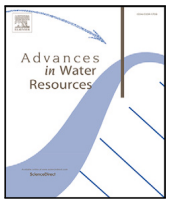
Advances in Water Resources 211 (2026) 105261

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres



Hybrid AI for estimating water levels at ungauged river locations via graph neural networks and terrain-aware edges

Anh Minh Truong *, Guangan Chen, Michiel De Baets, Michiel Vlaminck, Brian Booth, Hiep Luong

IPI-TELIN, Ghent University-IMEC, Technicum - Block 3, Sint-Pietersnieuwstraat 41, Gent, 9000, East Flanders, Belgium

ARTICLE INFO

Dataset link: Floodify dataset (Original data)

Keywords:
Graph neural network
Deep learning
Water-level estimation

ABSTRACT

Monitoring river water levels is critical for flood risk management, water-resource planning, and early warning systems. However, deploying dense gauge networks across extensive river systems is often infeasible due to logistical and financial constraints, and existing stations may fail or provide intermittent data. In this work, we propose **HIGNN (Hydrological Interpolation based on Graph Neural Network)**, a graph-based framework for estimating water-level changes at virtual sensor locations (i.e., ungauged sites or locations with missing observations) by leveraging information from neighboring telemetry stations and terrain characteristics. In HIGNN, nodes represent observation sites, waterway intersections, or virtual stations, while edges represent hydrological connectivity (e.g., upstream–downstream relations) and are characterized by topographic attributes such as elevation profiles, slope statistics, and flow direction. The model employs message passing to propagate water-level change signals through the river network, modulated by physically meaningful edge attributes. Across all water-level change brackets, HIGNN achieves the lowest mean RMSE, outperforming interpolation- and regression-based baselines. These results demonstrate that HIGNN can effectively estimate water-level changes at ungauged or temporarily unmonitored locations.

1. Introduction

Monitoring water levels is an increasingly important task due to the intensifying impacts of climate change, which results in more frequent and extreme rainfall events and increases the likelihood of river and canal overflows (Hirabayashi et al., 2013; Blöschl et al., 2017). Waterways such as rivers, canals, and streams are particularly critical to monitor, as they are the first to respond to excessive rainfall and can rapidly flood surrounding areas. The information gathered from this monitoring supports the implementation of immediate mitigation measures and long-term flood management strategies, thereby reducing casualties and economic losses due to flood events (Willner et al., 2018). However, deploying a dense network of physical telemetry is often impractical due to the substantial costs of sensor hardware, power supply, data transmission infrastructure, and ongoing maintenance. An alternative approach is to set up "virtual" stations, where water levels are estimated using data from neighboring telemetry stations. These serve as a cost-effective means to provide broader spatial coverage for critically monitoring areas. Furthermore, they can also provide useful information to decide where physical sensors may be most effectively deployed, by revealing areas with consistently high estimation uncertainty or strong hydrological variability, which indicate where direct measurements would most improve monitoring accuracy.

In recent years, virtual stations have emerged as a scalable and cost-efficient alternative to traditional sensor networks for estimating river water levels. Despite their potential, accurately estimating water levels at these locations remains a major challenge due to the intricate spatiotemporal behavior of hydrological systems, which are shaped by numerous environmental drivers such as precipitation, evaporation, land cover, and catchment characteristics. These processes interact in complex and often nonlinear ways, making it difficult to model how water levels evolve over time. Among these factors, precipitation plays a particularly dominant role, with recent rainfall events consistently linked to short-term water level increases, as demonstrated in prior empirical studies (Kenda et al., 2020; Ahmed et al., 2022). However, beyond temporal dynamics, the spatial relationships between virtual stations are also challenging to capture due to variations in terrain, river topology, and hydrological connectivity. These spatial heterogeneities hinder the development of generalized models that can be applied reliably across different geographic regions. As a result, generating

* Corresponding author.
E-mail address: anhminh.truong@ugent.be (A.M. Truong).

https://doi.org/10.1016/j.advwatres.2026.105261
Received 28 December 2025; Received in revised form 22 February 2026; Accepted 25 February 2026
Available online 3 March 2026

accurate predictions becomes especially difficult in sparsely monitored areas, where ground-truth data is limited or unevenly distributed.

Furthermore, physical sensors may malfunction, for example due to power issues, environmental damage, or calibration drift (Nagahama et al., 2024), and they may also suffer from data gaps and irregular sampling frequencies. These problems can introduce bias and degrade model performance. Combined with the varying number and spatial configuration of nearby stations, they make it difficult to develop models that are both robust and generalizable for predicting water levels at ungauged locations.

A widely used family of approaches for estimating values at unsampled locations is spatial interpolation (Yasin et al., 2024; Lee et al., 2023; Khan et al., 2023; Paiva et al., 2015; Li and Heap, 2014). Non-geostatistical methods such as inverse distance weighting (IDW) use distance-based rules to combine nearby observations, while geostatistical methods such as Ordinary Kriging (OK) estimate values using an explicit model of spatial correlation and can provide uncertainty estimates (Journel and Huijbregts, 1978; Cressie, 1993). Although effective in some settings, purely spatial interpolation often relies on smoothness or stationarity assumptions and may not capture river-network connectivity, terrain-driven anisotropy, or abrupt changes caused by local controls (e.g., structures and operations). This motivates incorporating auxiliary information (e.g., precipitation) and learning spatial dependencies beyond simple distance-based weighting.

Beyond interpolation, river water-level prediction is also addressed by several complementary directions. Physics-based hydrodynamic models simulate flow and water levels using governing equations, but typically require detailed channel geometry and parameterization (e.g., roughness) and therefore substantial calibration (Jiang et al., 2021). Data-driven time-series models (e.g., LSTM-based predictors) can capture nonlinear temporal patterns from historical observations and have been widely explored for river stage/level forecasting, but extending them to ungauged locations depends on how spatial relationships are represented and on robustness to missing data (Luo et al., 2025). Another line of work treats water level (or related hydrologic variables) as a state variable and applies data assimilation methods to update model states as new observations arrive; these approaches can be effective but rely on careful specification of model and observation uncertainties and are commonly coupled with a dynamical model (Sun et al., 2016; Matgen et al., 2010).

One notable approach that combines precipitation-driven modeling with spatial error correction is the Bayesian spatiotemporal model developed by Nagahama et al. (2024), which aims to predict water levels at both gauged and ungauged stations using precipitation as the primary driver. The method operates in two main stages. In the first stage, water levels at all locations, including those without sensors, are estimated by applying a regression model that uses precipitation and other environmental covariates as predictors. In the second stage, the model computes the residuals between the predicted and observed water levels at the gauged stations. These residuals, which capture spatially structured errors, are then modeled as a Gaussian process using a Nearest-Neighbor Gaussian Process (NNGP) framework. By incorporating temporal autocorrelation and leveraging spatial proximity, the NNGP interpolates the residuals across space to estimate the likely prediction errors at ungauged stations. These estimated residuals are then used to adjust the initial predictions, resulting in more accurate and spatially consistent water level estimates. Similarly, Tucci (2023) integrated spatial interpolation with a neural network to estimate hourly water levels in a hydroelectric basin, using variables such as temperature, humidity, wind speed, and precipitation.

In parallel with these developments, graph neural networks (GNNs) have emerged as a robust framework for modeling water systems. GNNs are particularly well-suited for hydrological applications because observations are often distributed across spatially interconnected networks like rivers, groundwater wells, or water distribution systems. By representing monitoring locations as nodes and their hydrological connections or flow pathways as edges, GNNs can effectively encode spatial structures while simultaneously incorporating temporal data and additional variables.

A recent study in the Netherlands used spatial–temporal graph neural networks (ST-GNNs) (Taccari et al., 2024) that combined groundwater level time series with auxiliary data such as precipitation, evaporation, river stage, and pumping well operations. This graph-based architecture enabled the model to learn both spatial interconnectivity and temporal evolution, resulting in improved prediction accuracy and robustness over traditional numerical methods, particularly under missing-data conditions.

While the ST-GNN model proposed by Taccari et al. (2024) effectively integrates groundwater and auxiliary hydrological data within a unified graph, our approach introduces several key innovations to better capture the hydrological dynamics of river networks:

- **Directed hydrological graph.** Rather than building an undirected graph with fixed, proximity-based edge weights, we construct a directed graph that follows the flow direction toward the virtual sensor locations, which serve as the prediction targets.
- **Physically informed edge features.** Each edge incorporates detailed topographic and hydrological attributes, such as elevation differences and flow direction indicators, enabling the model to represent hydrologically meaningful pathways.
- **Adaptive message passing.** These edge features are not manually weighted; instead, they are used within a learnable message-passing framework that adaptively infers the relative importance of each neighbor.
- **Residual-based interpolation.** Unlike Taccari et al. who did not explicitly model prediction errors, we treat residuals at telemetry stations as node features and interpolate them through the network, thereby incorporating both physical terrain information and predictive uncertainty.

Beyond open-channel hydrology, GNNs have also been applied to water infrastructure systems, e.g., for pressure reconstruction and leakage localization in water distribution networks under sparse sensing (Zhang and Fink, 2024). These studies highlight the broader utility of graph learning in water resources; however, river networks differ fundamentally from pressurized pipe systems in flow directionality and the role of terrain-driven connectivity, motivating a model that explicitly encodes directed river topology and topographic edge attributes.

In this work, we extend the standard message-passing Graph Neural Network (GNN) framework to estimate water levels in river networks by representing the waterway system as a graph, where nodes correspond to telemetry stations with observations or *virtual stations* without measurements, and edges encode hydrologically and topographically relevant information. The proposed method builds on Nagahama et al. (2024), replacing their linear regression component with a multilayer perceptron (MLP) that estimates water-level changes from precipitation data. The MLP is followed by a GNN that refines these estimates by propagating information through the river graph. Both components are trained end-to-end, enabling a unified optimization of temporal forcing (precipitation) and spatial dependencies.

We formulate the prediction task as residual learning: starting from the MLP baseline, the GNN learns a correction term by integrating spatial patterns from neighboring telemetry stations, guided by the physical and topographical relationships encoded in the graph. We hypothesize that combining historical water level data with graph-based hydrological connectivity allows the model to interpolate water-level changes at unmonitored locations, providing a cost-effective way to extend monitoring coverage without additional sensor deployments.

The remainder of this paper is organized as follows. Section 2 describes the data sources used in this study. Section 3 introduces the proposed GNN-based interpolation method, and Section 4 details the experimental setup and evaluation metrics. Section 5 presents and discusses the results. Finally, Section 6 concludes the paper and highlights potential directions for future research.

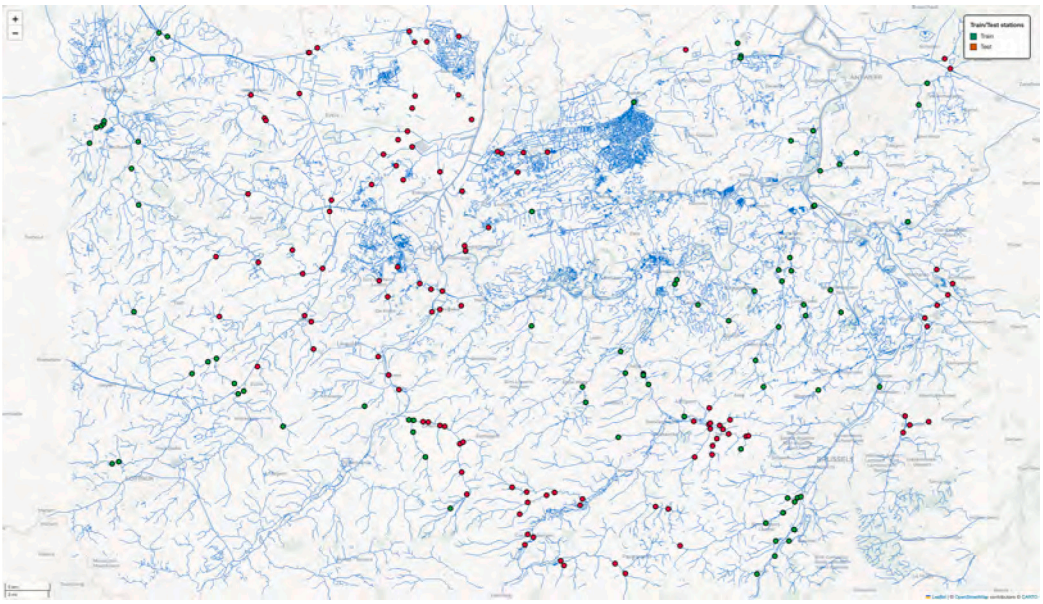


Fig. 1. Train and test station locations overlaid on the waterway network used to construct station neighborhoods. Green circles denote training stations and red circles denote test stations.

2. Study data

The data used in this study were collected from 395 measurement stations distributed across Flanders, Belgium. The stations are managed by the Flemish government and are publicly accessible via Waterinfo Vlaanderen.[1] Most stations are equipped with limnigraphs, i.e., instruments that continuously measure and record river stage (water level) (Vlaamse Milieumaatschappij (VMM) et al., 2025). At each limnigraph location, a stage–discharge (Q–H) relationship is established from regular in situ flow measurements, yielding a rating curve that links a water level H to a corresponding discharge Q. This relationship provides the basis for converting continuous water level observations into flow estimates when needed.

Depending on station configuration, water levels are measured using either acoustic or pressure sensors. Each sensor reports the height of the water surface relative to a local reference (gauge datum) that is fixed for the station. Because these local reference points are tied to the surrounding land surface, their absolute elevations can differ substantially across the network, making raw water level values not directly comparable between stations. To reduce the influence of these station-specific elevation offsets and to emphasize relative fluctuations, we normalize each station's water level time series by subtracting its median value. This transformation centers the data around zero, allowing the model to focus on water level changes over the time rather than the height of the water surface, which may be influenced by local topography and station installation characteristics.

Fig. 1 shows the train and test stations overlaid on the waterway network together with a labeled basemap. The stations are distributed along waterways across Flanders and span a wide range of locations, including areas around *Antwerp*, *Brussels*, *Gent*, *Brugge*, *Kortrijk*, *Roeselare*, *Mechelen*, *Aalst*, *Dendermonde*, and *Sint-Niklaas*. Green circles indicate stations assigned to the training set and red circles indicate stations assigned to the test set, while the blue polylines depict the waterway network used to define station neighborhoods in our graph construction.

In addition, the raw observations are not provided at a uniform sampling interval across stations (e.g., 15 min, 30 min, or 1 h depending on the station). To ensure consistent model inputs and evaluation, we align all time series to a common hourly timestamp set. No temporal interpolation is performed; if a station has no observation for a given hour, the value is treated as missing.

In addition to the water level and precipitation records, we use a Digital Terrain Model (DTM) of Flanders obtained from Vlaanderen Open Data (Agentschap Digitaal Vlaanderen, 2019). The DTM is provided at 1 m horizontal resolution and was derived from LiDAR acquisitions collected under the Digital Height Model Flanders II (DHMV II) campaign (2013–2015) (Agentschap Digitaal Vlaanderen, 2019).

Fig. 2 shows the DTM mosaic together with the locations of the monitoring stations used in this work. Green markers indicate stations assigned to the training set and red markers indicate stations assigned to the test set.

The map spans elevations from approximately −10.5 m to 156.2 m, with median elevation 14.7 m (5th–95th percentile: 3.6–76.5 m). Training and test stations cover similar elevation ranges: training stations range from 1.5 m to 82.5 m (median 13.3 m), while test stations range from 2.6 m to 49.1 m (median 9.7 m). The map highlights a clear elevation gradient across Flanders, with low-lying terrain in the coastal and northern areas and higher elevations toward the southern and southeastern parts of Flanders.

To ensure consistency across stations and to capture meaningful temporal dynamics, we first resample each water-level time series to a uniform one-hour interval, since different stations record measurements at varying frequencies. We then divide each resampled series into non-overlapping weekly segments. This procedure standardizes the input length for modeling and enables fine-grained filtering of short-term temporal patterns.

We then apply a simple filtering step to remove weekly segments affected by sensor or telemetry failures that introduce unrealistically large offsets in the recorded water level. For the remaining segments, we target short isolated and large spikes in the time series, that are inconsistent with the local water-level evolution, while preserving sustained changes that may correspond to real hydrological events. For each weekly segment, we remove short sensor spikes using a Hampel filter (Hampel, 1974). The Hampel filter compares each value to the local median computed within a sliding window centered at that time step. A sample is flagged as an outlier if its deviation from the local median exceeds a threshold scaled by the local median absolute deviation (MAD), which provides a robust estimate of the typical variability

[1] https://www.waterinfo.vlaanderen.be/Themas#item=flood/current%20situation.

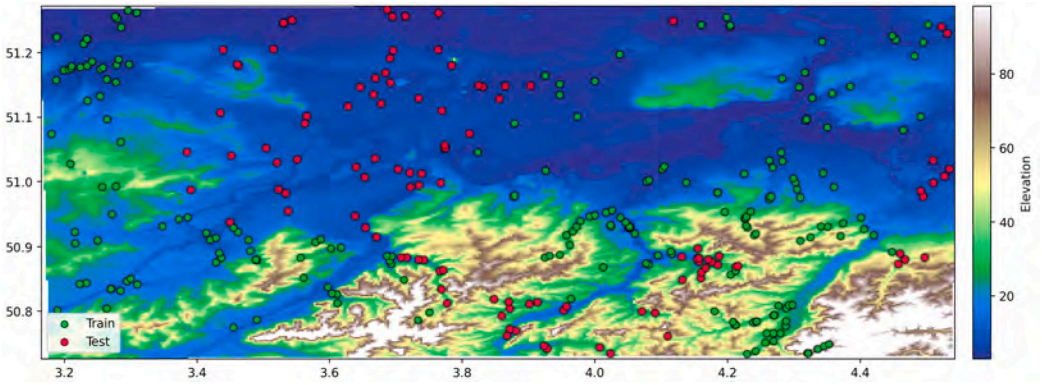
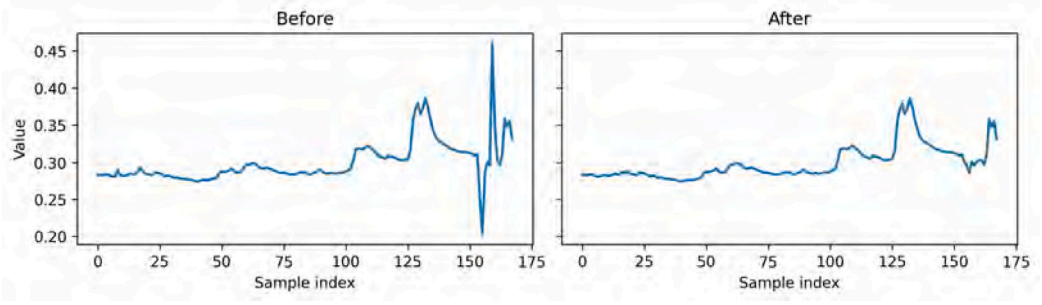
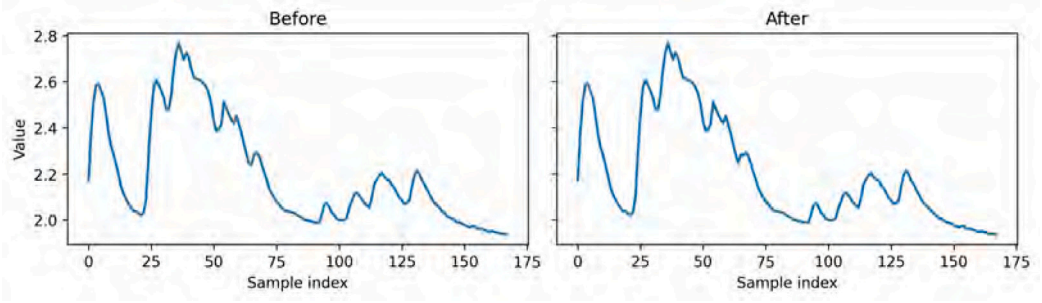


Fig. 2. Downsampled digital terrain model (DTM) of Flanders with monitoring stations used in this study. Green circles denote training stations and red circles denote test stations.



(a) Example A



(b) Example B

Fig. 3. Qualitative examples of Hampel filtering. Each panel shows the same weekly segment before (left) and after (right) filtering with a shared y-axis scale. The Hampel filter removes short isolated spikes while preserving sustained water-level evolution.

within the window. Flagged samples are then replaced with the local median. This procedure suppresses isolated spikes while largely preserving genuine hydrological changes that occur over multiple hours.

Fig. 3 shows two representative weekly segments before and after applying the Hampel filter. In both cases, the filter suppresses short, isolated spikes that are inconsistent with the local evolution, while preserving longer-duration variations that may correspond to genuine hydrological changes.

The processed dataset compiled for this study, including selected time series of water level and associated metadata, will be made publicly available at https://github.com/tmanh/floodify-sensor-fusion-data.

3. Proposed method

In this section, we propose a spatiotemporal framework that builds upon the two-stage approach of Nagahama et al. (2024). In that work, water levels at both gauged and ungauged locations are first predicted from precipitation using a simple regression model. The residual errors at gauged stations are then modeled and propagated with an NNGP to correct predictions at ungauged locations.

We extend (Nagahama et al., 2024) by replacing the first-stage regressor with an end-to-end trainable architecture that explicitly models both spatial and temporal dependencies. Specifically, precipitation inputs are first encoded by an MLP, and the resulting node-wise features are refined by a GNN operating on the waterway-based graph. The GNN aggregates information across stations according to the river-network topology and edge attributes, yielding spatially informed node representations at each time step. A gated recurrent unit (GRU) (Cho et al., 2014) is then applied along the temporal dimension to these GNN-refined representations to produce baseline water-level predictions over time. This separation clarifies the role of each module: the GNN performs spatial mixing conditioned on hydrological connectivity and topography, while the GRU captures temporal evolution and memory effects in the water-level dynamics.

3.1. Background: Graph attention networks

Graph Attention Networks (GATs) (Veličković et al., 2018) are message-passing graph neural networks that learn adaptive, data-dependent aggregation weights over a node's neighborhood. Let $\mathbf{h}_i \in \mathbb{R}^F$ denote the input feature of node i, and let $\mathcal{N}_i$ be its (one-hop) neighborhood. A GAT layer first applies a shared linear transformation $\mathbf{W} \in \mathbb{R}^{F' \times F}$ and then computes *masked* self-attention coefficients only for neighbors $j \in \mathcal{N}_i$. In the original formulation, an attention logit is obtained from the concatenated transformed features using a shared weight vector $\mathbf{a} \in \mathbb{R}^{2F'}$:

$$e_{ij} = a(\mathbf{W}\mathbf{h}_i, \mathbf{W}\mathbf{h}_j) = \mathrm{LeakyReLU}\left(\mathbf{a}^\top \left[\mathbf{W}\mathbf{h}_i \parallel \mathbf{W}\mathbf{h}_j\right]\right), \tag{1}$$

followed by neighborhood-wise softmax normalization,

$$\alpha_{ij} = \mathrm{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik})}. \tag{2}$$

The output feature of node i is then computed by an attention-weighted aggregation of transformed neighbor features:

$$\mathbf{h}'_i = \sigma\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}\, \mathbf{W}\mathbf{h}_j\right), \tag{3}$$

where $\sigma(\cdot)$ denotes a nonlinearity.

In waterway systems, the influence of a neighbor is not defined only by similarity but by physical context instead such as along-channel distance, elevation, and flow orientation. Thus, a natural extension is therefore to *condition the attention logits on edge attributes* $\mathbf{e}_{ij}$. In our model, we adopt this edge-conditioned attention principle so that message passing can emphasize hydrologically consistent connections (e.g., toward the target, short along-waterway distance) and suppress weak or non-physical relations, while still operating on a unified local graph per virtual station.

3.2. Overall architecture

Building on attention-based message passing as in Graph Attention Networks (Section 3.1), we propose an end-to-end spatiotemporal framework that combines (i) an MLP encoder for precipitation, (ii) a GNN operating on a local waterway graph to model spatial interactions, and (iii) a GRU head to capture temporal dynamics. An overview of the proposed computation flow is illustrated in Fig. 4. Let $\mathbf{y}_i = [y_{i,1}, y_{i,2}, \ldots, y_{i,T}]$ denote the vector of water-level observations at location i over time, and let $\mathbf{p}_i = [p_{i,1}, p_{i,2}, \ldots, p_{i,T}]$ denote the corresponding precipitation sequence. First, for each location i, $\mathbf{p}_i$ is fed into an MLP to produce a coarse latent representation $\widehat{\mathbf{f}}_i^{\mathrm{coarse}}$. This step provides a nonlinear mapping from raw precipitation inputs to a compact feature space that is used by the subsequent spatiotemporal modules.

Second, the coarse representations are refined on a river graph using a GNN together with spatial relationship features $\mathbf{F}_E^{\mathrm{spatial}}$. In our setting, the model operates on a *local* graph constructed for each virtual station. For a given virtual station v, we build a graph $\mathcal{G}_v = (\mathcal{V}_v, \mathcal{E}_v)$ by (i) selecting neighboring stations of v, (ii) identifying the waterway paths that connect each selected neighbor to v whenever such a connection exists, and (iii) including the shared junctions and intersections of these paths as additional nodes. The node set $\mathcal{V}_v$ therefore contains the virtual station v, its selected neighboring stations, and the intersection nodes along the connecting waterway paths, while $\mathcal{E}_v$ follows the waterway segments along these paths, preserving the geometry of the underlying network (Fig. 6). The proposed network is trained on the collection of graphs constructed from the training stations. During evaluation, inference is performed by applying the trained model to the graph $\mathcal{G}_v$ associated with each virtual station in the test set.

For message passing, the edge set is treated as *undirected*, so $\mathcal{E}_v$ specifies which pairs of locations exchange information. Directional and connectivity information is not represented by edge orientation, but is encoded in the associated edge features $\mathbf{f}_{(u,v)}^{\mathrm{spatial}}$ (Table 4). In particular, the *flow direction* feature takes values 1, −1, or 0 to indicate flow toward the virtual station, flow away from the virtual station, or the absence of a hydrological link, respectively. This design allows the model to include both hydrologically connected neighbors and broader spatial neighbors in a unified graph, while enabling the edge-attention mechanism to learn when and how each edge should contribute.

Table 1
Structure of the MLP used for precipitation feature extraction and for residual-weight prediction.

Layer	Operation	Output Dim.
Input	Scalar precipitation $p_{i,t}$	1
Layer 1	Linear + GELU	8
Layer 2	Linear	8

Third, the GNN output at each node is passed through a prediction module to obtain baseline water-level predictions. Concretely, the GNN produces a sequence of node representations $\mathbf{h}_{i,1:T}$, where each $\mathbf{h}_{i,t}$ summarizes information aggregated from neighboring stations at time t according to the learned attention weights and edge attributes. We then apply a GRU along the temporal dimension for each node,

$$\mathbf{z}_{i,1:T} = \mathrm{GRU}(\mathbf{h}_{i,1:T}), \tag{4}$$

followed by a nonlinear activation and a linear projection to obtain the baseline prediction sequence,

$$\widehat{\mathbf{y}}_{i,1:T}^{\mathrm{base}} = \mathrm{Linear}(\mathrm{GELU}(\mathbf{z}_{i,1:T})). \tag{5}$$

Here, the GNN performs spatial mixing using the river-network topology and edge features, while the GRU captures temporal dependencies and memory effects in the water-level dynamics based on the spatially refined representations.

Fourth, to account for residual errors that remain after baseline estimation, we compute residuals $\mathbf{r}_i$ at gauged stations as the differences between the observed values $\mathbf{y}_i$ and the corresponding baseline predictions $\widehat{\mathbf{y}}_i^{\mathrm{base}}$. These residuals capture local discrepancies and are subsequently used to estimate the expected error at the virtual station, which is then applied to correct and refine its water-level prediction.

Finally, a second MLP generates interpolation weights $\mathbf{w}_v$ for each virtual station based on the spatial relationship features between neighboring gauged stations and the virtual station. These weights determine the contribution of residuals from neighboring gauged stations, yielding an interpolated residual $\tilde{\mathbf{r}}_v$ at the virtual station. The interpolated residual is then added to the baseline prediction to yield the final adjusted estimate $\widehat{\mathbf{y}}_v$.

3.3. Water level regression from precipitation

We employ a lightweight multilayer perceptron (MLP) to extract features from precipitation inputs. The MLP consists of two fully connected layers with GELU activations (Hendrycks and Gimpel, 2016) (Table 1). At each time step t, the MLP maps the precipitation value $p_{i,t}$ at location i to a coarse feature representation $\widehat{\mathbf{f}}_{i,t}^{\mathrm{coarse}}$. These coarse features provide the node-wise signals that will be propagated spatially over the waterway network.

To enable spatial propagation, we represent the formulation waterway network as a graph (a simplified example can be found in Fig. 5). For each virtual station v (i.e., a prediction target), we construct a *local* graph $\mathcal{G}_v = (\mathcal{V}_v, \mathcal{E}_v)$ that captures the relevant network context around v (Fig. 6). The node set $\mathcal{V}_v$ includes the virtual station, its selected neighboring telemetry stations, and intermediate junction or intersection nodes along waterway paths that connect these neighbors to v whenever such connections exist. The edge set $\mathcal{E}_v$ follows the corresponding waterway segments along these paths, preserving the

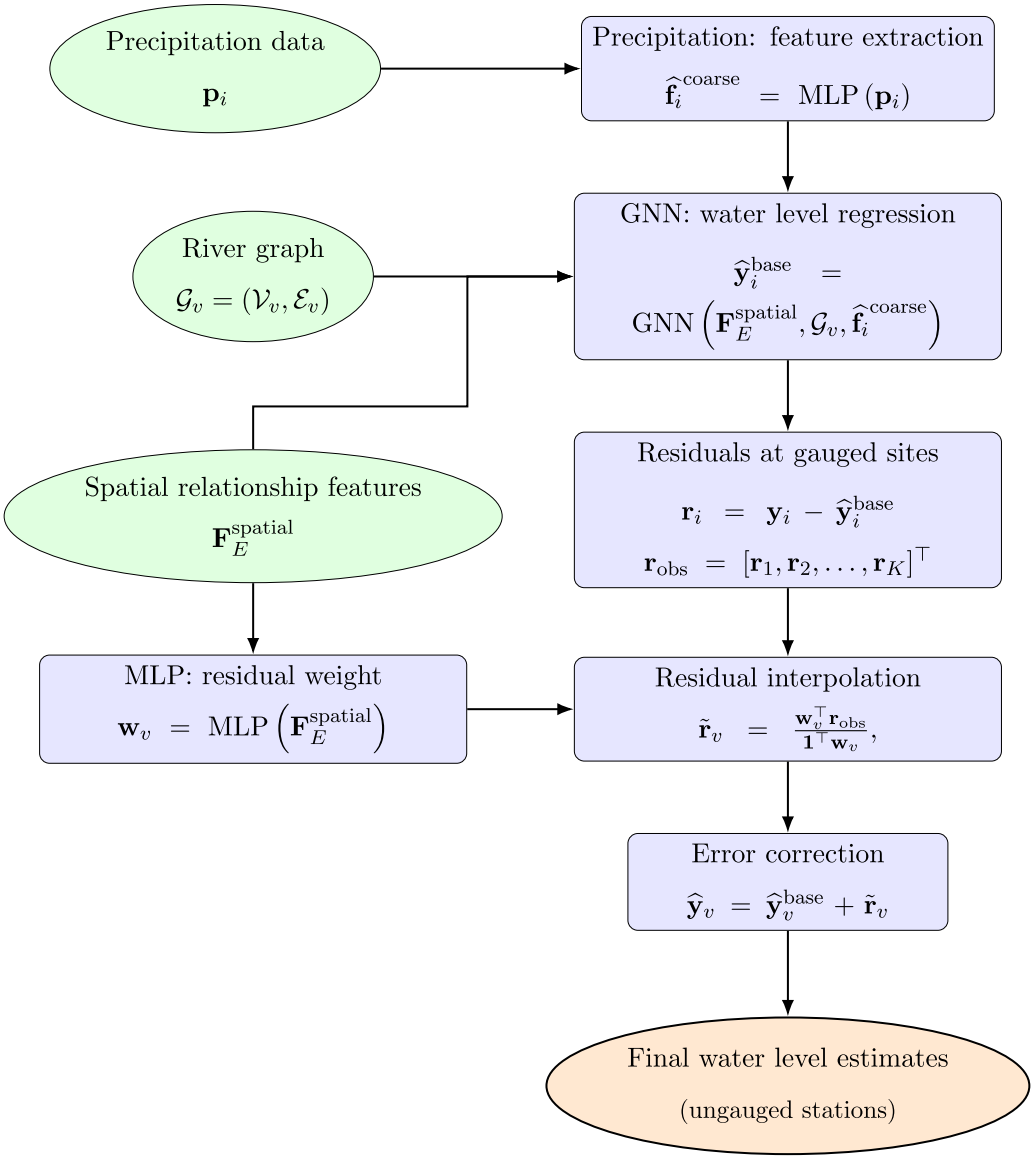


Fig. 4. Overview of the computation flow.

geometry of the underlying network. Each edge $(u, v) \in \mathcal{E}_v$ is associated with spatial relationship features $\mathbf{e}_{uv}$ that encode hydrological connectivity and topographic context (Table 4).

Given the coarse precipitation features and the corresponding local graph $\mathcal{G}_v$, we refine the node representations using a GNN with attention-weighted message passing. Intuitively, this stage mixes information across stations according to the waterway topology and the edge attributes, producing spatially refined features that are then passed to the temporal prediction head to generate baseline water-level estimates over time.

Let $\mathbf{f}_{i,1:T}^{\text{coarse}}$ denote the coarse feature sequence of station i produced by the MLP. For message passing, we treat $\mathcal{E}_v$ as *undirected*, and encode directional and connectivity information in the edge attributes $\mathbf{e}_{uv}$. An edge-attention network maps $\mathbf{e}_{uv}$ to a scalar logit s_{uv} (Table 2), which is normalized over the neighborhood of each node to obtain attention weights

$$\alpha_{uv} = \frac{\exp(s_{uv})}{\sum_{u' \in \mathcal{N}(v)} \exp(s_{u'v})}. \tag{6}$$

At each time step t, node features are updated by aggregating attention-weighted messages from neighbors,

$$\mathbf{m}_{v,t} = \sum_{u \in \mathcal{N}(v)} \alpha_{uv}\, \mathbf{f}_{u,t}^{\text{coarse}}, \qquad \mathbf{h}_{v,t} = \mathbf{f}_{v,t}^{\text{coarse}} + \mathbf{m}_{v,t}, \tag{7}$$

where the residual connection stabilizes learning and preserves local precipitation information. Directional and hydrological effects are introduced through $\mathbf{e}_{uv}$; for example, edges with flow-direction $= 1$ can be emphasized, edges with flow-direction $= -1$ can be down-weighted, and edges with flow-direction $= 0$ can be suppressed by the learned attention, without requiring a directed adjacency. The refined sequence $\mathbf{h}_{v,1:T}$ is then passed to the GRU-based prediction head to model temporal dependencies and produce $\widehat{\mathbf{y}}_{v,1:T}^{\text{base}}$.

Finally, the updated node representations are passed through a prediction head to obtain baseline water-level predictions. The prediction head consists of a GRU (Cho et al., 2014) applied along the temporal dimension, followed by a GELU activation and a linear projection to a

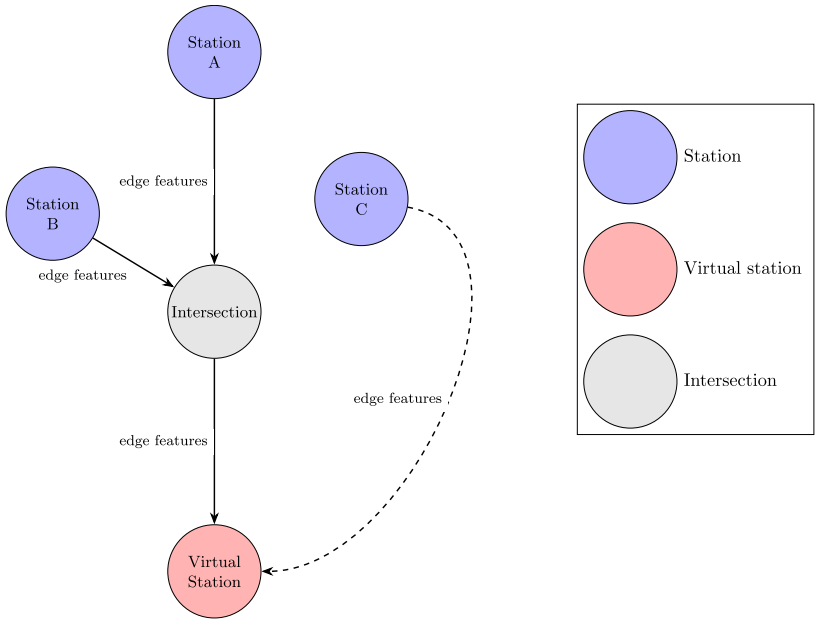


Fig. 5. Schematic illustration of the waterway network as a graph for GNN-based learning. Blue nodes represent telemetry stations with water-level observations, the gray node denotes a waterway intersection, and the red node indicates a virtual station where water levels are to be estimated. Solid lines show candidate connections along the main waterway, while the dashed line represents a nearby station that is not physically linked via the main waterway but is included to capture broader spatial relationships.

Table 2
Edge-attention network used to compute per-edge logits s_{uv} from edge attributes $\mathbf{e}_{uv}$. Here d_e is the edge feature dimension and E is the number of edges.

#	Layer	Input shape	Output shape
1	Linear	$[E, d_e]$	$[E, 16]$
2	GELU	$[E, 16]$	$[E, 16]$
3	Linear	$[E, 16]$	$[E, 16]$
4	GELU	$[E, 16]$	$[E, 16]$
5	Linear	$[E, 16]$	$[E, 1]$

Table 3
Architecture of the prediction head. T denotes the temporal length and N the number of nodes.

Layer	Input shape	Output shape
GRU	$[N, T, 8]$	$[N, T, 8]$
GELU	$[N, T, 8]$	$[N, T, 8]$
Linear	$[N, T, 8]$	$[N, T, 1]$

scalar output. The architecture of this prediction head is summarized in Table 3.

Each edge $(u, v) \in \mathcal{E}_v$ represents a candidate spatial or hydrological relationship between two locations and is treated as undirected for message passing. To preserve directional and connectivity information without enforcing a directed adjacency, we encode hydrological orientation and linkage directly in the edge attributes. Specifically, the *flow direction* feature takes values 1 if the waterway flow is toward the virtual station, −1 if it is away from the virtual station, and 0 if no valid waterway connection exists between the two locations. This design allows the model to include nearby stations even when no direct waterway path exists, while enabling the attention mechanism to emphasize hydrologically consistent connections when appropriate.

To capture physical and topographical relationships, we associate each edge with geometric descriptors such as the distance along the water path and the displacement vector, as well as terrain-based attributes computed from the DTM of Flanders (Agentschap Digitaal Vlaanderen, 2019). Elevation values are sampled along each waterway segment at regular intervals to compute summary statistics (mean, standard deviation, median, minimum, and maximum), and we include the elevation difference between endpoints. Finally, we incorporate a quantized river-width attribute extracted from OpenStreetMap (Geofabrik GmbH, 2025). Together, these edge features provide a compact representation of hydrological connectivity, topography, and channel characteristics that can be leveraged by the attention mechanism during message passing.

Table 4
Summary of edge features used in the graph model. Edges are treated as *undirected*; directional and connectivity information (toward, away, not connected) is encoded in the *flow direction* feature.

Feature name	Description
Distance along water path	Length of the waterway segment from source to target node.
Displacement vector	2D vector from source to target node coordinates.
Mean elevation	Mean elevation along the connecting segment.
Std elevation	Standard deviation of elevation along the segment.
Median elevation	Median elevation along the segment.
Min elevation	Minimum elevation along the segment.
Max elevation	Maximum elevation along the segment.
Flow direction	Hydrological flow indicator (1: toward virtual station, 0: not connected, −1: reversed).
Delta elevation	Elevation difference between source and target nodes.
Quantized width	Quantized width information of the waterways.

3.4. Error correction

After obtaining baseline predictions $\hat{\mathbf{y}}_i^{\text{base}}$, we apply a residual-based correction stage to further improve accuracy at virtual (ungauged) stations. This stage is inspired by the error-propagation idea of Nagahama et al. (2024), but is implemented within our learned framework so that residual transfer is guided by spatial relationship features rather than a fixed geostatistical model.

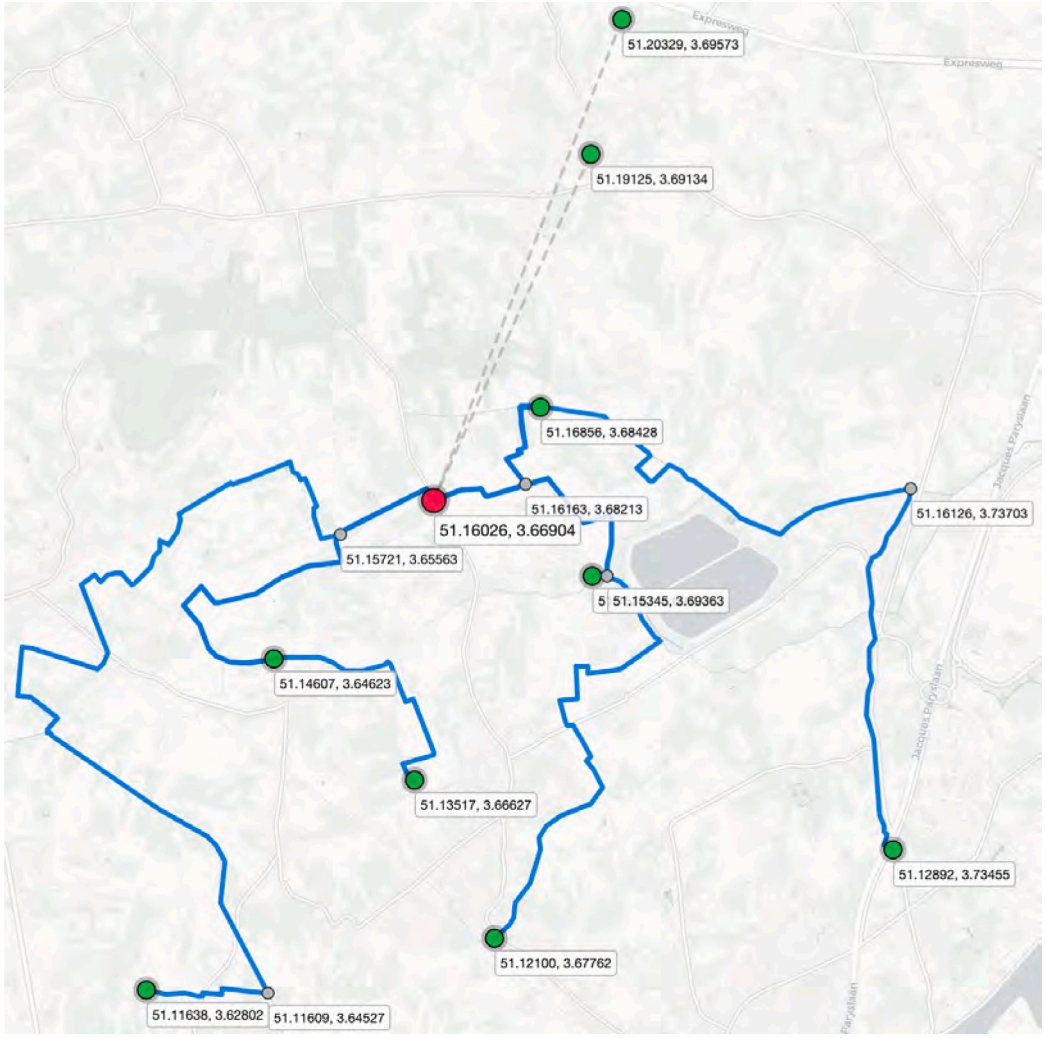


Fig. 6. Example local graph $\mathcal{G}_v$ constructed for a single virtual station v. The red node denotes the virtual station (prediction target) and green nodes denote its selected neighboring stations. Colored polylines follow the detailed waterway paths used to form edges, and gray markers indicate intermediate nodes along these paths (e.g., junctions and intersections).

Residual computation. For each gauged station i, we compute the residual

$$\mathbf{r}_i = \mathbf{y}_i - \hat{\mathbf{y}}_i^{\text{base}}, \tag{8}$$

where $\mathbf{y}_i$ denotes the observed water-level sequence at station i. The residual sequence $\mathbf{r}_i$ represents the station-specific discrepancy that remains after the baseline model, capturing local effects that are not fully explained by precipitation-driven prediction and spatial message passing.

Residual interpolation. To estimate the remaining prediction error at a virtual station, we transfer residual information from nearby gauged stations. Specifically, we use a second MLP to convert the spatial relationship features between each gauged neighbor and the virtual station into nonnegative interpolation weights. For a virtual station v, let $\mathbf{r}_{\text{obs}} = [\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_K]^\top$ denote the residual sequences at its K neighboring gauged stations. The model predicts a weight vector $\mathbf{w}_v \in \mathbb{R}^K$ and computes an interpolated residual as a normalized weighted combination,

$$\tilde{\mathbf{r}}_v = \frac{\mathbf{w}_v^\top \mathbf{r}_{\text{obs}}}{\mathbf{1}^\top \mathbf{w}_v}. \tag{9}$$

This formulation is analogous to distance-weighted interpolation, but replaces fixed distance kernels with learned weights that can adapt to hydrological connectivity, terrain attributes, and other edge features.

Error correction and final prediction. The interpolated residual is then added to the baseline prediction at the virtual station,

$$\hat{\mathbf{y}}_v = \hat{\mathbf{y}}_v^{\text{base}} + \tilde{\mathbf{r}}_v, \tag{10}$$

yielding the corrected estimate. In this way, systematic local errors observed at gauged stations are propagated to nearby virtual stations according to learned, feature-conditioned influence weights. The final output $\hat{\mathbf{y}}_v$ therefore integrates precipitation-driven baseline prediction, spatial refinement on the waterway graph, and residual-based error correction.

3.5. Optimization

The model was implemented in `PyTorch` (v2.7) with Python 3.11 and trained using the Adam optimizer. In Adam, β_1 and β_2 are the exponential decay rates for the moving averages of the first and second moments of the gradients, respectively. For all experiments in this work, we set $\beta_1 = 0.9$, $\beta_2 = 0.999$, and the learning rate to 10^{-3}.

Multi-target supervision. During training, we apply supervision not only at the virtual (target) station but also at the neighboring gauged stations available in the local graph. This provides additional learning signals and encourages the model to produce consistent predictions across connected nodes. Concretely, we optimize three terms: (i) the baseline

prediction at the virtual station, (ii) the corrected (final) prediction at the virtual station, and (iii) the baseline predictions at the gauged neighbors. The overall loss is

$$\mathcal{L} = \mathrm{MSE}\big(\hat{\mathbf{y}}_v^{\mathrm{base}}, \mathbf{y}_v\big) + \mathrm{MSE}\big(\hat{\mathbf{y}}_v, \mathbf{y}_v\big) + \sum_{j\in\mathcal{N}} \mathrm{MSE}\big(\hat{\mathbf{y}}_j^{\mathrm{base}}, \mathbf{y}_j\big), \tag{11}$$

where $\hat{\mathbf{y}}_v^{\mathrm{base}}$ and $\hat{\mathbf{y}}_v$ are the baseline and corrected predictions at the virtual station v, respectively, and $\hat{\mathbf{y}}_j^{\mathrm{base}}$ denotes the baseline prediction at a neighboring gauged station $j \in \mathcal{N}$. The corresponding observation sequences are $\mathbf{y}_v$ and $\mathbf{y}_j$.

4. Experiments

4.1. Experimental setup

In our experiments, the final dataset includes 97 training stations and 112 test stations. For each virtual sensor location, we identify and select a set of neighboring measurement stations located within a fixed spatial radius of 15 km. This radius-based selection reflects a realistic interpolation scenario, where information from nearby observed stations is used to estimate water levels at uninstrumented or sparsely monitored locations. The number of neighboring stations per virtual sensor varies depending on the local station density, typically ranging from 3 to 10. This variability allows the model to adapt to both dense and sparse regions of the monitoring network, thereby improving generalizability.

4.2. Assessment metrics

We evaluate model performance using two complementary metrics: the Pearson correlation coefficient and the mean squared error (MSE). All metrics are computed over the full one-year evaluation period (January–December 2024). The Pearson correlation coefficient is computed between the predicted and observed sequences over the entire time series. It measures how well the model reproduces the temporal dynamics and overall trends of water-level variation, largely independent of any constant offset.

To quantify absolute accuracy, we also report the MSE. Because prediction difficulty can vary with hydrological conditions, we additionally compute MSE within three ranges of observed (normalized) water level: (i) 0–0.3 m, (ii) 0.3–0.6 m, and (iii) 0.6–1.2 m. This stratification allows us to separate performance under low-level conditions from periods of larger fluctuations, and to examine whether errors concentrate at higher water levels where rapid changes are more likely to occur.

4.3. Baseline methods

To assess the effectiveness of the proposed approach, we compare it with a set of established baselines for spatial interpolation, including classical geostatistical methods, a learning-based weighting model, and a Gaussian-process-based error-correction method.

- **IDW:** Inverse Distance Weighting estimates the value at a target location as a weighted average of neighboring observations, with weights decaying with spatial distance. It is simple and widely used, but it does not explicitly model spatial correlation beyond distance-based similarity.
- **OK:** Ordinary Kriging is a geostatistical interpolation method that models spatial dependence through a variogram. The variogram characterizes how similarity between observations decreases with separation distance, enabling statistically grounded interpolation. OK is commonly used in environmental applications and serves as a strong reference baseline.
- **KED:** Kriging with External Drift extends kriging by incorporating an external covariate that is available at both observed and target locations. In our experiments, rainfall is used as the drift variable so that the mean water-level field can vary with rainfall, while residual spatial dependence is captured by a covariance model. To keep inference efficient for long time series, we cache geometry-dependent components (e.g., neighbor covariance terms) and update only the drift-related terms over time.
- **OCK:** Ordinary CoKriging jointly models the primary variable (water level) and an auxiliary variable (rainfall) to exploit both spatial correlation and cross-correlation between variables. We adopt an intrinsic coregionalization model (ICM) to ensure a valid cross-covariance structure.

MLP weight prediction (learning-based IDW). We also include a learning-based baseline that predicts interpolation weights from the same edge features used in our method (Tucci, 2023). Concretely, an MLP takes as input the edge features between the virtual station and each neighboring station and outputs a set of normalized weights. The virtual-station estimate is then computed as a weighted average of neighboring water-level observations. This baseline evaluates how much can be gained by learning adaptive weights from engineered edge features, without using graph message passing or temporal decoding.

NNGP error correction. Finally, we include the method of Nagahama et al. (2024) as a Gaussian-process-based baseline. Their approach first predicts water levels from rainfall using a simple regression model and then applies a Nearest Neighbor Gaussian Process (NNGP) to propagate residual errors from gauged stations to ungauged locations. This baseline represents a scalable probabilistic framework for spatial error correction in large monitoring networks.

Best-possible weighted average (upper bound). In addition to the above baselines, we report an upper bound for weighted-average interpolation. For each virtual station and each evaluation segment, we compute the set of nonnegative weights over its K neighbors that minimizes the mean squared error with respect to the ground-truth target sequence, subject to a sum-to-one constraint. This yields the best possible *time-invariant* convex combination of neighbor observations for that segment. Although this upper bound uses the ground truth and is therefore not a deployable method, it provides a useful reference for quantifying the maximum achievable performance of any constant-weight interpolation scheme given the selected neighbors.

4.4. Performance analysis by water level brackets

To better understand performance across different hydrological conditions, we arrange the test samples by the ground-truth water level changes at each time step into three brackets: **low** (0–0.3 m), **medium** (0.3–0.6 m), and **high** (0.6–1.2 m). This approach is used only for evaluating and does not affect training. Table 5 reports the mean RMSE for each method within each bracket, as well as the overall RMSE aggregated over the full 0–1.2 m range.

Across the medium and high water-level change brackets, HIGNN achieves the lowest RMSE among deployable methods, with clear margins over classical interpolation, kriging-based baselines, and the MLP/NNGP alternatives. This indicates that combining waterway-aware spatial aggregation with temporal modeling is particularly beneficial when water levels exhibit larger variability and more dynamic behavior.

In contrast, in the low bracket (0–0.3 m), classical baselines remain more accurate and HIGNN yields higher RMSE than the best geostatistical and distance-based methods. This regime corresponds to smaller fluctuations where simple spatial interpolation is often sufficient and the advantage of a higher-capacity spatiotemporal model is less pronounced.

Table 5
RMSE by water-level bracket. Columns **Low**, **Med**, and **High** report mean RMSE in the ranges 0–0.3 m, 0.3–0.6 m, and 0.6–1.2 m, respectively. **Overall** reports RMSE over 0–1.2 m. Best and second best are computed *excluding* the upper-bound method; best is in **bold** and second best is underlined.

Method	Low	Med	High	Overall
IDW	0.0841	0.2369	0.4314	**0.1135**
OK	0.0873	0.2289	0.4487	0.1160
MLP	0.0859	0.2304	0.4482	0.1149
NNGP	0.0852	0.2265	0.4437	0.1137
KED	**0.0827**	0.2461	0.4610	0.1144
OKCK	0.0861	0.2419	0.4325	0.1158
HIGNN (ours)	0.0974	**0.2090**	**0.3649**	0.1193
Best-possible weighted average (upper bound)	0.0394	0.1368	0.2703	0.0585

When aggregating over the full 0–1.2 m range, the strongest overall RMSE is achieved by IDW/NNGP, while HIGNN is slightly worse in the aggregate despite its substantial gains in the medium and high brackets. Finally, the best-possible weighted-average upper bound is significantly lower than all practical methods in every bracket, suggesting that there remains headroom if a model can better approximate the optimal combination of neighboring stations.

Fig. 7 reports the average gain of HIGNN relative to each baseline within the three water-level brackets. For each time step, the gain is defined as the reduction in squared error achieved by HIGNN compared with a baseline,

$$\Delta = (y - \hat{y}_{\text{baseline}})^2 - (y - \hat{y}_{\text{HIGNN}})^2, \tag{12}$$

where y is the observed value, and $\hat{y}_{\text{baseline}}$ and $\hat{y}_{\text{HIGNN}}$ denote the baseline and HIGNN predictions, respectively. Positive values of Δ indicate that HIGNN achieves lower error than the baseline.

A consistent pattern emerges across all comparisons. In the low range $[0, 0.3]$, the gain is close to zero and slightly negative for most baselines, indicating that HIGNN is comparable to classical methods when water levels are small and variations are limited. In contrast, HIGNN shows clear positive gains in the medium $(0.3, 0.6)$ and high $[0.6, 1.2]$ ranges for every baseline. The improvements increase with water level magnitude: gains in the high bracket are substantially larger than those in the medium bracket, suggesting that HIGNN becomes increasingly advantageous as the dynamics become more pronounced.

The largest gains are observed against the simpler interpolation baselines (IDW and MLP), which rely on fixed or purely local weighting and therefore struggle to capture network-driven propagation effects under larger fluctuations. HIGNN also yields consistent improvements over the geostatistical baselines (OK, KED, and OCK) and over NNGP, indicating that the learned combination of waterway-aware spatial message passing and temporal modeling provides benefits beyond both classical variogram-based interpolation and GP-based residual propagation. Overall, these results align with the bracketed RMSE analysis: HIGNN's main advantage appears in the medium and high water level changes, where accurately modeling both hydrological connectivity and temporal evolution is most critical.

Figs. 8, 9, 10, 11, and 12 provide qualitative comparisons between the proposed HIGNN method, several established baselines (IDW, OK, MLP, NNGP, KED, and OCK), the interpolation upper bound, and the ground truth (GT) for five representative stations. The examples span diverse temporal behaviors, including sharp peak events (station 27 and station 96), step-like transitions with multiple regimes (station 80), a single dominant pulse followed by a decaying limb (station 15), and long, structured fluctuations with alternating plateaus (station 5). Each plot visualizes hourly water-level variations relative to the station median, enabling direct comparison of timing and magnitude across methods.

Table 6
Overall Pearson correlation coefficient (ρ) between predicted and observed water levels at virtual stations for each method.

Method	Correlation (ρ)	Std. Dev.
IDW	0.7431	0.3033
NNGP	0.7522	0.2586
OK	0.7410	0.2555
MLP	0.7462	0.2558
KED	0.7274	0.2357
OCK	0.6984	0.2484
HIGNN	**0.7040**	**0.2690**

Across all cases, classical spatial baselines (IDW and OK) generally capture the overall trend but tend to smooth rapid transitions and underestimate peaks. Learning-based baselines (MLP and NNGP) are more flexible but can misestimate peak magnitudes and sometimes introduce overshoot or temporal lag. KED and OCK show larger deviations in several examples, including pronounced positive bias in station 80 and negative offset behavior in station 15, highlighting sensitivity to the drift.

HIGNN more consistently tracks the onset and decay of events. For stations 27 and 96, HIGNN follows the main peak shape more closely than the baselines and typically remains near the interpolation upper bound during the rapid rise and subsequent recession. At station 15, HIGNN captures the main pulse timing well, but all methods exhibit larger discrepancies after the peak, suggesting reduced neighborhood informativeness during the post-event period. For station 5, HIGNN aligns well with GT across repeated transitions and plateaus, while IDW/OK remain overly smoothed and learning-only models show larger drift. Overall, these qualitative results complement the quantitative metrics: HIGNN yields clearer benefits in dynamic conditions where accurate peak timing and magnitude matter, and it tends to stay closest to the interpolation upper bound when the neighborhood information is sufficient.

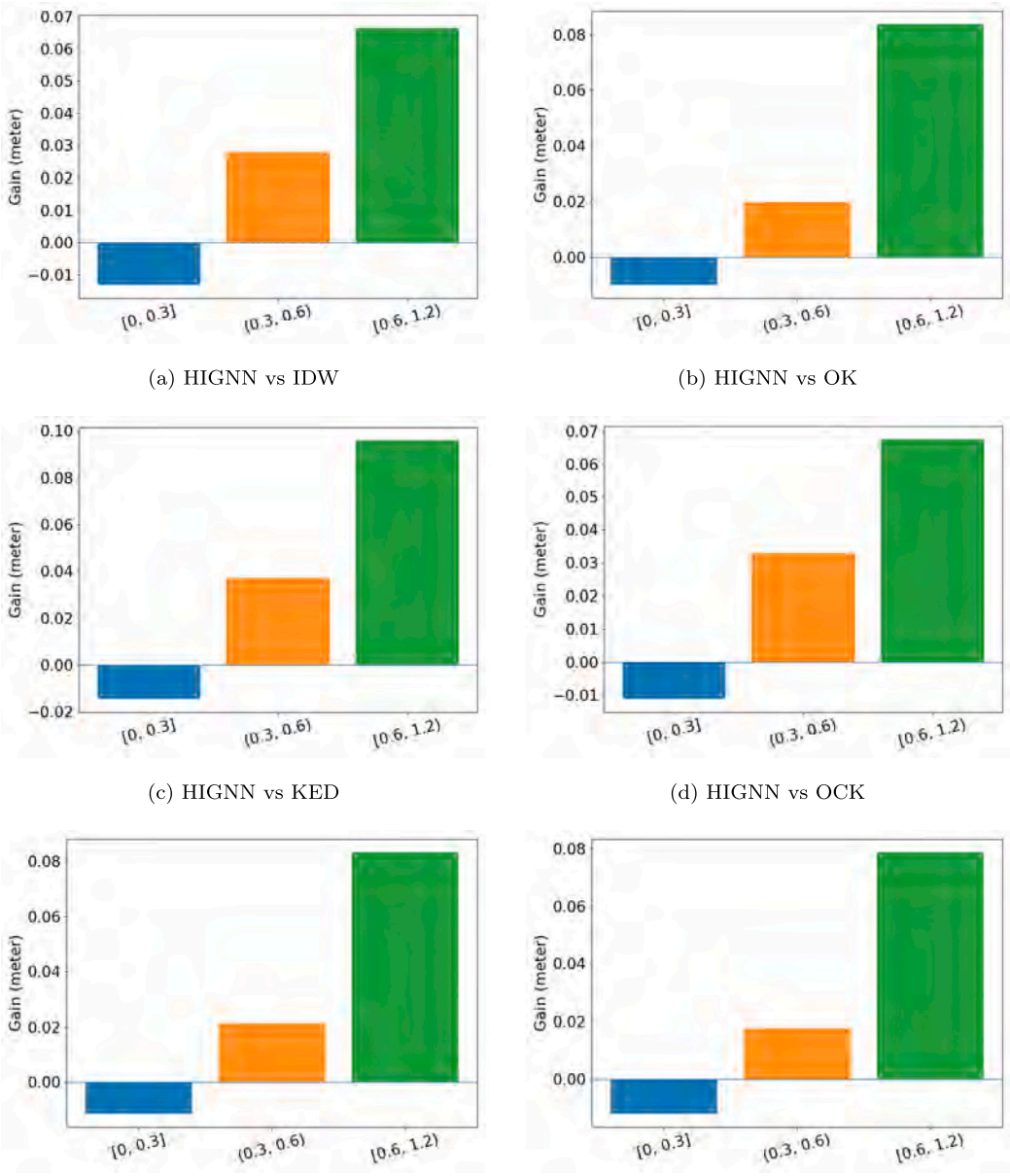
4.5. Temporal consistency of water level predictions

Table 6 summarizes the overall Pearson correlation coefficient (ρ) between predicted and observed water levels at virtual stations. The correlations are consistently high across all methods (approximately 0.70–0.75), indicating that both classical interpolation and learning-based approaches generally track the dominant temporal evolution of the water-level signal. Among the baselines, NNGP achieves the highest correlation ($\rho = 0.7522$), followed by MLP ($\rho = 0.7462$), IDW ($\rho = 0.7431$), and OK ($\rho = 0.7410$). KED remains competitive ($\rho = 0.7274$), while OKCK attains the lowest correlation in this comparison ($\rho = 0.6984$).

HIGNN achieves a comparable correlation of 0.7040, indicating that its predictions remain temporally consistent with the observed dynamics, though it does not maximize ρ relative to the strongest baselines. This aligns with the bracketed RMSE analysis: HIGNN is optimized to reduce amplitude-sensitive errors during higher-variability periods (medium and high change regimes), which can yield substantial RMSE improvements without necessarily increasing correlation, since ρ primarily reflects trend alignment rather than absolute error magnitude. Overall, the correlation results confirm that HIGNN preserves the temporal structure of the water-level signal while offering stronger accuracy gains under more dynamic hydrological conditions.

4.6. Statistical significance of error differences

To assess whether the performance differences between HIGNN and the baseline methods are statistically significant, we conduct two-sided paired t-tests on the per-sample squared errors within each water-level bracket. For each bin, we compare the error sequence produced by HIGNN with the error sequence produced by a baseline on the *same*



(a) HIGNN vs IDW (b) HIGNN vs OK

(c) HIGNN vs KED (d) HIGNN vs OCK

(e) HIGNN vs MLP (f) HIGNN vs NNGP

Fig. 7. Per-bin average gain of HIGNN against other methods (positive indicates lower error than the baseline).

set of samples. We compute the test statistics using the `ttest_rel` function from `scipy.stats`.

Because we perform multiple statistical tests (each baseline compared with HIGNN in each bin), we apply a multiple-testing correction to reduce false positives. Specifically, we adjust all p-values jointly using the Holm procedure via `statsmodels.stats.multitest.multipletests`, and declare significance at $\alpha = 0.05$ based on the adjusted p-values.

Table 7 reports the resulting t-statistics. In our implementation, the paired test is computed as $\texttt{ttest\_rel}(e_{\text{HIGNN}}, e_{\text{baseline}})$. Therefore, a *negative* t-statistic indicates that HIGNN has a lower mean squared error than the baseline (i.e., HIGNN performs better), whereas a *positive* t-statistic indicates the opposite. After Holm correction, all comparisons are statistically significant in all three bins.

The direction of the effects is consistent with the bracketed RMSE analysis. In the medium $(0.3, 0.6)$ and high $[0.6, 1.2)$ brackets, HIGNN significantly outperforms every practical baseline, as reflected by large negative t-statistics across IDW, OK, MLP, NNGP, KED, and OKCK. This confirms that HIGNN provides robust error reductions during moderate to high water levels, where water-level dynamics are more pronounced and exploiting waterway-aware spatial structure and temporal modeling is most beneficial.

In the low bracket $[0, 0.3]$, the t-statistics are positive for all baselines, indicating that classical methods achieve lower mean squared error than HIGNN in this case. This behavior is expected in periods of small water levels and limited variation, where simple local interpolation can be sufficient and the advantage of a higher-capacity spatiotemporal model is less pronounced. Finally, the best-possible

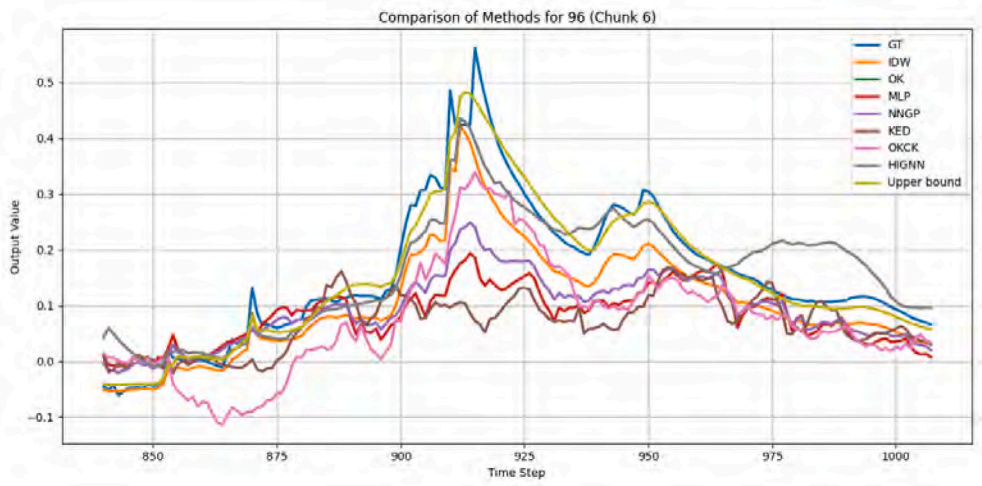


Fig. 8. Station 96 (GPS: 51.181709°N, 3.460833°E): Comparison of ground truth and different methods for a sample at this station (Chunk 6). Each time step corresponds to one hour. The output represents water level variations with respect to the median value, i.e., the median water level at each station has been subtracted.

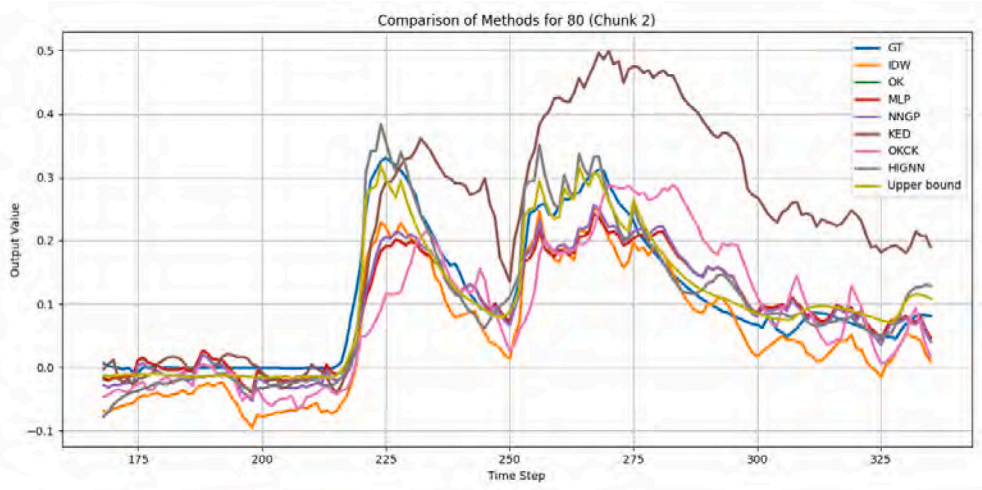


Fig. 9. Station 80 (GPS: 51.109942°N, 3.769376°E): Comparison of ground truth and different methods for a sample at this station (Chunk 2). Each time step corresponds to one hour. The output represents water level variations with respect to the median value, i.e., the median water level at each station has been subtracted.

weighted-average upper bound is significantly better than HIGNN in all bins (large positive t), as expected since it is computed using the ground truth and serves only as a reference rather than a deployable method.

4.7. Evaluation on interpolation-feasible subset

In practice, not every virtual station in the test set is equally amenable to neighbor-based interpolation. Some targets have weak, sparse, or hydrologically inconsistent neighbors, so even an ideal weighted average of the available stations cannot approximate the target well. To separate these *neighborhood-limited* cases from those where interpolation is plausible, we perform an additional analysis on an *interpolation-feasible subset* of the test samples.

We use the best-possible weighted-average upper bound as a diagnostic of neighborhood quality. For each time step and virtual station, let $\hat{y}_{\mathrm{UB}}$ denote the prediction of the best-possible weighted average (computed from the available neighboring stations with weights chosen to minimize instantaneous error) and y the ground truth. Before constructing the feasible subset, we first use $\hat{y}_{\mathrm{UB}}$ to quantify a fundamental limitation of neighbor-based interpolation for large-change events. Fig. 13 reports the *event recall* for large-change events, defined as ground-truth peaks with water-level change $y \geq 0.6\,\mathrm{m}$. A predicted event is counted as successfully recalled if it contains a peak within a ± 3 hour window of the ground-truth peak time and the peak magnitude matches within a tolerance of 0.1 m, i.e.,

$$\exists\, t' \in [t^\star - 3,\, t^\star + 3] \text{ s.t. } |\hat{y}(t') - y(t^\star)| \leq 0.1 \text{ m}, \tag{13}$$

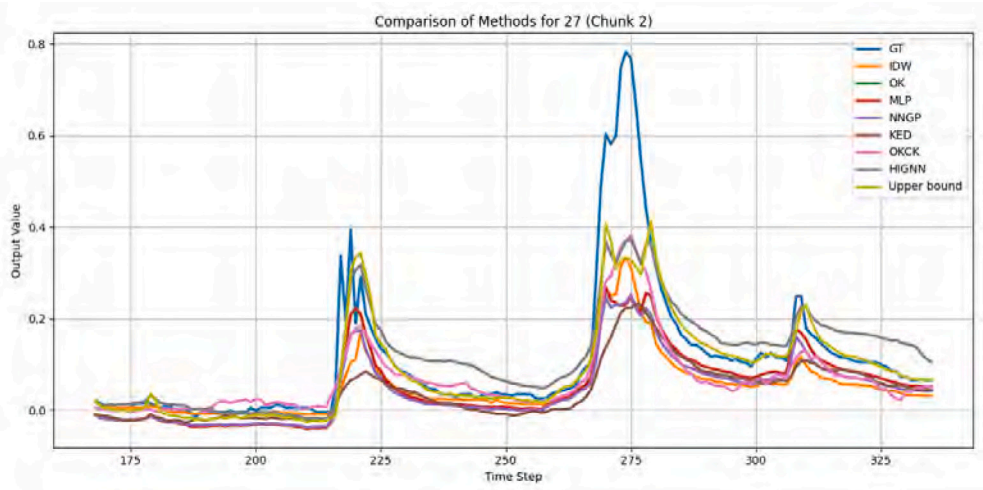


Fig. 10. Station 27 (GPS: 50.869533°N, 4.215837°E): Comparison of ground truth and different methods for a sample at this station (Chunk 2). Each time step corresponds to one hour. The output represents water level variations with respect to the median value, i.e., the median water level at each station has been subtracted.

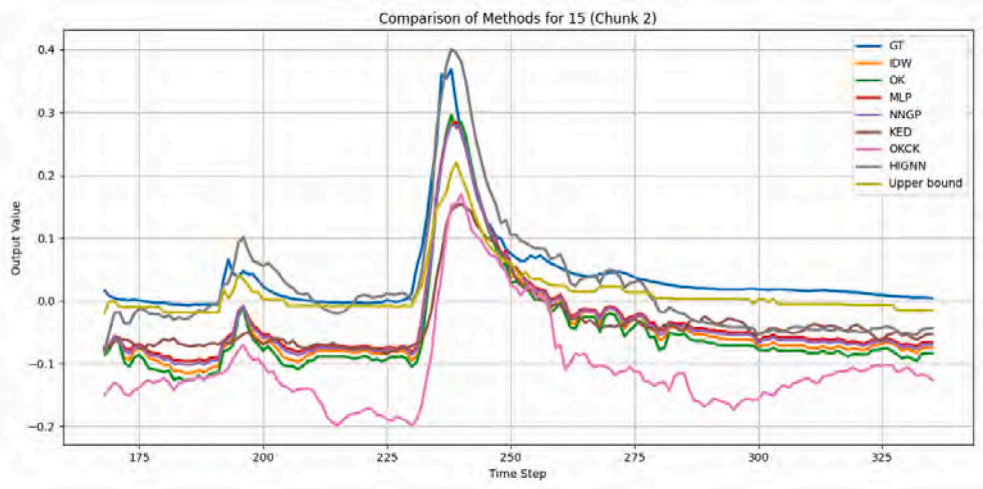


Fig. 11. Station 15 (GPS: 50.811944°N, 3.777113°E): Comparison of ground truth and different methods for a sample at this station (Chunk 2). Each time step corresponds to one hour. The output represents water level variations with respect to the median value, i.e., the median water level at each station has been subtracted.

Table 7
Paired t-test statistics comparing per-sample squared errors of HIGNN with baseline methods across water-level brackets (in meters). The tests are computed as $\texttt{ttest\_rel}(e_{\text{HIGNN}}, e_{\text{baseline}})$. Negative t indicates that HIGNN has lower mean squared error than the baseline; positive t indicates higher mean squared error. All results are statistically significant after Holm correction ($\alpha = 0.05$).

Comparison	[0, 0.3]	(0.3, 0.6)	[0.6, 1.2)
HIGNN vs. IDW	24.061	−23.065	−28.746
HIGNN vs. OK	33.313	−18.035	−35.431
HIGNN vs. MLP	31.338	−18.188	−34.351
HIGNN vs. NNGP	34.452	−15.787	−35.393
HIGNN vs. KED	31.001	−23.537	−31.705
HIGNN vs. OKCK	25.838	−22.041	−23.969
HIGNN vs. Best (upper bound)	134.555	56.237	35.682

where $t^\star$ denotes the time index of the ground-truth peak. Notably, even the upper bound attains only about 0.6 recall, indicating that roughly 40% of large-change events cannot be reproduced from the available neighbors using any weighted-average interpolation. This motivates evaluating methods separately on samples where the neighborhood is sufficiently informative.

We retain only samples for which this upper bound attains a small absolute error,

$$|\hat{y}_{\text{UB}} - y| \le \tau, \tag{14}$$

with $\tau = 0.1\,\text{m}$. This filtering step selects samples for which the available neighbors are sufficiently informative so that interpolation-based approaches are meaningful. We then recompute RMSE for each method on this same filtered subset, and report results by water-level bracket.

Table 8 highlights that the largest differences emerge in the high bracket $[0.6, 1.2)$, where interpolation is both more challenging and

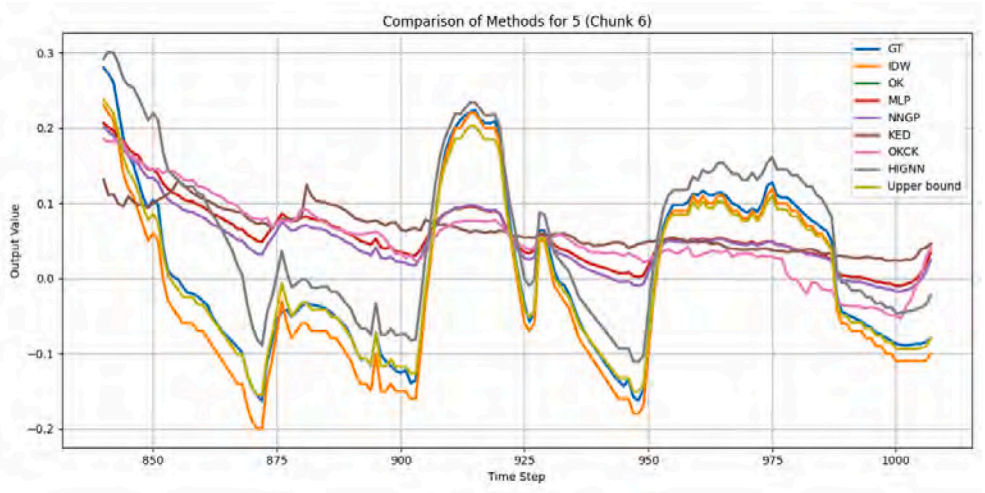


Fig. 12. Station 5 (GPS: 50.762012°N, 3.867335°E): Comparison of ground truth and different methods for a sample at this station (Chunk 6). Each time step corresponds to one hour. The output represents water level variations with respect to the median value, i.e., the median water level at each station has been subtracted.

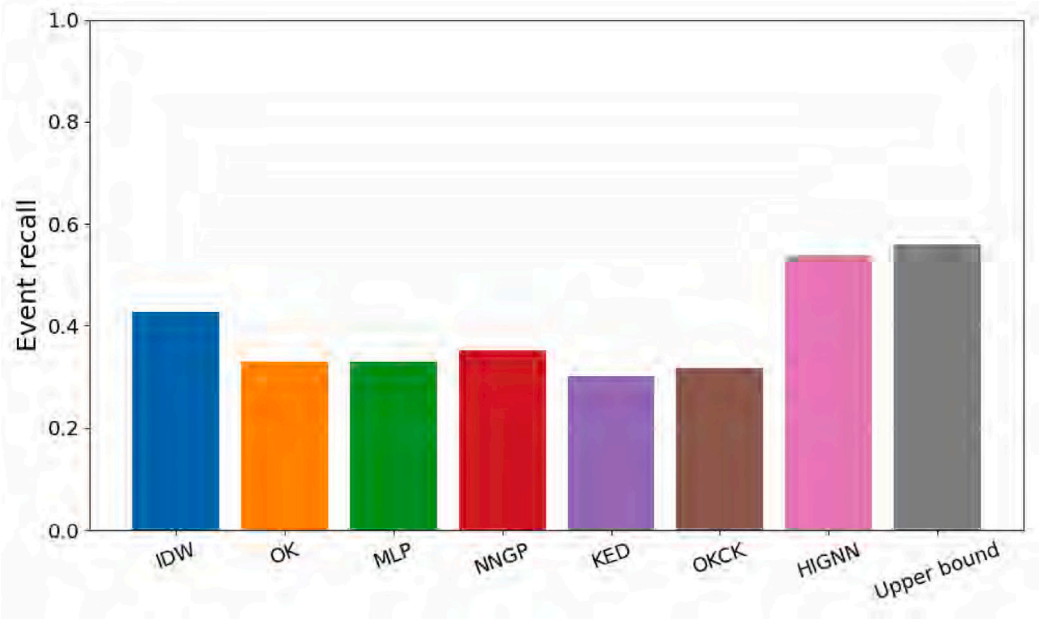


Fig. 13. Event recall on the interpolation-feasible subset for large-change events ($y \geq 0.6$ m). A prediction is counted as a hit if its absolute error is within 0.1 m. "Upper bound" denotes the best-possible weighted average computed using ground truth to select neighbor weights.

more consequential. In this case, HIGNN achieves the lowest RMSE among all deployable methods (0.187 m), outperforming the strongest classical baselines such as IDW (0.226 m), as well as kriging-based approaches (e.g., OK: 0.292 m; and KED: 0.290 m), and learned-weight baselines (MLP: 0.263 m; and NNGP: 0.259 m). These results indicate that, when the neighborhood contains informative stations, HIGNN can better exploit waterway-aware connectivity and temporal context to reduce errors under higher water levels, compared with methods that rely primarily on fixed spatial kernels or static interpolation weights.

We also evaluate whether methods can *detect large-change events*. Fig. 13 reports the event recall for cases with $y \geq 0.6$ m, where an event is counted as correctly captured if the prediction matches the magnitude within 0.1 m. HIGNN achieves higher recall than the classical and learned baselines, indicating improved sensitivity to large water-level changes when informative neighbors are available. As expected, the upper-bound weighted average attains the highest recall, reflecting the

Table 8
RMSE in the high water-level bracket $[0.6, 1.2)$ on the interpolation-feasible subset ($|\hat{y}_{\mathrm{UB}} - y| \leq 0.1$ m). Best and second best are computed *excluding* the upper-bound method; best is in **bold** and second best is underlined.

Method	RMSE
IDW	0.2265
OK	0.2924
KED	0.2896
OKCK	0.2751
MLP	0.2626
NNGP	0.2594
HIGNN (ours)	**0.1874**
Best-possible weighted average (upper bound)	0.0396

Table 9
Computational time comparison for different models (CPU only). Note that each segment corresponds to one week of data.

Method	Total elapsed time (s)	Average time per segment (ms)
IDW	0.381	0.823
MLP	0.426	0.919
OK	24.879	5.415
HIGNN	3.434	0.742
NNGP	1.232	2.662
KED	7.580	1.637
OCK	4.941	1.067

remaining intrinsic limitations imposed by neighborhood information even after filtering.

4.8. Computational time

Table 9 reports CPU runtimes on an Intel Core i7-8086K (4.00 GHz) for processing the full evaluation set, together with the average latency per weekly segment. Since all methods run fast enough for offline evaluation, this comparison mainly reflects relative scalability when inference must be repeated many times (e.g., across many virtual stations or frequent re-runs). IDW and the MLP baseline are the most efficient, requiring 0.823 ms and 0.919 ms per segment, respectively. HIGNN remains similarly lightweight, with an average of 0.742 ms per segment, indicating that its graph-based spatial modeling can be executed with low overhead in our setting. Among geostatistical baselines, OCK (1.067 ms) and KED (1.637 ms) are moderately more expensive, while OK has the highest latency (5.415 ms per segment) due to repeated kriging solves. NNGP is also slower than the lightweight baselines (2.662 ms), but remains substantially faster than OK. Overall, HIGNN achieves a favorable accuracy–efficiency trade-off, delivering strong performance in the most consequential change ranges while maintaining sub-millisecond per-segment latency.

4.9. Ablation on different components

To better understand the contribution of each component in HIGNN, we design the following ablation versions:

- **V1 (Full HIGNN)**: The complete proposed model, including precipitation feature extraction with MLP, water-level regression with a GNN enhanced by a GRU, and residual interpolation with an MLP.
- **V2 (HIGNN–IDW)**: Replaces the residual interpolation MLP with IDW. This tests whether learned interpolation provides benefits over a simple distance-based heuristic.
- **V3 (HIGNN–MLP)**: Replaces the GRU in the GNN with a feedforward MLP. This ablation evaluates the role of temporal modeling in capturing water-level dynamics.
- **V4 (MLP only)**: Removes the GNN entirely, predicting water levels directly from precipitation with an MLP. This serves as a simplified baseline to assess the value of graph-based spatial modeling.
- **V4 (HIGNN–NON)**: Uses the same architecture as V1 but *removes topographic edge features* from the graph. Specifically, we exclude elevation-based attributes (mean, std, median, min, max elevation, and Δ elevation) and retain only non-topographic edge information (distance along water path, displacement vector, flow direction, and quantized width). This ablation isolates the contribution of terrain-derived edge features.

Table 10
Ablation study of HIGNN variants. Mean RMSE at virtual stations for different water-level-change brackets. Best and second best are computed *excluding* the upper-bound method; best is in **bold** and second best is underlined.

Version	0–0.3 m	0.3–0.6 m	0.6–1.2 m	0–1.2 m
V1 (Full HIGNN)	**0.0974**	0.2090	**0.3649**	**0.1193**
V2 (HIGNN–IDW)	0.2144	**0.2008**	0.4368	0.2213
V3 (HIGNN–MLP)	0.0983	0.2097	0.3953	0.1214
V4 (MLP only)	0.1561	0.2135	0.5539	0.1771
V5 (HIGNN–NON)	0.1236	0.2036	0.3749	0.1415
Upper bound	0.0394	0.1368	0.2703	0.0585

Overall RMSE across brackets. Table 10 reports RMSE across water-level-change brackets. In the high bracket (0.6–1.2 m), the full model (V1) achieves the lowest RMSE (0.3649), indicating that combining waterway-aware spatial mixing, temporal recurrence (GRU), and learned residual correction is most effective when changes are large. Removing the nonlinear/topographic edge features (V5, HIGNN–NON) increases error in all brackets, most notably in the low bracket (0.1236 vs. 0.0974), and also degrades performance in the high bracket (0.3749 vs. 0.3649), suggesting that these features provide complementary cues beyond geometric distance and connectivity.

Replacing the residual interpolation module with an IDW-style variant (V2, HIGNN–IDW) leads to a severe degradation across all brackets (0.2144, 0.2008, 0.4368), highlighting that fixed distance-based weighting is insufficient within our graph-temporal framework and that feature-conditioned residual weighting is critical for reliable corrections. Replacing the GRU with an MLP (V3, HIGNN–MLP) yields similar performance to the full model in the low and medium brackets (0.0983 and 0.2097) but substantially worsens the high bracket (0.3953), supporting the benefit of recurrent temporal modeling for capturing event evolution and sustained dynamics.

Finally, the MLP-only baseline (V4) performs markedly worse, especially in the high bracket (0.5539), confirming that explicit graph-based spatial modeling and waterway-aware aggregation are essential once spatial heterogeneity and large water-level changes are present. As expected, the best-possible weighted-average upper bound remains significantly lower than all deployable methods in every bracket.

Peak-event recall. To complement RMSE with an event-oriented metric, Fig. 14 reports event recall for ground-truth peaks with magnitude ≥ 0.6 m. A prediction is counted as a hit if it contains a peak within a ± 3 hour window of the ground-truth peak time and the peak magnitude matches within 0.1 m. The full HIGNN (V1) attains the highest recall among deployable variants, while both V2 and V3 reduce recall and V4 performs markedly worse. This indicates that the full model is not only more accurate on average for large changes, but is also more reliable at recovering peak timing and magnitude. The upper bound reaches only about 0.6 recall, showing that a substantial fraction of large peaks are not explainable from the available neighborhood using any weighted-average interpolation; within this intrinsic limitation, V1 closes much of the gap.

Results on the interpolation-feasible subset. Finally, we repeat the ablation on the interpolation-feasible subset and focus on the high bracket $[0.6, 1.2)$, where improvements are most practically relevant (Table 11). On this subset, the full model again performs best (0.1874 m). Replacing the GRU with an MLP (V3) substantially increases error (0.2483 m), reinforcing that recurrent temporal modeling helps track the evolution of high-magnitude events even when spatial information is available. Removing topographic edge features (V5, HIGNN–NON) also causes a clear degradation (0.2520 m), suggesting that terrain-derived edge attributes remain important for accurately propagating large changes along the waterway graph under interpolation-feasible conditions.

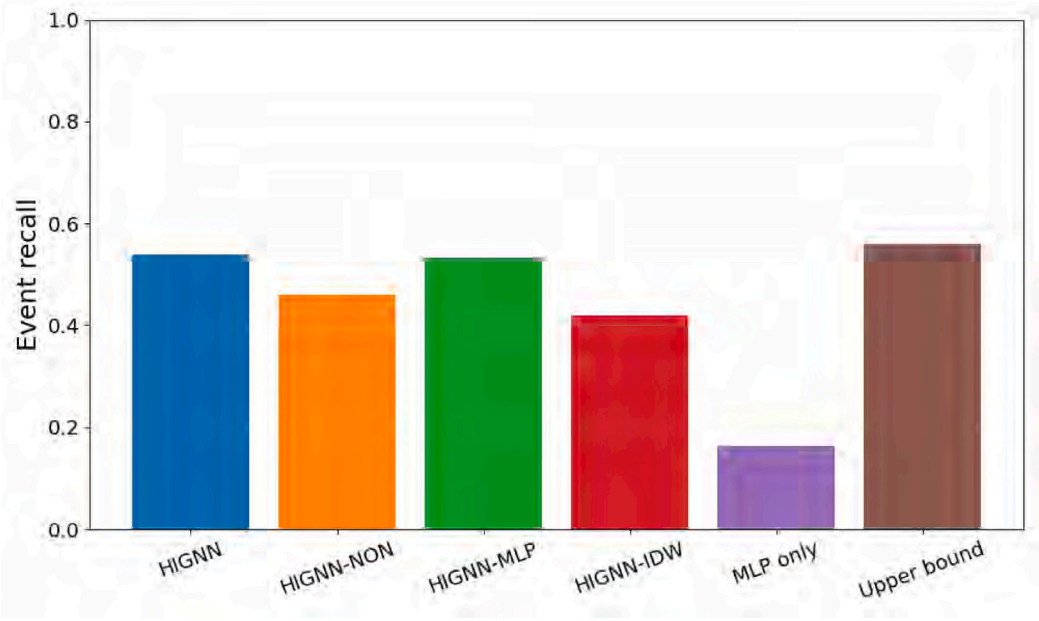


Fig. 14. Event recall for HIGNN variants on large-change events. Ground-truth peaks satisfy $y \geq 0.6$ m. A prediction is counted as a hit if it contains a peak within ± 3 h of the ground-truth peak and matches the peak magnitude within 0.1 m. "Upper bound" denotes the best-possible weighted-average interpolation from neighboring stations.

Table 11
Ablation study of HIGNN variants. RMSE in the high water-level bracket $[0.6, 1.2)$ on the interpolation-feasible subset. Best and second best are computed *excluding* the upper-bound method; best is in **bold** and second best is underlined.

Version	0.6–1.2 m
V1 (Full HIGNN)	**0.1874**
V2 (HIGNN–IDW)	0.3158
V3 (HIGNN–MLP)	0.2483
V4 (MLP only)	0.4928
V5 (HIGNN–NON)	0.2520
Upper bound	0.0396

In contrast, replacing the residual interpolation module with the IDW-based variant (V2) leads to a much larger performance drop (0.3158 m), indicating that even under best-hit (informative-neighborhood) conditions, fixed distance-based residual weighting is not a sufficient substitute for feature-conditioned residual correction within our framework. Finally, the MLP-only baseline (V4) remains substantially worse (0.4928 m), confirming that the primary gains come from combining graph-based spatial modeling with temporal prediction rather than relying on a purely non-spatial baseline. As expected, the best-possible weighted-average upper bound remains far lower (0.0396 m), serving only as a reference.

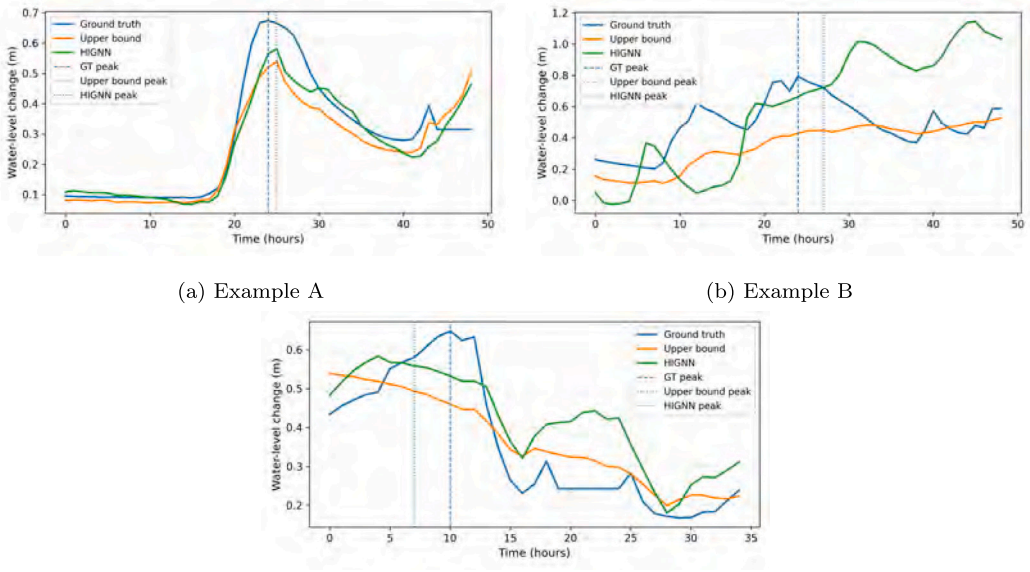
5. Discussion

The experimental results highlight several practical insights about HIGNN relative to established baselines. First, HIGNN provides the largest gains in the medium and high water-level-change ranges, where accurately capturing rapid rises and peak magnitudes is most important for flood monitoring and warning. This improvement is consistent with the model design: waterway-aware message passing aggregates information from hydrologically relevant neighbors, while the GRU refines predictions using temporal context, which helps to better track time-evolving events than purely spatial interpolation or purely feedforward predictors. In the low range, classical methods such as kriging-based baselines remain competitive, suggesting that when changes are small and series are smoother, simpler geostatistical assumptions can be sufficient.

The ablation study clarifies the role of individual components. Replacing learned residual interpolation with IDW generally increases errors in the medium and high brackets, indicating that residual weighting benefits from data-driven adaptation to spatial heterogeneity beyond distance alone. Replacing the GRU with an MLP also leads to degraded performance under larger changes, supporting the importance of temporal recurrence for modeling delayed responses and event evolution. The MLP-only variant performs substantially worse for larger changes, showing that precipitation-only prediction without graph-based spatial propagation is insufficient at many virtual stations. Taken together, these results suggest that the strongest performance is achieved when spatial structure (waterway graph), temporal structure (GRU), and learned residual correction are combined.

Furthermore, we evaluate the ability to recover large-change peak events using the peak-matching criterion (magnitude tolerance 0.1 m within a ± 3 h window). Fig. 15 provides representative examples where the best-possible weighted-average upper bound (computed from neighbors) underestimates the ground-truth peak amplitude, while HIGNN produces a closer peak magnitude within the matching window. These cases illustrate that, even when neighboring stations do not directly reproduce the peak via a weighted average, HIGNN can partially compensate by leveraging precipitation-driven signals and temporal context to adjust the peak shape and amplitude. This behavior helps explain the improved peak recall of HIGNN relative to simpler interpolation-based baselines, and also motivates our separate analysis of neighborhood quality (Section 4.7), since a substantial fraction of peaks remain difficult to reconstruct from neighbors alone.

Computational efficiency is also relevant for operational use. While several baselines have short runtimes in our experimental setting, the relative differences remain informative when inference must be repeated across many virtual stations and long time series. As summarized in Table 9, HIGNN is substantially faster than optimization-heavy geostatistical and Bayesian baselines (e.g., OK and NNGP), while maintaining competitive accuracy in the most consequential change ranges. Although IDW and a pure MLP are faster, they exhibit noticeably reduced accuracy for larger changes, whereas HIGNN achieves higher



(a) Example A (b) Example B

(c) Example C

Fig. 15. Representative peak-reconstruction examples where the best-possible weighted average underestimates the ground-truth peak, while HIGNN provides a closer peak magnitude. Vertical dashed lines indicate the ground-truth peak time; dotted lines indicate the peak time selected within the ± 3 h matching window for each method. In these cases, HIGNN reduces peak underestimation relative to the upper bound, improving event recall under the 0.1 m tolerance criterion.

accuracy with only a modest additional computational cost, making it suitable for large-scale or near real-time monitoring scenarios.

Despite these advantages, several limitations remain. First, HIGNN relies on terrain- and waterway-derived features to construct informative spatial relationships. Errors in the DEM, or in the derivation of flow direction and drainage connectivity, can propagate into edge attributes and affect attention weights, particularly in low-relief areas where flow directions can be ambiguous. Improving robustness to such uncertainty (e.g., uncertainty-aware edge features or sensitivity tests under alternative flow-direction estimates) is an important direction for future work. Second, our formulation assumes a *static* river graph. In practice, effective connectivity and travel-time relationships can vary over time due to seasonal vegetation, changing hydraulic conditions, sediment transport, and human interventions. Extending the model toward *dynamic graphs*—where edge features, and potentially adjacency, evolve over time—could better reflect these effects.

Third, HIGNN can still struggle at stations whose dynamics are strongly influenced by human operations that are not represented in our inputs, such as reservoir releases, pumping, and floodgate control. This limitation is consistent with the spatial error patterns observed for large changes: Fig. 16 maps per-station RMSE in the high bracket [0.6, 1.2), m and shows that the largest errors are geographically concentrated rather than uniformly spread.

Many of these high-error locations occur in reaches where regulation and local hydraulic interventions can decouple a station from nearby gauges, making neighborhood signals less informative during peak events. For example, along the Dijle near Mechelen, documented pumping interventions transfer Dijle water into wetland areas (Mechels Broek) to support rewetting, which can modify local water-level dynamics relative to upstream/downstream stations (Groen Mechelen, 2023). Similarly, the Mark catchment includes flood-mitigation measures such as retention basins and controlled overflow areas designed to attenuate peak flows, directly altering peak timing and magnitude along the river (Coördinatiecommissie Integraal Waterbeleid (CIW), 2026).

More broadly, regulation infrastructure in urban regions can produce abrupt, non-meteorological fluctuations: in Ghent, water levels along the Scheldt system are actively managed through sluices and pumping as part of the Sigma Plan (sigmaplan, 2017, 2024; City of Ghent, 2016; Institute for European Environmental Policy (IEEP), 2023); in Leuven, the Dyle is affected by engineered structures such as locks and adjustable weirs (Vlaamse Milieumaatschappij (VMM), 2015; Turkelboom et al., 2018); and in Brussels, locks and pumping operations along the Senne and the Willebroek Canal balance flood protection, navigation, and urban drainage (Port of Brussels, 2025; Tractebel, 2021). These examples help explain why some locations exhibit higher errors under large changes and suggest that incorporating operational indicators (or suitable proxy covariates) and refining graph/edge-feature construction in regulated reaches are promising directions to further reduce the remaining high-error hotspots highlighted in Fig. 16.

Finally, our evaluation is limited to a single year and a single study region. Future work should assess transferability across longer time spans and basins with different topographic characteristics, regulation intensity, and gauging density to better understand generalization and any needed re-calibration of graph construction and masking thresholds.

6. Conclusion

In this work, we introduced HIGNN, a spatiotemporal learning framework for estimating water levels at ungauged stations. By combining precipitation-driven feature extraction, graph-based spatial propagation, residual interpolation, and temporal refinement through a GRU, HIGNN effectively integrates meteorological, hydrological, and topographical information into a unified architecture.

Extensive experiments against classical geostatistical and machine learning baselines demonstrated that HIGNN achieves better accuracy in medium and high water-level ranges, where capturing dynamic fluctuations is most critical for flood monitoring and early warning. The ablation study confirmed the contribution of each component, showing that both residual interpolation and temporal modeling play key roles in improving predictive performance. Furthermore, runtime analysis indicated that HIGNN offers a favorable balance between accuracy and computational efficiency, making it suitable for large-scale or near real-time deployment.

Future work will explore integration with additional environmental indicators, incorporation of human-operation signals, and evaluation

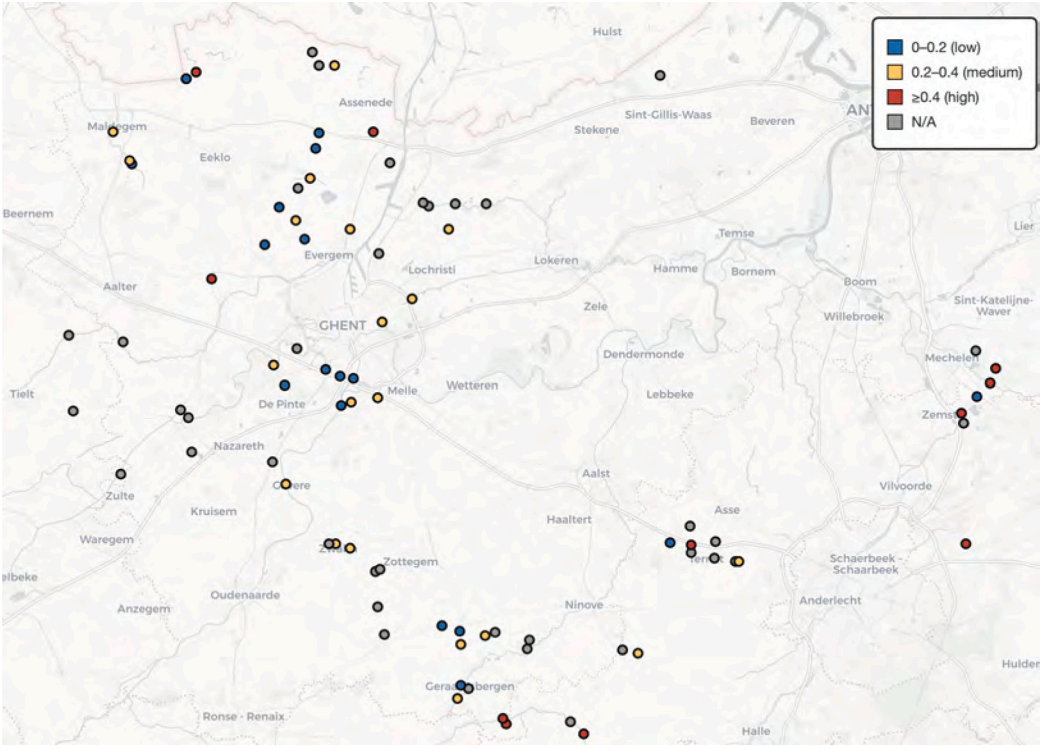


Fig. 16. Spatial distribution of HIGNN errors in the high water-level-change bracket $[0.6, 1.2)$ m. Markers are colored by per-station RMSE (blue: 0–0.2 m, orange: 0.2–0.4 m, red: ≥0.4 m); gray indicates insufficient data in this bracket.

across longer temporal horizons and diverse basins. By advancing the ability to estimate water levels in ungauged regions, HIGNN contributes toward more resilient and data-driven water management strategies.

CRediT authorship contribution statement

Anh Minh Truong: Writing – original draft, Visualization, Validation, Software, Methodology. **Guangan Chen:** Writing – review & editing, Resources. **Michiel De Baets:** Visualization, Resources. **Michiel Vlaminck:** Writing – review & editing, Supervision. **Brian Booth:** Writing – review & editing, Supervision, Conceptualization. **Hiep Luong:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was financially supported by the imec ICON VLAIO project "Floodify", Belgium (HBC.2023.0627), a collaboration between Ghent University; imec; HydroScan; Geographic Information Management (GIM); MyCSN; Aquafin; CityMesh; and the firefighters of the City of Ghent. We wish to thank all the partners for their assistance in technical discussions and data collection.

Data availability

Data will be made available on request.

Floodify dataset (Original data) (floodify-sensor-fusion-data)

References

Agentschap Digitaal Vlaanderen, 2019. Hoogte – DTM (Digitaal terreinmodel Vlaanderen II). URL: https://www.vlaanderen.be/datavindplaats/catalogus/hoogte-dtm-0. Version 2014.01; download services available via WMS/WMTS.

Ahmed, A.N., Yafouz, A., Birima, A.H., Kisi, O., Huang, Y.F., Sherif, M., Sefelnasr, A., El-Shafie, A., 2022. Water level prediction using various machine learning algorithms: A case study of Durian Tunggal River, Malaysia. Eng. Appl. Comput. Fluid Mech. 16 (1), 422–440. http://dx.doi.org/10.1080/19942060.2021.2019128.

Blöschl, G., Hall, J., Parajka, J., Perdigão, R.A.P., Merz, B., Arheimer, B., Aronica, G.T., Bilibashi, A., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G.B., Claps, P., Fiala, K., Frolova, N., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T.R., Kohnová, S., Koskela, J.J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Rogger, M., Salinas, J.L., Sauquet, E., Šraj, M., Szolgay, J., Viglione, A., Volpi, E., Wilson, D., Zaimi, K., Živković, N., 2017. Changing climate shifts timing of European floods. Science 357 (6351), 588–590. http://dx.doi.org/10.1126/science.aan2506.

Cho, K., van Merriënboer, B., Gülçehre, Ç., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNN encoder–decoder for statistical machine translation. arXiv preprint. URL: https://arxiv.org/abs/1406.1078. arXiv:1406.1078.

City of Ghent, 2016. Ghent climate adaptation plan 2016–2019. URL: https://stad.gent/sites/default/files/page/documents/Ghent%20Climate%20Adaptation%20Plan%202016-2019.pdf.

Coördinatiecommissie Integraal Waterbeleid (CIW), 2026. Wateroverlast (Mark). URL: https://www.integraalwaterbeleid.be/nl/bekkens/denderbekken/gebiedsgerichte-werking/mark/wateroverlast. Overview of flood-risk context and mitigation measures in the Mark catchment, including retention basins (wachtbekkens) and controlled overflow areas.

Cressie, N.A.C., 1993. Statistics for Spatial Data. John Wiley & Sons, http://dx.doi.org/10.1002/9781119115151.

Geofabrik GmbH, 2025. OpenStreetMap data extract for Belgium. URL: https://download.geofabrik.de/europe/belgium.html.

Groen Mechelen, 2023. Natuurpunt installeert zonnepomp om Mechels Broek te vernatten. URL: https://www.groenmechelen.be/natuurpunt_installeert_zonnepomp_om_mechels_broek_te_vernatten. Article dated 27 April 2023; describes pumping Dijle water into Mechels Broek for rewetting.

Hampel, F.R., 1974. The influence curve and its role in robust estimation. J. Amer. Statist. Assoc. 69 (346), 383–393. http://dx.doi.org/10.2307/2285666, URL: https://www.jstor.org/stable/2285666.

Hendrycks, D., Gimpel, K., 2016. Gaussian error linear units (GELUs). arXiv preprint. URL: https://arxiv.org/abs/1606.08415. arXiv:1606.08415.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., Kanae, S., 2013. Global flood risk under climate change. Nat. Clim. Chang. 3 (9), 816–821. http://dx.doi.org/10.1038/nclimate1911.

Institute for European Environmental Policy (IEEP), 2023. Water policy fitness check. URL: https://ieep.eu/wp-content/uploads/2023/01/Water_Policy_Fitness_Check.pdf.

Jiang, L., et al., 2021. Calibrating 1D hydrodynamic river models in the absence of cross-section geometry using satellite observations of water surface elevation and river width. Hydrol. Earth Syst. Sci. 25, 6359–6381. http://dx.doi.org/10.5194/hess-25-6359-2021.

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, London.

Kenda, K., Peternelj, J., Mellios, N., Kofinas, D., Čerin, M., Rožanec, J., 2020. Usage of statistical modeling techniques in surface and groundwater level prediction. J. Water Supply: Res. Technology—AQUA 69 (3), 248–265. http://dx.doi.org/10.2166/aqua.2020.143.

Khan, M., Almazah, M.M.A., Ellahi, A., Niaz, R., Al-Rezami, A.Y., Zaman, B., 2023. Spatial interpolation of water quality index based on ordinary kriging and universal kriging. Geomatics, Nat. Hazards Risk 14 (1), 2190853. http://dx.doi.org/10.1080/19475705.2023.2190853.

Lee, C.H., Kim, K.D., Lyu, S., Kim, D.S., Kim, Y.D., 2023. Analysis of mixing patterns of river confluences through 3D spatial interpolation of sensor measurement data. Water 15 (5), 925. http://dx.doi.org/10.3390/w15050925.

Li, J., Heap, A.D., 2014. Spatial interpolation methods applied in the environmental sciences: A review. Environ. Model. Softw. 53, 173–189. http://dx.doi.org/10.1016/j.envsoft.2013.12.008.

Luo, J., et al., 2025. Classification-enhanced LSTM model for predicting river levels. J. Hydrol. 650, 132535. http://dx.doi.org/10.1016/j.jhydrol.2024.131931.

Matgen, P., Montanari, M., Hostache, R., Pfister, L., Hoffmann, L., Plaza, D., Pauwels, V.R.N., De Lannoy, G.J.M., 2010. Towards the sequential assimilation of SAR-derived water stages into hydraulic models using the ensemble Kalman filter. Water Resour. Res. 46 (W08403), http://dx.doi.org/10.1029/2009WR008213.

Nagahama, V.H., Sweeney, J., Cahill, N., 2024. A scalable Bayesian spatiotemporal model for water level predictions using a nearest neighbor Gaussian process approach. arXiv preprint. URL: https://arxiv.org/abs/2412.06934. arXiv:2412.06934.

Paiva, R.C.D., Durand, M.T., Hossain, F., 2015. Spatiotemporal interpolation of discharge across a river network by using synthetic SWOT satellite data. Water Resour. Res. 51 (1), 430–449. http://dx.doi.org/10.1002/2014WR015618.

Port of Brussels, 2025. Port of Brussels manages canal infrastructure and regulates water flow to prevent flooding. URL: https://port.brussels/en/node.

2017. Flood control areas in the sigma plan include sluices for draining when water subsides. URL: https://www.sigmaplan.be/sites/default/files/2024-05/170817-sigmabrochure-2017-en-lr.pdf.

2024. Sigma plan includes new pumping stations and buffer canals to protect residents. URL: https://cdn.life-sparc.eu/sites/2/2024/04/22110742/Sigmaplan_FICHE_Vlassenbroek_ENG.pdf.

Sun, L., Seidou, O., Nistor, I., Liu, K., 2016. Review of the Kalman-type hydrological data assimilation. Hydrol. Sci. J. 61 (13), 2348–2366. http://dx.doi.org/10.1080/02626667.2015.1127376.

Taccari, M.L., Wang, H., Nuttall, J., Chen, X., Jimack, P.K., 2024. Spatial-temporal graph neural networks for groundwater data. Sci. Rep. 14 (1), 24564. http://dx.doi.org/10.1038/s41598-024-75385-2.

Tractebel, 2021. Charleroi–Brussels Canal: Modernization of locks and pumping stations. URL: https://tractebel-engie.be/en/news/2021/charleroi-brussels-canal-modernization-of-locks-and-pumping-stations.

Tucci, M., 2023. Hourly water level forecasting in an hydroelectric basin using spatial interpolation and artificial intelligence. Sensors 23 (1), 203. http://dx.doi.org/10.3390/s23010203.

Turkelboom, F., Demeyer, R., Vranken, L., Coucke, L., 2018. Green versus grey solution for flood control of Leuven City (Belgium). URL: https://iale-europe.eu/iale2017/green-versus-grey-solution-flood-control-leuven-city-belgium. Hydrological modeling of flood control solutions using floodplains vs. reservoirs.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y., 2018. Graph attention networks. In: International Conference on Learning Representations. URL: https://openreview.net/forum?id=rJXMpikCZ.

Vlaamse Milieumaatschappij (VMM), 2015. The River Dyle in Leuven (Belgium): A blessing and a curse. URL: https://en.vmm.be/publications/the-river-dyle-in-leuven-belgium-a-blessing-and-a-curse.

Vlaamse Milieumaatschappij (VMM), et al., 2025. Woordenlijst (Glossary). URL: https://waterinfo.vlaanderen.be/default.aspx?path=NL/Algemene_Info/Woordenlijst&KL=en. Waterinfo.be / Vlaamse Milieumaatschappij (VMM).

Willner, S.N., Otto, C., Levermann, A., 2018. Global economic response to river floods. Nat. Clim. Chang. 8 (7), 594–598. http://dx.doi.org/10.1038/s41558-018-0173-2.

Yasin, K.H., Gelete, T.B., Iguala, A.D., Kebede, E., 2024. Optimal interpolation approach for groundwater depth estimation. MethodsX 13, 102916. http://dx.doi.org/10.1016/j.mex.2024.102916.

Zhang, Z., Fink, O., 2024. Algorithm-informed graph neural networks for leakage detection and localization in water distribution networks. arXiv preprint. URL: https://arxiv.org/abs/2408.02797. arXiv:2408.02797.