	Ranged between 0.2 and 2.0
Initial Infiltration	Ranged between 2 mm/h and 20 mm/h

The parameter ranges in Table 2 were defined in consultation with water engineering experts to ensure hydraulically realistic conditions. Using this classification, we define our data split as follows:

- **Training:** Bui#01 (low), Bui#03 (low), Bui#05 (medium), Bui#07 (high), Bui#09 (extreme)
- **Validation:** Bui#04 (low), Bui#06 (medium)
- **Testing:** Bui#02 (low), Bui#08 (high), Bui#10 (extreme)

This split provides a balanced representation of rainfall intensities across all subsets while maintaining strict separation between scenarios, allowing us to test how well the model generalizes to unseen hydraulic conditions.

4.1.2 Shunqing. For the second case study, this paper uses a real-world urban drainage network from the Shunqing district in Sichuan Province, China, to validate the GNN-based hydraulic estimation model under authentic operational conditions [19]. The network is modeled in the SWMM, a widely used physics-based simulator for urban drainage systems. The model accurately represents the city's stormwater infrastructure, with structural details provided in Table 3. This established model provides a high-fidelity environment for generating realistic hydraulic data. The dataset is generated by simulating the network's response to 148 distinct, real-world rainfall events monitored on-site. These events vary significantly in duration, ranging from 6 to 24 hours, and capture a wide spectrum of meteorological conditions representative of the region's climate. The Shunqing scenarios are based entirely on these observed rainfall events, with the SWMM model parameters held constant. This approach allows for evaluating the model's performance against realistic, historically recorded conditions. The SWMM simulations were run at a high temporal resolution, with hydraulic states recorded every 1 minute. The target variable for the model is the nodal flood volume (measured in m^3), which represents the volume of water exceeding the capacity of a given node at each timestep. This high-resolution data allows for a detailed analysis of the system's dynamic flooding behavior. From these simulations, a comprehensive dataset of 10,522,560 hydraulic snapshots was generated. A key characteristic of this real-world dataset is its sparsity; only 21% of the snapshots (2,186,206) correspond to active flooding events (i.e., non-zero flood volumes), making the estimation task more challenging and realistic. To ensure a robust evaluation of the model's generalization capabilities, the 148 rainfall events were chronologically split into training, validation, and testing sets. The first 70% of the events were allocated for training, the next 15% for validation, and the final 15% for testing. This chronological split ensures that the model is evaluated on its ability to estimate unseen events, which is critical for real-world deployment. Unlike the Heusden dataset, the Shunqing dataset is based on observed rainfall events rather than systematic parameter exploration.

Table 3: Structural details of the network within Shunqing, China.

Nodes		*Links*	
Entities	***Amount***	***Entities***	***Amount***
Junctions	105	Conduits	131
Outfalls	8		
Total Nodes	113	***Total Links***	131

4.2 Pre-processing

Simulation exports are handled per dataset: the Heusden model produced by *InfoWorks ICM* [21, 22] and the Shunqing model produced by SWMM [19] are each transformed into graph snapshots for separate GNN training. In both cases, every record in the export corresponds to the hydraulic state of one network element at a single timestep. We apply two preprocessing steps, with minor dataset-specific adjustments where needed.

Outfall-level imputation. For the Heusden dataset, *InfoWorks ICM* does not compute valid values for the water level at outfall nodes, which appear as `NaN` or negative entries. To avoid disconnecting parts of the graph, missing values are completed using a k-nearest-neighbor ($k = 5$) imputation based on nearby nodes, and a binary mask is retained to mark imputed entries. The Shunqing dataset, in contrast, provides complete hydraulic states directly from SWMM simulations and does not require this step.

Feature normalisation. For Heusden, the target is **flood depth**, which accounts for terrain elevation. It is computed as follow:

$$\text{FloodDepth}(t) = \text{Level}(t) - \text{FloodLevel}. \tag{1}$$

For Shunqing, the target variable is **cumulative flood volume** as provided in the dataset [19]. For both datasets, values are standardised to zero mean and unit variance using statistics from the training split only, which are reused during validation and testing to avoid leakage.

Terminology (Stormwater vs. Wastewater). In urban drainage, we distinguish (i) *stormwater* networks that convey rainfall–runoff from streets and roofs to receiving waters, with intermittent, event-driven flows; (ii) *wastewater* (sanitary) networks that convey domestic/industrial sewage to treatment, with continuous diurnal flows; and (iii) *combined* systems that carry both. In this paper, the Shunqing case is a stormwater drainage model (runoff only), whereas the Heusden case is an combined wastewater and stormwater drainage system with wet-weather response. Both are gravity-driven pipe networks; the differing hydraulics motivate our distinct targets (flood volume vs. water level) and help interpret spatial error patterns near control structures.

4.3 Methodology

This study adapts the **GATRes** model [16], originally developed for water distribution networks, to wastewater and stormwater systems. GATRes was selected because prior work [16] shows that it achieves state-of-the-art performance compared to several alternative GNN architectures on water distribution networks; since urban drainage networks share similar sparse, graph-structured topologies and hydraulic dependencies, we assume that the same inductive biases will transfer well. We use GATRes as a surrogate learning framework, focusing on its applicability to multiple urban drainage systems. For architectural details, we refer readers to the original paper.

Graph representation. Each network is represented as a directed graph $G = (V, E)$, where nodes correspond to manholes, pumps, or outfalls, and edges to physical links such as pipes, conduits, or channels. Node values represent observed or unobserved hydraulic states (flood depth for Heusden, cumulative volume for Shunqing).

Learning setup and masking. To emulate sparse sensing, a fixed subset of nodes $\mathcal{V}_{\text{obs}} \subset V$ is randomly chosen as sensors, and their values are provided as input. The model is trained to estimate the states of all other nodes $v_j \in V \setminus \mathcal{V}_{\text{obs}}$. A fixed masking strategy ensures consistency across timesteps within each scenario. Training minimizes the mean absolute error between estimated and simulated values.

This setup trains and evaluates the surrogate separately on each dataset under the same masking protocol and loss, enabling a consistent comparison across simulation frameworks and targets and an assessment of robustness under sparse sensing.

5 Results

For both datasets explained in Section 4, we apply the same training/validation protocol, sensor-availability settings (10–40% of nodes observed), and evaluation metrics (MAE, RMSE). Each dataset is trained and assessed *independently* with its own model instance.

All experiments are conducted using the same reproducible configuration: GATRes models with either small (15 blocks, 64 hidden channels, ~300K parameters) or medium (20 blocks, 128 hidden channels, ~500K parameters) architectures, trained with the Adam optimizer (learning rate 1×10^{-3}, weight decay 1×10^{-5}), batch sizes of 4–8 depending on network size, for up to 500 epochs with early stopping (patience of 50 epochs, minimum delta 1×10^{-6}). Training is performed on GPU-enabled nodes (NVIDIA A100) using PyTorch 2.1.2 with CUDA 12.1.1. Data is split into 70% training, 15% validation, and 15% test sets, with node features normalized using StandardScaler (Z-score normalization) with statistics computed on the training set only. The random seed is fixed to 42 for reproducibility across all experiments.

We report three complementary views: (i) performance as sensor availability decreases, (ii) spatial distribution of errors, and (iii) ability to reproduce hydraulic dynamics.

5.1 Effect of Sensor Availability

Heusden (water level, m). We investigate the MAE achieved by the GatRes model across different sensor coverage levels, as shown in Figure 3. Under a 70% masking rate, corresponding to 30% sensor availability, the model achieves a low MAE of 0.1059 meters. Figure 3 also illustrates the performance degradation as sensor coverage decreases. However, even with only 10% sensor availability, the model achieves a mean absolute error of 0.1753 meters (17.53 centimeters) across all masked nodes, which remains low and within the acceptable range defined by water engineering experts.

Shunqing (flood volume, m^3). Under 70% masking (30% sensors), the model achieves an MAE of $0.453\,\mathrm{m}^3$ and an RMSE of $2.145\,\mathrm{m}^3$ (Table 4). Given observed flood volumes up to $164.36\,\mathrm{m}^3$ with a mean of $20.48\,\mathrm{m}^3$, the MAE corresponds to $\sim 2.2\%$ of the mean event magnitude. Figure 2 shows that performance decreases gradually as sensor availability is reduced from 40% to 10%; even with only 10% of sensors, the MAE remains modest (e.g., $0.619\,\mathrm{m}^3$) relative to typical event sizes.

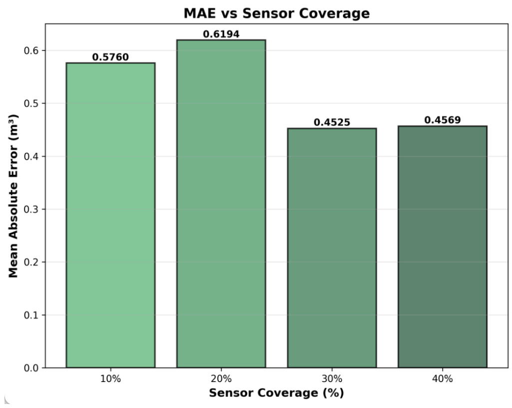

Figure 2: Impact of sensor coverage on MAE for the Shunqing network.

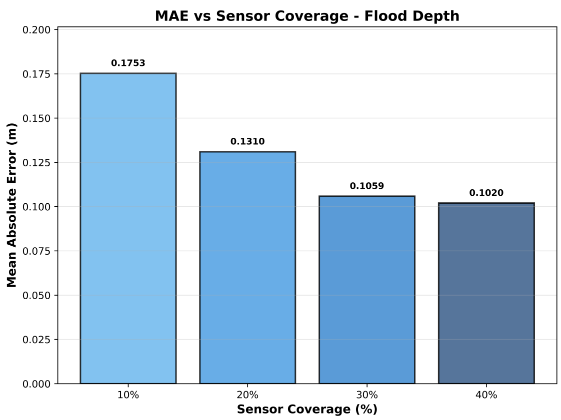

Figure 3: Impact of sensor coverage on MAE for the Heusden network.

Table 4: Performance of GATRes on the Shunqing dataset (70% masking).

Sensor Availability	MAE (m^3)	RMSE (m^3)
30% (70% masking)	**0.453**	2.145

5.2 Spatial Distribution of Errors

Heusden. With 30% sensor coverage (268 sensors out of 894 nodes), most nodes show low MAE values (Figure 4), ranging from 0.0162 to 0.3554 meters. This indicates that the GNN model can effectively capture the relationships between connected nodes. However, closer inspection of the areas with higher errors reveals that they are concentrated near hydraulic control elements (e.g., pumps, weirs) and outfalls, where nonlinear behavior due to the flow-altering effects of these components.

Shunqing. At the same 30% coverage, node-level errors are well contained and more uniformly distributed across the network (Figure 5). This pattern reflects the smaller stormwater network and the volume-based target, while still illustrating the surrogate's ability to interpolate across sparsely sensed areas.

5.3 Hydraulic Dynamics

Heusden. Figure 6 compares GNN's per-timestep estimation with *InfoWorks ICM* simulation results at two representative nodes. Figure 6(a) shows that the GNN is able to estimate the increase and decrease of water level indicating that the GNN model able to learn well from the existing sensor data. However, the node nears the flow-altering structures, Figure 6(b), shows a significant difference especially on a high water level conditions. This result indicates that our GNN model are still not be able to fully capture the changing dynamics around complex hydraulic structures.

Shunqing. In the event-driven stormwater setting, the target is cumulative flood volume per timestep and per node. The low MAE/RMSE in Table 4, together with the bounded spatial errors in Figure 5, indicate accurate magnitude estimates across 148 rainfall events with diverse durations and intensities.

Summary. Across two networks, target variables (level vs. volume), and sensor-availability settings, the surrogate yields accurate, spatially coherent sequences of per-timestep estimates from sparse observations, providing a practical basis real-time monitoring needs in urban drainage.

6 Discussion

The experiments highlight the value of graph neural networks as surrogates for wastewater digital twins. Four themes emerge: calibration and learning challenges, behavior under extreme versus routine conditions, applicability across different network architectures and simulation frameworks, and broader implications for smart infrastructures.

6.1 Implications for Calibration and Learning

The results reveal important insights for calibration and learning in wastewater digital twins. Estimation accuracy is not uniform across the network, with errors clustering near pumps, weirs, and outfalls. This pattern reflects localized nonlinear dynamics and the limited influence of sparse sensors on complex hydraulic structures. The spatial error distribution suggests that calibration strategies for wastewater DTs should explicitly incorporate hydraulic structures and flow paths when retraining or adapting models, potentially requiring weighted loss functions or targeted training-data augmentation near critical infrastructure elements.

The learning process benefits from the graph structure's ability to propagate information across the network topology. However, the concentration of errors near control structures indicates that the current approach may need additional training diversity or specialized attention mechanisms to better capture the nonlinear behavior at these locations. Future work could explore adaptive masking strategies that provide more training exposure to challenging regions. In addition, extending GATRes to a *heterogeneous* variant that explicitly models different node and edge types (e.g., pumps,

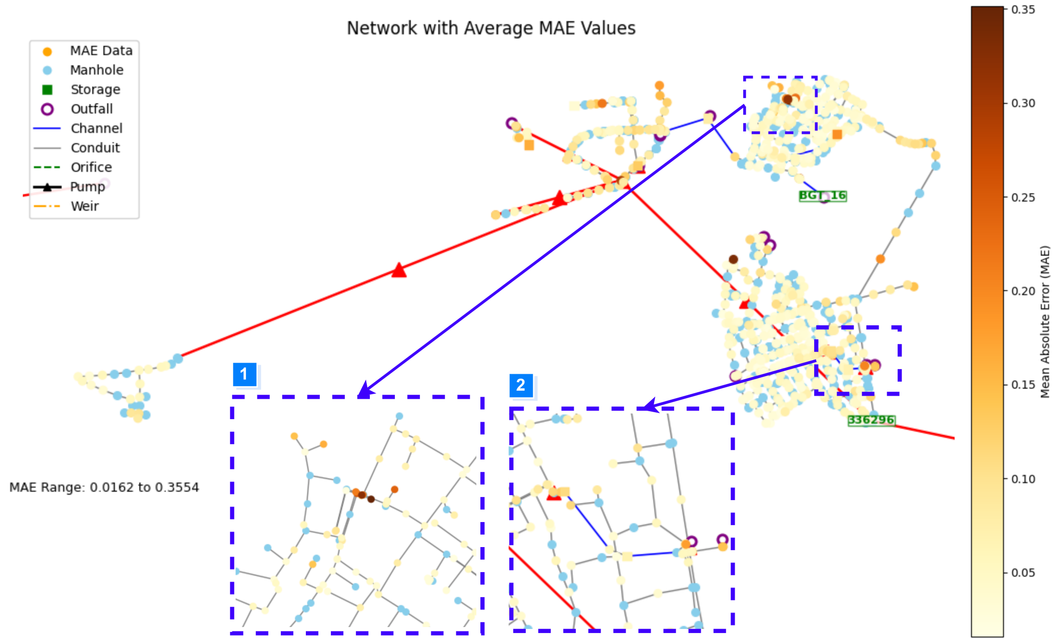

Figure 4: Spatial distribution of MAE across the Heusden network under 30% sensor coverage.

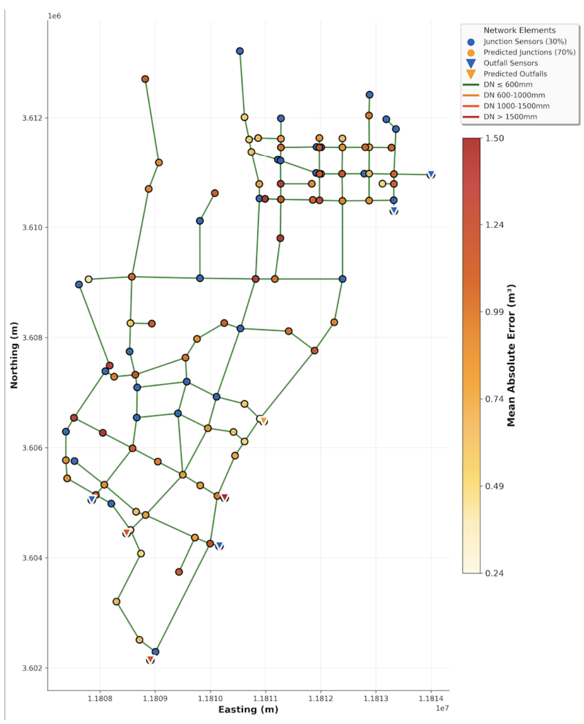

Figure 5: Node-level MAE in Shunqing under 70% masking (30% sensors).

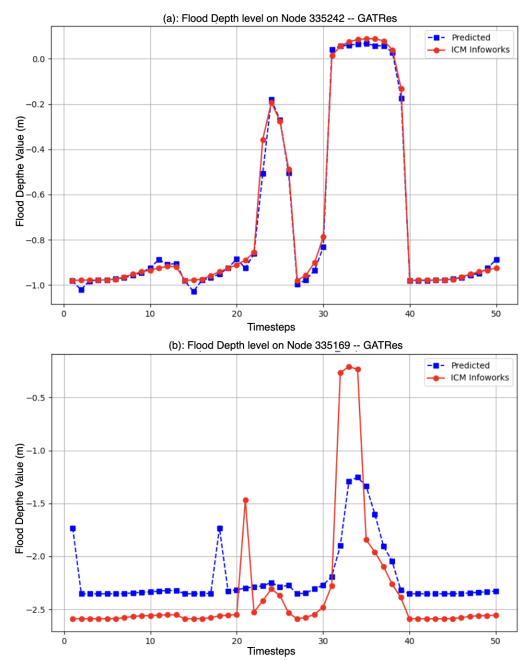

Figure 6: Heusden: per-timestep GNN estimation versus *InfoWorks ICM* simulations at two nodes.

weirs, outfalls versus conduits and manholes) could further improve estimation accuracy by incorporating the distinct hydraulic roles of these components [16].

Equally important, the ability to produce reliable estimates with only 30% sensor coverage illustrates the surrogate's value as a virtual sensing mechanism. This capability supports cost-efficient monitoring strategies, where installing dense sensor networks is often infeasible for municipalities managing large wastewater infrastructures.

6.2 Performance under Extreme and Routine Conditions

The analysis reveals distinct performance characteristics under different operational conditions. From an operational perspective, peak conditions are critical for emergency response and flood management. The surrogate produces estimates that align well with flood-wave timing and overall propagation patterns in the reference simulations, but amplitude errors remain during sharp transitions, particularly at nodes with complex hydraulic behavior. These limitations point to the need for training on a wider variety of rainfall scenarios, particularly extreme events with rapid intensity changes that challenge the model's ability to capture nonlinear dynamics.

Under routine conditions, the surrogate delivers consistent and stable estimates, supporting applications in daily monitoring and operational decision-making. During normal flow conditions, the estimates remain smooth and physically plausible as conditions evolve within a scenario, maintaining accuracy throughout. This stability is crucial for continuous monitoring applications where false alarms or missed detections could lead to unnecessary interventions or missed maintenance opportunities.

The contrast between routine and extreme performance highlights a key trade-off in surrogate modeling: while the approach excels at capturing typical system behavior, it benefits from additional training diversity to handle the full spectrum of operational conditions. Implementations should therefore prioritize comprehensive scenario coverage during training, including both frequent low-intensity events and rare high-intensity storms.

6.3 Applicability across Networks and Frameworks

We evaluate the same pipeline, trained *independently* on each dataset using an identical protocol, on two markedly different networks: a large gravity-driven wastewater system in Heusden (The Netherlands) modeled in *InfoWorks ICM*, and a smaller stormwater system in Shunqing (China) modeled in SWMM [21, 22, 19]. Despite differences in topology, scale, and target variables (water level vs. cumulative flood volume), the surrogate attains reliable accuracy under sparse sensing in both settings. This indicates that the graph-based formulation captures hydraulic patterns relevant across infrastructures and is compatible with multiple simulation environments.

At the same time, the results emphasize that dataset-specific preprocessing remains necessary, particularly when switching between target variables (flood depth in Heusden versus cumulative volume in Shunqing). Future work should explore transfer-learning strategies that reduce manual adjustments while preserving robustness across infrastructures.

6.4 Broader Impact for Smart Cities and Critical Infrastructures

By bridging detailed physics-based simulations with graph-based learning, the surrogate provides a scalable and adaptive monitoring tool for smart city infrastructures. It enables reliable state estimation without relying exclusively on computationally demanding hydraulic models, thereby supporting cost-effective and responsive infrastructure management.

The approach directly addresses three practical challenges in urban wastewater management: limited sensor coverage, the computational cost of detailed hydraulic modeling, and the need for methods that can be deployed across diverse infrastructures. By producing accurate, network-wide estimates from sparse inputs, the surrogate offers a cost-efficient alternative to dense instrumentation or continuous full-physics simulations, which is particularly valuable for municipalities operating under budget constraints or with legacy monitoring systems.

7 Conclusion

This study demonstrates the potential of graph neural networks as surrogates for wastewater monitoring within digital-twin ecosystems. The proposed pipeline was evaluated on two distinct urban drainage networks, Heusden (The Netherlands) and Shunqing (China), using identical training, masking, and evaluation protocols. In Heusden, water-level estimation errors fall within 20–50 cm across multiple rainfall scenarios; in Shunqing, flood-volume estimation achieves low absolute errors relative to event magnitudes. Taken together, these results show that the approach is accurate under sparse sensing and applicable across different network architectures and simulation tools.

By enabling accurate, scalable, and cost-effective state estimation, this work contributes to the development of smart, resilient urban drainage systems. Graph-based surrogates can play a central role in bridging the gap between detailed hydraulic simulations and practical real-time monitoring.

8 Limitations and Future Work

While promising, several aspects offer opportunities for further development.

- **Training labels and verification scope.** We use calibrated hydraulic simulations to obtain dense network labels that are unavailable in the field, enabling broad coverage of operating regimes and controlled stress testing. At present, model training relies exclusively on these simulation-derived labels; historical sensor time series are not used directly, because existing instrumentation is sparse and unevenly distributed across the network. In this work, verification is performed via held-out rainfall scenarios and spatial masking (hide-a-sensor) across two networks. In live deployments, formal verification is necessarily confined to instrumented nodes; to address this, future work will incorporate on-line checks against available sensors, realistic noise/outage models derived from telemetry, and explicit noise injection into simulation outputs to mimic sensor uncertainty and outages during training, together with periodic co-calibration/fine-tuning to track system changes.

- **Temporal modeling.** The current setup treats system states as independent snapshots. Incorporating temporal information through spatio-temporal GNNs or sequence-based extensions could improve peak estimation and flood-wave tracking.
- **Model design.** Extending GATRes toward a heterogeneous and temporally aware variant could better capture nonlinear dynamics around pumps, weirs, and outfalls by encoding node/edge types explicitly.

Future work will explore heterogeneous, time-aware GATRes architectures, integrate richer feature sets, and evaluate continuous deployment scenarios with live telemetry streams and noise-augmented training regimes that more closely reflect real-world sensor behaviour.

Artifact Availability

Code for the GatRes implementations is provided as part of the replication package. It is available at *https://github.com/DDTClean-Project/gatres-shunqing-ddtclean* and *https://github.com/DDTClean-Project/gatres-heusden*

References

[1] Esther Ugo Alum, Emmanuel Ifeanyi Obeagu, and Okechukwu Paul-Chima Ugwu. 2024. Enhancing quality water, good sanitation, and proper hygiene is the panacea to diarrhea control and the attainment of some related sustainable development goals: a review. *Medicine*, 103, 38, (Sept. 2024). doi:10.1097/md.0000000000039578.

[2] G. Baskar, A. N. Parameswaran, and R. Sathyanathan. 2025. Optimizing papermaking wastewater treatment by predicting effluent quality with node-level capsule graph neural networks. *Environmental Monitoring and Assessment*, 197, 2. doi:10.1007/s10661-024-13581-3.

[3] Yan Chen, Daoping Huang, Jing Wu, Gang Fang, Yan Li, Abid Aman, and Yiqi Liu. 2025. Towards sludge bulking diagnosis via dynamic attention graph neural network. *Journal of Water Process Engineering*, 69, 106774. doi:10.1016/j.jwpe.2024.106774.

[4] Anastasis Christou et al. 2024. Sustainable wastewater reuse for agriculture. *Nature Reviews Earth amp; Environment*, 5, 7, (June 2024), 504–521. doi:10.1038/s43017-024-00560-y.

[5] F. H. L. R. Clemens. 2001. *Hydrodynamic Models in Urban Drainage: Application and Calibration*. Ph.D. Dissertation. Delft University of Technology. ISBN: 90-407-2163-7.

[6] Victoria Degeler, Mostafa Hadadian, Erkan Karabulut, Alexander Lazovik, Hester van het Loo, Andrés Tello, and Huy Truong. 2025. Ditec: Digital Twin for evolutionary changes in water distribution networks. In *Leveraging Applications of Formal Methods, Verification and Validation. Application Areas*. Springer Nature Switzerland, 62–82. ISBN: 978-3-031-75390-9.

[7] A. Garzón, Z. Kapelan, J. Langeveld, and R. Taormina. 2022. Machine learning-based surrogate modeling for urban water networks: review and future research directions. *Water Resources Research*, 58, 5. doi:10.1029/2021wr031808.

[8] MohammadHossein Homaei, Agustín Javier Di Bartolo, Mar Ávila, Óscar Mogollón-Gutiérrez, and Andrés Caro. 2024. Digital transformation in the water distribution system based on the Digital Twins concept. (2024). https://arxiv.org/abs/2412.06694 arXiv: 2412.06694 [cs.CY].

[9] Offir Inbar, Moni Shahar, and Dror Avisar. 2024. 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. doi:10.1039/d4ew00111g.

[10] Tove A. Larsen, Sabine Hoffmann, Christoph Lüthi, Bernhard Truffer, and Max Maurer. 2016. Emerging solutions to the water challenges of an urbanizing world. *Science*, 352, 6288, 928–933. doi:10.1126/science.aad8641.

[11] Zilin Li, Haixing Liu, Chi Zhang, and Guangtao Fu. 2024. Real-time water quality prediction in water distribution networks using graph neural networks with sparse monitoring data. *Water Research*, 250, 121018. doi:10.1016/j.watres.2023.121018.

[12] J. Meirlaen, B. Huyghebaert, F. Sforzi, L. Benedetti, and P. Vanrolleghem. 2001. Fast, simultaneous simulation of the integrated urban wastewater system using mechanistic surrogate models. *Water Science and Technology*, 43, 7, 301–309. doi:10.2166/wst.2001.0439.

[13] Duc-Thinh Ngo, Ons Aouedi, Kandaraj Piamrat, Thomas Hassan, and Philippe Raipin-Parvédy. 2023. Empowering Digital Twin for future networks with graph neural networks: overview, enabling technologies, challenges, and opportunities. *Future Internet*, 15, 12, 377. doi:10.3390/fi15120377.

[14] RIONED. 2024. Ontwerpbuien met statistische herhalingstijd (bui01–bui10). (2024). https://www.riool.net/kennisbank/onderzoek/modelleren-hydraulisch-functioneren/stap-iv-hydraulische-belasting-bepalen/hemelwaterafvoer-hwa/ontwerpbuien-met-statistische-herhalingstijd/bui01-bui10.

[15] Naghmeh Shafiee Roudbari, Shubham Rajeev Punekar, Zachary Patterson, Ursula Eicker, and Charalambos Poullis. 2024. From data to action in flood forecasting leveraging graph neural networks and Digital Twin visualization. *Scientific Reports*, 14, 1. doi:10.1038/s41598-024-68857-y.

[16] Huy Truong, Andrés Tello, Alexander Lazovik, and Victoria Degeler. 2024. Graph neural networks for pressure estimation in water distribution systems. *Water Resources Research*, 60, 7. doi:10.1029/2023WR036741.

[17] Guido Vaes, Patrick Willems, and Jean Berlamont. 2002. The use of reservoir models for the assessment of the input from combined sewer overflows into river models. *Global Solutions for Urban Drainage*, 1–16. doi:10.1061/40644(2002)139.

[18] Charlotte Weil, Simon Elias Bibri, Régis Longchamp, François Golay, and Alexandre Alahi. 2023. Urban Digital Twin challenges: a systematic review and perspectives for sustainable smart cities. *Sustainable Cities and Society*, 99, 104862. doi:10.1016/j.scs.2023.104862.

[19] Zhiyu Zhang, Wenchong Tian, Chenkaixiang Lu, Zhenliang Liao, and Zhiguo Yuan. 2024. Graph neural network-based surrogate modelling for real-time hydraulic prediction of urban drainage networks. *Water Research*, 263, 122142. doi:10.1016/j.watres.2024.122142.

[20] Guang-yao Zhao, Shunya Suzuki, Jia-hao Deng, and Masafumi Fujita. 2022. Machine learning estimation of biodegradable organic matter concentrations in municipal wastewater. *Journal of Environmental Management*, 323, 116191. doi:10.1016/j.jenvman.2022.116191.

[21] Jinjun Zhou, Yali Pang, Hao Wang, Wentao Li, Jiahong Liu, Zhuoran Luo, Weiwei Shao, and Haijia Zhang. 2023. Sewage network operational risks based on InfoWorks ICM with nodal flow diurnal patterns under NPIs for COVID-19. *Water Research*, 246, 120708. doi:10.1016/j.watres.2023.120708.

[22] Qiande Zhu et al. 2024. Simulation and analysis of water quality improvement measures for plain river networks based on InfoWorks ICM model: case study of baoying county, china. *Water*, 16, 18, 2698. doi:10.3390/w16182698.