

Physics-informed Graph Neural Networks for Operational Flood Modeling

Carlo Malapad Acosta¹, Herath Mudiyanselage Viraj Vidura Herath², Jia Yu Lim¹, Abhishek Saha³, Sanka Rasnayaka¹ and Lucy Marshall²

¹Department of Computer Science, School of Computing, National University of Singapore

²Faculty of Engineering, The University of Sydney

³Delft Institute of Applied Mathematics, Delft University of Technology
e1391136@u.nus.edu, viraj.herath@sydney.edu.au

Abstract

Flood models inform strategic disaster management by simulating the spatiotemporal hydrodynamics of flooding. While physics-based numerical flood models are accurate, their substantial computational cost limits their use in operational settings where rapid predictions are essential. Models designed with graph neural networks (GNNs) provide both speed and accuracy while having the ability to process unstructured spatial domains. Given its flexible input and architecture, GNNs can be leveraged alongside physics-informed techniques with ease, significantly improving interpretability. This study introduces a novel flood GNN architecture, DUALFloodGNN, which embeds physical constraints at both global and local scales through explicit loss terms. The model jointly predicts water volume at nodes and flow along edges through a shared message-passing framework. To improve performance for autoregressive inference, model training is conducted with a multi-step loss enhanced with dynamic curriculum learning. Compared with standard GNN architectures and state-of-the-art GNN flood models, DUALFloodGNN achieves substantial improvements in predicting multiple hydrologic variables while maintaining high computational efficiency. The model is open-sourced at https://github.com/acostacos/dual_flood_gnn.

1 Introduction

Floods pose a substantial socioeconomic threat to modern society, resulting in losses of more than 3 trillion USD in the past 25 years and affecting up to 1.81 billion individuals globally [Rentschler *et al.*, 2022; UNDRR and CRED, 2020]. Furthermore, the frequency and severity of such events are only projected to increase due to climate change [Tabari, 2020]. Effective mitigation of these repercussions requires detailed flood information. In such cases, hydrodynamic flood models play a crucial role by simulating water behavior across space and time, enabling data-driven disaster management and flood-resilient engineering designs [Kumar *et al.*, 2023]. Conventional physics-based numerical models capture flood

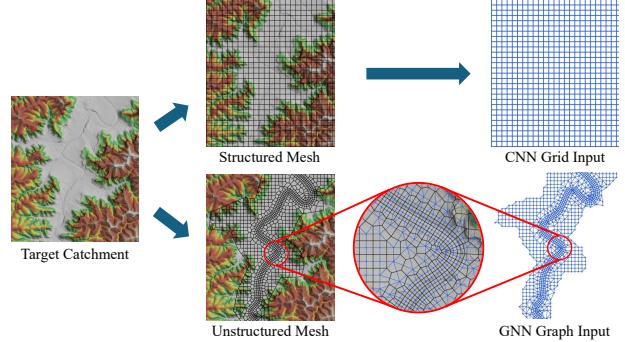


Figure 1: Discretization of a target catchment. A structured mesh produces a uniform grid suitable for Convolutional Neural Networks (CNNs). A more flexible unstructured mesh can be translated into a graph for Graph Neural Networks (GNNs). The centroid of each mesh cell is represented as a node, while connections between adjacent cell centers define the edges as shown in the magnified section.

dynamics with high accuracy by solving the governing equations of fluid flow. However, they require substantial computation time, limiting their effectiveness for real-time or operational use [Sun *et al.*, 2023].

To address this, data-driven surrogate models forecast flooding through statistical correlations, resulting in much faster computation speed. In particular, deep learning (DL) has the ability to automatically learn and extract complex non-linear relationships from data, making it suitable for modeling fluid motion. Research continues to assess the efficacy of various architectures including multilayer perceptrons (MLPs) [Ahmadlou *et al.*, 2021; Xie *et al.*, 2021], recurrent neural networks (RNNs) [Fang *et al.*, 2021; Rahimzad *et al.*, 2021] and convolutional neural networks (CNNs) [Herath *et al.*, 2025; Hosseiny, 2021; Neo *et al.*, 2025]. Although suitable for real-time forecasting, DL models are often criticized as "black boxes", due to a limited understanding of their predictive paradigm [Devia *et al.*, 2015]. Such an opacity presents a significant disadvantage in physics-based disciplines, where explainability is essential to ensure consistency with underlying physical laws.

Physics-informed DL models combine the strengths of data-driven and physics-based approaches by embedding physical constraints into DL architectures, enabling them to

learn the underlying behaviour of physical systems [Sharma *et al.*, 2023]. The main challenge of this approach lies in the complexity of aligning existing algorithms in both domains. An example of such a case is illustrated in figure 1. Physics-based numerical approximations involve the discretization of the area into a mesh structure, which is typically integrated in DL models through CNNs, given their aptitude with spatial input [Gao *et al.*, 2022]. However, this methodology is also fundamentally constrained by the nature of CNNs, which are restricted to processing structured grid meshes. This limits physical fidelity, as this discretization scheme cannot adequately represent the irregular topographies and non-uniform fluid dynamics characteristic of flooding scenarios. Graph neural networks (GNNs) were developed to directly address this limitation by generalizing CNN-based convolution for arbitrary spatial data represented as graphs [Bronstein *et al.*, 2017]. This representation enables a more flexible description of hydrodynamic quantities through an unstructured mesh and has been shown to improve generalization in physical systems [Thangamuthu *et al.*, 2023].

Integration of physics-informed techniques with GNN flood models is still an emerging field of research. For example, HydroGraphNet [Taghizadeh *et al.*, 2025] incorporates the principle of mass conservation at a global graph-level through a regularization term in the loss function. However, local mass conservation is only implicitly assumed, as global mass balance does not guarantee the same behavior in smaller regions. We address this research gap by incorporating an additional loss term explicitly enforcing node-level mass balance. To enable this, we diverge from previous approaches [Bentivoglio *et al.*, 2023; Bentivoglio *et al.*, 2024] by mapping the time-varying water flow as an edge attribute, thus preserving its vector characteristics. This graph mapping provides more intuitive physical meaning and closely mimics the capabilities of physics-based numerical models. However, this also introduces the challenge of predicting both node and edge features. Existing domain-agnostic GNNs created for this purpose [Jiang *et al.*, 2019; Yang and Li, 2020] require the generation of a line graph, which typically contains more nodes and edges, increasing computational overhead. We develop a novel architecture that performs simultaneous node and edge prediction given only the original graph utilizing two separate decoders. We coin our approach **Double-target Universally And Locally constrained Flood GNN (DUALFloodGNN)**.

Our contributions can be summarized as follows:

- We develop a physics-informed loss function that enforces mass conservation at both global and local scales through multiple regularization terms, preserving conservation within localized areas of the computational mesh and ensuring physically realistic flood dynamics.
- To facilitate computation of local mass balance, we design an architecture capable of simultaneously predicting features for nodes (water volume) and edges (water flow). Joint modeling is performed through shared message generation, which utilizes their physical correlation while requiring only the original graph structure.
- To the best of our knowledge, this is the first study

to benchmark performance across different GNN-based flood models. Experimental results validate the effectiveness of DUALFloodGNN, which outperforms both standard GNN baselines and state-of-the-art GNN flood models [Bentivoglio *et al.*, 2023; Taghizadeh *et al.*, 2025].

2 Background and Related Work

Graph Neural Networks. GNNs have found widespread use in the processing of graph-structured data [Hamilton *et al.*, 2018; Kipf and Welling, 2017; Veličković *et al.*, 2018]. GNNs generalize CNN-based convolution for non-Euclidean domains [Bronstein *et al.*, 2017] through the message passing framework [Gilmer *et al.*, 2017]. Given a node i in the graph, its node embedding $h_i^{(l)}$ for the current layer l is updated to $h_i^{(l+1)}$ only based on its local neighborhood $\mathcal{N}(i)$ through Eq. 1:

$$h_i^{(l+1)} = U(h_i^{(l)}, A(M(h_i^{(l)}, h_j^{(l)}))); \forall j \in \mathcal{N}(i) \quad (1)$$

where h_j are the neighbor embeddings and M , A and U are the message, aggregate and update functions. This is repeated for each node in the graph, creating an updated graph representation after one full cycle. Each layer in a GNN corresponds to one message passing iteration, which can be repeated to enable further message propagation.

Node and Edge Prediction. CensNet [Jiang *et al.*, 2019] predicts both node- and edge-level features through alternating spectral graph convolutions using the original graph for node embeddings and the line graph for edge embeddings. NENN [Yang and Li, 2020] extends this approach to spatial-based message passing and incorporates attention to both node- and edge-level layers.

GNNs for Flood Modeling. Initial studies adopted existing architectures for the hydrological context [Farahmand *et al.*, 2023; Oliveira Santos *et al.*, 2023]. Further studies incorporated RNN components to explicitly capture the temporal aspect of flood dynamics [Kazadi *et al.*, 2024; Roudbari *et al.*, 2024; Zhao *et al.*, 2020], with some utilizing heterogeneous graph structures to encode diverse relationships within the system [Jiang *et al.*, 2024; Luo *et al.*, 2024].

Physics-informed Flood GNNs. Physics-informed constraints were integrated either through the model architecture or loss function. SWE-GNN [Bentivoglio *et al.*, 2023] modulates propagated messages by the latent gradient between neighboring nodes, which is extended for a multi-scale architecture in mSWE-GNN [Bentivoglio *et al.*, 2024]. GNN-fusion [Zhang *et al.*, 2024] computes flow at a node as the sum of predicted flow from its adjacent edges to reliably enforce mass conservation. While it supports node and edge prediction, their model generates these embeddings with different convolutions and has only been tested for 1D flow in drainage networks. HydroGraphNet [Taghizadeh *et al.*, 2025] introduces a physics-based loss that enforces global mass conservation through a forward and backward check, ensuring the predicted water volume is within bounds.

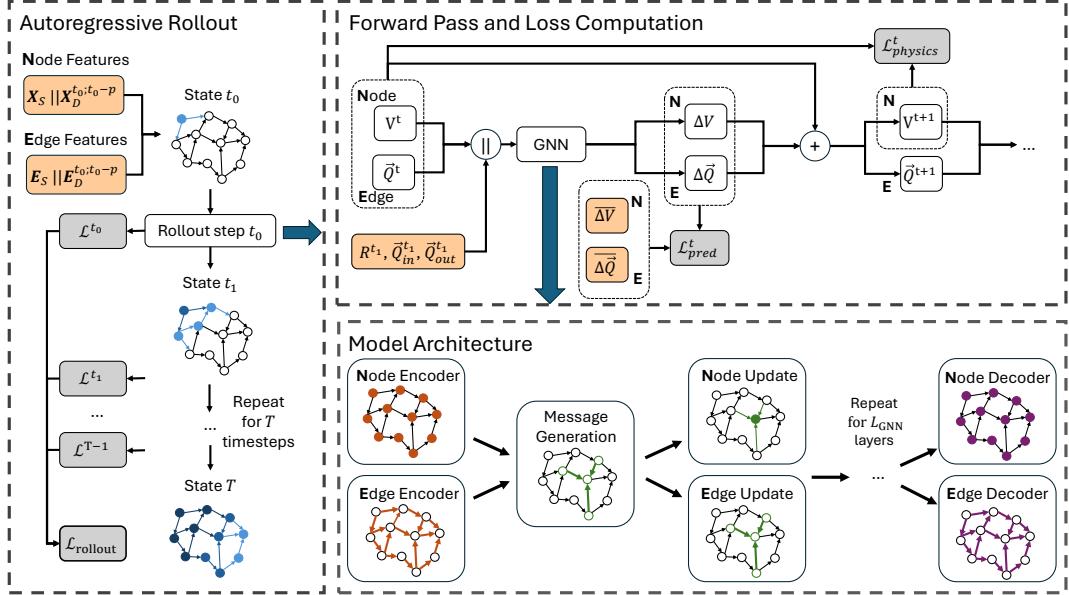


Figure 2: The DUALFloodGNN architecture. To predict T timesteps, an autoregressive rollout is performed while accumulating the loss at each timestep to compute the total rollout loss $\mathcal{L}_{rollout}$ (left). During each step, the model predicts node (ΔV) and edge ($\Delta \vec{Q}$) embeddings through joint modeling which shares generated messages (bottom right). From this, the prediction loss \mathcal{L}_{pred} and physics loss $\mathcal{L}_{physics}$ are calculated given the previous state (V^t, \vec{Q}^t), current state ($V^{t+1} = V^t + \Delta V, \vec{Q}^{t+1} = \vec{Q}^t + \Delta \vec{Q}$), and ground truth ($\overline{\Delta V}, \overline{\Delta \vec{Q}}$) (top right).

3 Preliminaries

Notations. Formally, a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is comprised of a set of nodes \mathcal{V} and edges \mathcal{E} . Nodes are described by a node feature matrix $\mathbf{X} = \mathbb{R}^{|\mathcal{V}| \times f_v}$ where f_v is the node feature dimension. Edges are typically represented in terms of an adjacency matrix $\mathbf{A} \in \{0, 1\}^{|\mathcal{V}| \times |\mathcal{V}|}$ where $\mathbf{A}_{ij} = 1$ if a connecting edge exists from node i to node j , otherwise, $\mathbf{A}_{ij} = 0$. A memory-efficient alternative, especially for sparse graphs, is the coordinate (COO) format $\mathbf{I} = \mathbb{R}^{2 \times |\mathcal{E}|}$ where $(i, j) \in \mathbf{I}$ if there is an edge from node i to node j [Fey and Lenssen, 2019]. Through this representation, edges can be described with multi-dimensional feature vectors, which combine to form the edge feature matrix $\mathbf{E} = \mathbb{R}^{|\mathcal{E}| \times f_e}$ where f_e is the edge feature dimension.

Problem Definition. A flood event can be described through a time series of directed graphs $\{G_0, G_1, \dots, G_T\}$ where nodes and edges have associated water volume (V) and flow (\vec{Q}) values respectively. Water flow is a vector quantity where $\vec{Q} > 0$ denotes flow along the edge's orientation and $\vec{Q} < 0$ indicates movement in the opposite direction. Nodes and edges can be further described by static structural properties of the catchment. The model performs a node- and edge-level regression task to predict future hydraulic states $\{V^{t+n}, \vec{Q}^{t+n} | n = 1, \dots, T\}$ at each structure given the initial state at the current timestep t and up to p previous states $\{V^{t-n}, \vec{Q}^{t-n} | n = 0, \dots, p\}$. Predictions are obtained in an autoregressive manner, where the predicted features are used recursively as dynamic features for subsequent model inference. Additionally, boundary conditions (e.g., rainfall R , up-

stream inflow \vec{Q}_{in} and downstream outflow \vec{Q}_{out}) are supplied at each timestep $\{R^{t+n}, \vec{Q}_{in}^{t+n}, \vec{Q}_{out}^{t+n} | n = 0, \dots, T - 1\}$, describing water flow entering and exiting the system.

4 Methodology

Overview. In this section, we discuss DUALFloodGNN, a novel physics-informed GNN architecture used for the simultaneous prediction of node water volume and edge water flow. It is comprised of three main components: (1) a model that performs shared message passing to predict both node and edge features, (2) a physics-informed loss function that enforces mass conservation at a global and local scale between consecutive predictions, and (3) an autoregressive training strategy enhanced by dynamic curriculum learning. Figure 2 provides a diagram illustrating this framework.

4.1 Model Architecture

Input and Output. The input of the model is a directed graph at timestep t containing static features (e.g., area, elevation) and dynamic features (e.g., volume, flow, rainfall, global inflow) up to p previous timesteps. The node feature matrix \mathbf{X}^t (Eq. 2) and edge feature matrix \mathbf{E}^t (Eq. 3) are created by concatenating all the features into a single matrix:

$$\mathbf{X}^t = \mathbf{X}_S || \mathbf{X}_D^{t-p} || \dots || \mathbf{X}_D^t \quad (2)$$

$$\mathbf{E}^t = \mathbf{E}_S || \mathbf{E}_D^{t-p} || \dots || \mathbf{E}_D^t \quad (3)$$

where $||$ is the concatenation operation. Instead of predicting future states directly, DUALFloodGNN predicts the change in water volume ΔV and water flow $\Delta \vec{Q}$ from the current timestep to the next timestep. This reduces the variation in the

output, achieving more stable and accurate long-horizon predictions. The model’s prediction can be added to the current dynamic features V^t and \vec{Q}^t to get the next timestep values V^{t+1} and \vec{Q}^{t+1} , which also acts as a residual connection.

To facilitate joint node and edge modeling, we adopt the general Encode-Process-Decode framework [Battaglia *et al.*, 2018]. This is well-suited for the dual regression task as node and edge features are transformed into a unified latent space, enabling a richer exchange of information between them during message passing.

Encoder. The encoder block transforms the input node features \mathbf{X}^t and edge features \mathbf{E}^t into their respective latent representations. These two are encoded separately as they represent different physical quantities. The encoding operations are defined in Eqs. 4 & 5 as

$$\mathbf{H}^t = \text{MLP}(\mathbf{X}^t) \quad (4)$$

$$\epsilon^t = \text{MLP}(\mathbf{E}^t) \quad (5)$$

where $\mathbf{H}^t = \mathbb{R}^{|\mathcal{V}| \times d}$ and $\epsilon^t = \mathbb{R}^{|\mathcal{E}| \times d}$ are the encoded node and edge representations, and d is the latent dimension size.

Processor: Shared Message Passing. The latent node and edge feature matrices, \mathbf{H}^t and ϵ^t , are then processed through L_{GNN} message passing layers that iteratively update both of these at each layer l . Given a node in the graph i , messages m_{ji} are generated for each of its neighboring nodes $\mathcal{N}(i)$:

$$m_{ji}^{(l)} = \text{MLP}(h_i^{(l)} || h_j^{(l)} || e_{ij}^{(l)}); \forall j \in \mathcal{N}(i) \quad (6)$$

where h_i and h_j are the node embeddings and its neighbor embeddings, respectively, and e_{ij} is the edge embedding between them. Messages are conditioned on both node and edge attributes, as they are high-dimensional representations of flood volume along their respective structures, suggesting their co-dependency in characterizing local fluid dynamics.

Following this interpretation, the same generated messages are used to simultaneously update both embedding types. To update the current embedding of node i , messages m_{ji} are aggregated through summation, improving the model’s expressive power in identifying the underlying graph structure [Xu *et al.*, 2019]. The aggregated value is fed to a parametrized fully connected layer to generate the updated node embeddings $h_i^{(l+1)}$:

$$h_i^{(l+1)} = \text{MLP}\left(\sum_{j \in \mathcal{N}(i)} m_{ji}^{(l)}\right) \quad (7)$$

The updated edge embeddings $e_{ij}^{(l+1)}$ are directly overwritten with the message m_{ji} at each edge, inspired by the approach of Ashraf *et al.* [2024].

$$e_{ij}^{(l+1)} = m_{ji}^{(l)}; \forall j \in \mathcal{N}(i) \quad (8)$$

Residual connections are added at each node, and edge updates are made to stabilize predictions. From a hydrological perspective, these update operations are physically meaningful, as messages m_{ji} symbolize flow between regions and their sum $\sum_{j \in \mathcal{N}(i)} m_{ji}$ for node i would represent the net water flux in that region.

Decoder. After the processor block, the final embedding matrices of the nodes and edges, \mathbf{H}^L and ϵ^L , are translated back to the prediction feature space through a decoder, generating the final output of the model. Similar to the encoder, two separate modules are used to produce predictions for each type of structure. This decoding operation can be summarized by Eqs. 9 & 10 as

$$\Delta V^{t+1} = \text{MLP}(\mathbf{H}^L) \quad (9)$$

$$\Delta \vec{Q}^{t+1} = \text{MLP}(\epsilon^L) \quad (10)$$

where $\Delta V^{t+1} = \mathbb{R}^{|\mathcal{V}|}$ and $\Delta \vec{Q}^{t+1} = \mathbb{R}^{|\mathcal{E}|}$ are the respective node and edge predictions.

Learnable Parameters. The encoders, decoders and message passing functions are all implemented as MLPs with L_{MLP} layers. A ReLU activation function is added after each layer l except for the final layer $L_{\text{MLP}} - 1$. This is mathematically expressed in Eqs. 11 & 12:

$$\mathbf{x}^{l+1} = \text{ReLU}(\mathbf{W}^l \mathbf{x}^l) \quad (11)$$

$$\mathbf{x}^{(L_{\text{MLP}})} = \mathbf{W}^{(L_{\text{MLP}}-1)} \mathbf{x}^{(L_{\text{MLP}}-1)} \quad (12)$$

where \mathbf{x} are the processed embeddings and \mathbf{W} is a matrix with learnable parameters. Following the design of SWE-GNN [Bentivoglio *et al.*, 2023], the bias term is omitted to prevent spurious water volume predictions in isolated nodes distant from the flood extent, a modification empirically demonstrated to improve flood classification.

4.2 Physics-informed Loss Function

Multi-task Prediction Loss. The base training objective is formulated as a multi-task loss consisting of the mean squared error of node and edge predictions, as defined in Eqs. 13 & 14.

$$\mathcal{L}_{\text{node}} = \frac{1}{|\mathcal{V}|} \sum_{i \in \mathcal{V}} (\overline{\Delta V_i} - \Delta V_i)^2 \quad (13)$$

$$\mathcal{L}_{\text{edge}} = \frac{1}{|\mathcal{E}|} \sum_{k \in \mathcal{E}} (\overline{\Delta \vec{Q}_k} - \Delta \vec{Q}_k)^2. \quad (14)$$

The total prediction loss can then be computed as a weighted sum between these two terms, expressed in Eq. 15:

$$\mathcal{L}_{\text{pred}} = \lambda_1 \mathcal{L}_{\text{node}} + \lambda_2 \mathcal{L}_{\text{edge}} \quad (15)$$

where the coefficients λ_1 and λ_2 dictate the contribution of $\mathcal{L}_{\text{node}}$ and $\mathcal{L}_{\text{edge}}$ in $\mathcal{L}_{\text{pred}}$.

Mass Balance Regularization. To ground model predictions in physical realism, we utilize additional regularization terms derived from the conservation of mass from the Shallow Water Equations, the fundamental principles governing 2D flood movement [Costabile *et al.*, 2017]. If assuming constant density, mass conservation is equivalent to volume conservation, enabling formulation using volumetric quantities aligned with the model’s predictions. The physics-informed loss is comprised of global mass loss $\mathcal{L}_{\text{global}}$ and local mass loss $\mathcal{L}_{\text{local}}$, which is written in Eq. 16 as

$$\mathcal{L}_{\text{physics}} = \lambda_3 \mathcal{L}_{\text{global}} + \lambda_4 \mathcal{L}_{\text{local}} \quad (16)$$

where the coefficients λ_3 and λ_4 balance the contribution of $\mathcal{L}_{\text{global}}$ and $\mathcal{L}_{\text{local}}$ in $\mathcal{L}_{\text{physics}}$. We now explain the formulation of $\mathcal{L}_{\text{global}}$ and $\mathcal{L}_{\text{local}}$.

Global-level Regularization. Global mass conservation considers the entire target catchment as the control volume. We define the global mass conservation loss as follows (Eq. 17):

$$\mathcal{L}_{global} = \left| \sum_{i \in \mathcal{V}} \Delta V_i^t - ((\vec{Q}_{in}^t - \vec{Q}_{out}^t) \cdot \Delta t + \sum_{i \in \mathcal{V}} R_i^t) \right| \quad (17)$$

where $\Delta V_i^t = V_i^{t+1} - V_i^t$ is the change in volume at node i from the current timestep t to the next timestep $t+1$ and Δt is the temporal resolution between consecutive timesteps. The regularization term is based on the absolute value of the mass balance residual, resulting in a well-defined convex objective for gradient-based optimization. The minimum $\mathcal{L}_{global} = 0$ corresponds to perfect conservation. Notably, this form of regularization is only applied to node volume predictions V as the other variables \vec{Q}_{in} , \vec{Q}_{out} and R are given to the model as boundary conditions.

Local-level Regularization. Mass conservation was also explicitly enforced on a local scale considering each individual node i as the control volume. We define the local mass conservation loss as follows (Eq. 18):

$$\mathcal{L}_{local} = \sum_{i \in \mathcal{V}} \left| \Delta V_i^t - ((\vec{Q}_{i+}^t - \vec{Q}_{i-}^t) \cdot \Delta t + R_i^t) \right| \quad (18)$$

where \vec{Q}_{i+} and \vec{Q}_{i-} are the respective inflow and outflow at a node i . Taking the absolute value is performed for the same reasons previously mentioned, where $\mathcal{L}_{local} = 0$ corresponds to perfect conservation at all nodes. Unlike the global regularization term, enforcing this condition balances both node water volume V and edge water flow \vec{Q} predictions.

The inflow \vec{Q}_{i+} and outflow \vec{Q}_{i-} at each node is computed directly from the edge predictions \vec{Q} . Given the COO adjacency matrix \mathbf{I} , let $\mathbf{I}' = \mathbb{R}^{2 \times |\mathcal{E}|}$ be the transposed adjacency matrix such that

$$\mathbf{I}' = \{(j, i) \mid (i, j) \in \mathbf{I}\}. \quad (19)$$

The undirected adjacency matrix $\mathbf{I}_{undir} = \mathbb{R}^{2 \times 2|\mathcal{E}|}$ can then be represented as in Eq. 20.

$$\mathbf{I}_{undir} = \mathbf{I} \parallel \mathbf{I}'. \quad (20)$$

Weights are generated for each edge in \mathbf{I}_{undir} representing the flow along its direction through Eq. 21.

$$\mathbf{I}_w = \mathbf{I}_{undir} \odot \text{ReLU}(\vec{Q} \parallel -\vec{Q}) \quad (21)$$

where \odot is the element-wise product. Applying $\text{ReLU}(\cdot)$ removes negative values of \vec{Q} , which indicate flow opposite to the edge direction, leaving only positive weights. The total inflow is then computed for each node i as in Eq. 22:

$$\vec{Q}_{i+} = \sum_{j \in \mathcal{N}_{in}(i)} \mathbf{I}_{w_{ij}} \quad (22)$$

where $\mathcal{N}_{in}(i) = \{j \mid (j, i) \in \mathbf{I}_{undir}\}$ are the edges pointing towards node i . Similarly, the total outflow at each node is denoted as in Eq. 23:

$$\vec{Q}_{i-} = \sum_{j \in \mathcal{N}_{out}(i)} \mathbf{I}_{w_{ij}} \quad (23)$$

where $\mathcal{N}_{out}(i) = \{j \mid (i, j) \in \mathbf{I}_{undir}\}$ are the edges pointing away from node i .

Overall Loss. The overall loss function is derived by taking the sum of the prediction loss and physics loss, defined in Eq. 24:

$$\mathcal{L} = \mathcal{L}_{pred} + \mathcal{L}_{physics}. \quad (24)$$

4.3 Autoregressive Training

Multi-step-ahead Loss. Multi-step predictions are obtained through autoregressive inference that introduces gradually accumulating errors. The typical teacher forcing training strategy is unable to handle such distribution shifts [Teutsch and Mäder, 2022]. We adopt the approach of Bentivoglio et al. [2023], which proposes the use of a multi-step-ahead loss to allow the model to develop robustness to its own noisy input. For each training batch, an autoregressive rollout is performed, after which the total loss used for backpropagation is calculated as in Eq. 25:

$$\mathcal{L}_{rollout} = \frac{1}{O} \sum_{o=1}^O \mathcal{L}(\Delta V^o, \Delta \vec{Q}^o) \quad (25)$$

where O is the rollout length and \mathcal{L} is the overall loss defined in Eq. 24. This procedure is particularly crucial for the application of the physics-informed loss, as during multi-step rollouts, the constrained hydrological states are given purely by the model’s predictions. Thus, not only does the model learn to correct its own errors, but to do so in a manner that adheres to mass conservation throughout the trajectory.

Dynamic Curriculum Learning. To improve training stability, the original implementation utilizes a curriculum learning strategy [Bentivoglio et al., 2023]. The model is initially trained on the simpler task of predicting 1-step-ahead loss, which is gradually increased at a 1-step increment after a set number of epochs until it finally reaches the target rollout length O . However, this deterministic approach fails to accommodate the non-uniform difficulty of predicting varying time horizons. To create a more adaptive progression, this fixed schedule is replaced with an early stopping mechanism, where the prediction horizon is extended only after the model’s performance has converged on the validation set for the current rollout length o . Furthermore, we design the step size between curriculum stages to be a tunable hyperparameter C , which controls the learning progression. To mitigate catastrophic interference, the learning rate is decayed by a factor γ after each curriculum step. This entire process can be interpreted as a sequence of fine-tuning steps, where a converged model for a given horizon is adapted for the subsequent, more challenging longer rollout duration.

5 Experiments

5.1 Experimental Setup

Dataset. The target catchment for this study was taken from a section of the Wollombi River watershed located in New South Wales, Australia. A dataset comprised of 56 flow-dominant flood events were generated using the HEC-RAS numerical model [Brunner, 2025]. The 2D unstructured mesh was converted to a directed graph composed of 1129 nodes and 2743 edges. Each event consists of 576 time steps with a resolution of 15-minutes.

Model	Node Water Volume			Edge Water Flow		
	RMSE (m^3) \downarrow	MAE (m^3) \downarrow	NSE \uparrow	RMSE (m^3/s) \downarrow	MAE (m^3/s) \downarrow	NSE \uparrow
GCN (node)	$9,849.98 \pm 2,485.73$	$6,340.51 \pm 1,859.60$	0.19 ± 0.51	-	-	-
GAT (node)	$13,232.86 \pm 11,835.18$	$4,583.68 \pm 1,969.27$	-8.45 ± 27.14	-	-	-
GraphSAGE (node)	$11,146.97 \pm 11,420.79$	$4,177.75 \pm 2,519.14$	-21.40 ± 55.87	-	-	-
GIN (node)	$5,023.18 \pm 1,450.37$	$2,587.05 \pm 798.38$	0.69 ± 0.24	-	-	-
GINE (node)	$3,437.68 \pm 693.94$	$1,860.03 \pm 406.78$	0.84 ± 0.13	-	-	-
HydroGraphNet	$7,584.51 \pm 3,068.54$	$3,343.52 \pm 1,392.34$	0.08 ± 0.84	-	-	-
GCN (edge)	-	-	-	90.99 ± 25.04	40.37 ± 10.47	-2.26 ± 8.83
GAT (edge)	-	-	-	263.40 ± 63.16	60.11 ± 13.90	-22.63 ± 21.39
GraphSAGE (edge)	-	-	-	44.02 ± 12.54	17.05 ± 4.48	-0.19 ± 2.92
GIN (edge)	-	-	-	51.21 ± 16.33	22.64 ± 7.61	-0.42 ± 3.64
GINE (edge)	-	-	-	249.31 ± 70.59	58.38 ± 15.56	-18.29 ± 16.18
DUALFloodGNN	$2,216.27 \pm 868.21$	$1,075.63 \pm 394.83$	0.96 ± 0.03	25.88 ± 9.64	11.40 ± 4.17	0.92 ± 0.17

Table 1: Average metrics (\pm std) for water volume and flow regression task. The best performing model is highlighted in bold while the second best result is underlined.

Baselines. The DUALFloodGNN architecture was compared to standard GNN models such as GCN [Kipf and Welling, 2017], GAT [Veličković *et al.*, 2018], GraphSAGE [Hamilton *et al.*, 2018], GIN [Xu *et al.*, 2019], and GINE [Hu *et al.*, 2020]. Variations of these models for edge regression were created by incorporating an MLP prediction head that leverages the features of adjacent nodes to predict edge values. The proposed architecture was also compared to domain-specific GNN models, including HydroGraphNet [Taghizadeh *et al.*, 2025] and SWE-GNN [Bentivoglio *et al.*, 2023]. Models were evaluated for the water volume, flow and depth prediction tasks through 14-fold cross validation.

Metrics. Performance for the regression tasks were measured using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Additionally, the Nash–Sutcliffe Efficiency coefficient (NSE) was used to measure the model accuracy relative to the mean of the observations at each timestep as defined in Eq. 26 [Nash and Sutcliffe, 1970].

$$\text{NSE} = 1 - \frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{\sum_{t=1}^T (y_t - \bar{y}_t)^2}. \quad (26)$$

To measure flood extent classification, Critical Success Index (CSI), which is defined in Eq. 27, was computed by classifying flooded nodes for the thresholds $\tau = 0.05 \text{ m}$ and $\tau = 0.3 \text{ m}$, following the convention in other literature [Bentivoglio *et al.*, 2023; Herath *et al.*, 2025; Taghizadeh *et al.*, 2025].

$$\text{CSI} = \frac{TP}{TP + FN + FP}. \quad (27)$$

5.2 Experimental Results

Overall Results. The results for the water volume and flow prediction task are shown in table 1. For node volume prediction, DUALFloodGNN yielded the best results out of all the models, surpassing the RMSE of the second best model, GINE, by 35.53%. The proposed model maintained its superior standing for edge flow regression compared to other benchmarked GNNs, improving the baseline RMSE by

Model	RMSE \downarrow (m)	NSE \uparrow	CSI \uparrow		Inference Time (s)
	$\tau = 0.05 \text{ m}$	$\tau = 0.3 \text{ m}$	$\tau = 0.05 \text{ m}$	$\tau = 0.3 \text{ m}$	
HydroGraphNet	0.76 ± 0.34	0.21 ± 0.73	0.69 ± 0.13	0.80 ± 0.05	3.25 ± 0.17
SWE-GNN	1.23 ± 0.35	0.46 ± 0.09	0.47 ± 0.11	0.49 ± 0.09	7.21 ± 0.20
DUALFloodGNN	0.21 ± 0.07	0.97 ± 0.02	0.90 ± 0.03	0.91 ± 0.03	4.10 ± 0.06

Table 2: Average metrics (\pm std) for water depth regression task. The best performing model is highlighted in bold while the second best result is underlined.

41.21%. Notably, baseline models that incorporated edge attributes (GAT and GINE) performed significantly worse in flow prediction, highlighting the importance of joint node and edge modeling for utilizing edge features. Moreover, DUALFloodGNN exhibited stable performance across different test sets, yielding a low standard deviation. Among all the models, only DUALFloodGNN had metrics for both water volume and flow as it is the only model capable of performing node and edge prediction within a single architecture.

Each flood GNN model’s performance was evaluated for water depth prediction given its prevalence in practical use-cases. Water depth was back calculated from DUALFloodGNN’s predictions with a volume-to-elevation mapping provided in the HEC-RAS simulation. As observed in table 2, the proposed model continued to outperform even domain-specific models, yielding a 72.37% improvement to the second best RMSE. In terms of classification, DUALFloodGNN accurately categorized flooded nodes by achieving a similar CSI of around 0.9 in both 0.05 and 0.3 m thresholds. SWE-GNN showed the same trend albeit at a much lower accuracy, while HydroGraphNet displayed a high delta between its CSI for the two cases. With regards to inference speed, DUALFloodGNN closely matches the low latency of HydroGraphNet with only a 0.85 second difference. Thus, DUALFloodGNN preserves the same 2 to 3 orders of magnitude speed-up compared to numerical solvers as reported by the previous studies [Bentivoglio *et al.*, 2023; Taghizadeh *et al.*, 2025]. For example, an event that would take almost a day to run for a numerical model would only take around 2 minutes for our model to simulate.

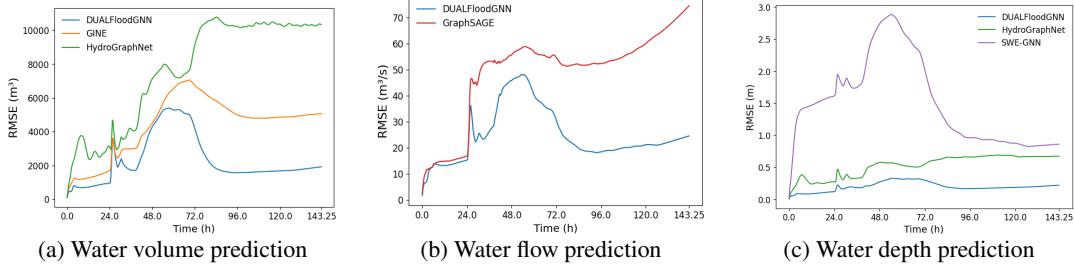


Figure 3: Task-specific RMSE of select benchmarked GNN models for each timestep in the case study event.

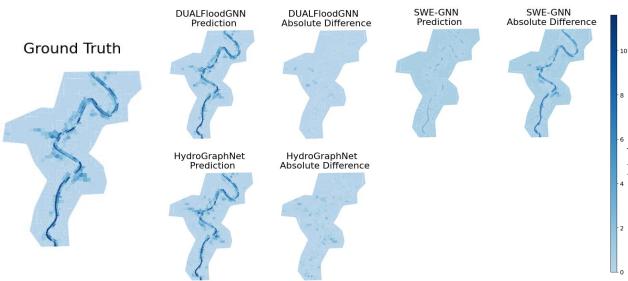


Figure 4: Maximum water depth map of flood-specific GNN models for the case study event. The maps of the ground truth, model predictions (left) and difference between the two (right) are shown.

Event Analysis. The water volume, flow and depth RMSE per timestep of select GNNs were plotted for a single case study event in figure 3. Across all tasks, DUALFloodGNN demonstrated the best temporal stability, consistently maintaining the lowest RMSE values throughout the entire rollout period. To visualize flood extent, a max flood depth map, which plots the maximum water depth for each node in the graph across all timesteps, was generated for the selected event (figure 4). Based on the ground truth, the nodes with the highest water depth are located near the river area where flow is most concentrated. SWE-GNN predicted mostly uniform flood propagation throughout the catchment, which explains its consistent CSI score in table 2 despite applying different thresholds. Although HydroGraphNet demonstrated an overall higher fidelity, it overestimated flooding for a few nodes far from the central river area. DUALFloodGNN produced the most accurate max flood map, which can be attributed to local constraints enabling better reflection of the flood’s microscale dynamics.

5.3 Ablation Studies

Table 3 summarizes the ablation studies performed to identify the key factors contributing to the model’s performance.

Global Inflow Feature. The exclusion of the global inflow Q_{in} significantly affected the predictive accuracy of the model. The inclusion of this global feature [Battaglia *et al.*, 2018] for all nodes is an approach directly adopted from HydroGraphNet [Taghizadeh *et al.*, 2025], which demonstrated the same degradation in performance when omitted. This could be attributed to the flow-dominant nature of the flood events in the dataset. We enhanced the global inflow feature

Model	Node RMSE ↓ (m^3)	Edge RMSE ↓ (m^3/s)
DUALFloodGNN	$2,216.27 \pm 868.21$	25.88 ± 9.64
No Inflow Feature	$20,562.71 \pm 8,054.11$	423.40 ± 246.11
Global Loss Only	$2,794.10 \pm 851.61$	48.09 ± 14.52
Local Loss Only	$2,116.45 \pm 845.44$	26.07 ± 8.55
No Physics Loss	$2,640.42 \pm 1011.80$	46.86 ± 15.81
HydroGraphNet	$7,584.51 \pm 3,068.54$	-
No Inflow Feature	$20,867.39 \pm 6697.08$	-

Table 3: Ablation study comparing RMSE across different model configurations. The best metrics are highlighted in bold.

by treating it as a dynamic feature, also including the previous boundary conditions as model input.

Physics-informed Loss Three additional physics-informed variants were considered: only global loss, only local loss and no physics-informed loss. In contrast to applying only global mass balance, enforcing local-level regularization resulted in a lower RMSE for both node and edge prediction, surpassing the accuracy of the purely data-driven model. This can be attributed to the co-dependent modeling of node and edge outputs in the local physics loss. A model with physics constraints on both scales demonstrates comparable overall performance to one with only local loss, though the superior variant varies across testing folds. Thus, selection between the two should be driven by empirical evaluation on the target dataset.

6 Conclusion

In conclusion, this study presents DUALFloodGNN, a novel physics-informed flood GNN model, that explicitly incorporates mass conservation at both a global and local scale through the loss function. To this end, it performs simultaneous node and edge prediction, jointly modeling these values through a shared message passing operation. To enhance the model’s robustness, it is autoregressively trained with a dynamic curriculum learning strategy, which adjusts the learning period in response to the current difficulty. DUALFloodGNN outperforms standard GNN architectures and state-of-the-art GNN flood models in the prediction of hydrodynamic variables and the classification of flooded locations. Additionally, the model maintains computational efficiency comparable to other flood-specific GNN approaches, demonstrating its suitability for operational use cases.

References

- [Ahmadlou *et al.*, 2021] Mohammad Ahmadlou, A'kif Al-Fugara, Abdel Rahman Al-Shabeb, Aman Arora, Rida Al-Adamat, Quoc Bao Pham, Nadhir Al-Ansari, Nguyen Thi Thuy Linh, and Hedieh Sajedi. Flood susceptibility mapping and assessment using a novel deep learning model combining multilayer perceptron and autoencoder neural networks. *Journal of Flood Risk Management*, 14(1):e12683, 2021.
- [Ashraf *et al.*, 2024] Inaam Ashraf, Janine Strotherm, Luca Hermes, and Barbara Hammer. Physics-informed graph neural networks for water distribution systems. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(20):21905–21913, 3 2024.
- [Battaglia *et al.*, 2018] Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, Caglar Gulcehre, Francis Song, Andrew Ballard, Justin Gilmer, George Dahl, Ashish Vaswani, Kelsey Allen, Charles Nash, Victoria Langston, Chris Dyer, Nicolas Heess, Daan Wierstra, Pushmeet Kohli, Matt Botvinick, Oriol Vinyals, Yujia Li, and Razvan Pascanu. Relational inductive biases, deep learning, and graph networks, 2018.
- [Bentivoglio *et al.*, 2023] R. Bentivoglio, E. Isufi, S. N. Jonkman, and R. Taormina. Rapid spatio-temporal flood modelling via hydraulics-based graph neural networks. *Hydrology and Earth System Sciences*, 27(23):4227–4246, 2023.
- [Bentivoglio *et al.*, 2024] R. Bentivoglio, E. Isufi, S. N. Jonkman, and R. Taormina. Multi-scale hydraulic graph neural networks for flood modelling. *EGUsphere*, 2024:1–28, 2024.
- [Bronstein *et al.*, 2017] Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: Going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, July 2017.
- [Brunner, 2025] Gary W. Brunner. *HEC-RAS Hydraulic Reference Manual*. Hydrologic Engineering Center, Institute for Water Resources, U.S. Army Corps of Engineers, 2025. Accessed: 2025-12-16.
- [Costabile *et al.*, 2017] Pierfranco Costabile, Carmelina Costanzo, and Francesco Macchione. Performances and limitations of the diffusive approximation of the 2-d shallow water equations for flood simulation in urban and rural areas. *Applied Numerical Mathematics*, 116:141–156, 2017. New Trends in Numerical Analysis: Theory, Methods, Algorithms and Applications (NETNA 2015).
- [Devia *et al.*, 2015] Gayathri K. Devia, B.P. Ganasri, and G.S. Dwarakish. A review on hydrological models. *Aquatic Procedia*, 4:1001–1007, 2015. International Conference on Water Resources, Costal AND Ocean Engineering (ICWRCOE'15).
- [Fang *et al.*, 2021] Zhice Fang, Yi Wang, Ling Peng, and Haoyuan Hong. Predicting flood susceptibility using lstm neural networks. *Journal of Hydrology*, 594:125734, 2021.
- [Farahmand *et al.*, 2023] Hamed Farahmand, Yuanchang Xu, and Ali Mostafavi. A spatial–temporal graph deep learning model for urban flood nowcasting leveraging heterogeneous community features. *Scientific Reports*, 13(1):6768, 4 2023.
- [Fey and Lenssen, 2019] Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric, 2019.
- [Gao *et al.*, 2022] Han Gao, Matthew J. Zahr, and Jian-Xun Wang. Physics-informed graph neural galerkin networks: A unified framework for solving pde-governed forward and inverse problems. *Computer Methods in Applied Mechanics and Engineering*, 390:114502, 2022.
- [Gilmer *et al.*, 2017] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry, 2017.
- [Hamilton *et al.*, 2018] William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs, 2018.
- [Herath *et al.*, 2025] Herath Mudiyanselage Viraj Vidura Herath, Lucy Marshall, Abhishek Saha, Sanka Rasnayaka, and Sachith Seneviratne. Subgrid informed neural networks for high-resolution flood mapping. *Journal of Hydrology*, 660, 10 2025.
- [Hosseiny, 2021] Hossein Hosseiny. A deep learning model for predicting river flood depth and extent. *Environmental Modelling & Software*, 145:105186, 2021.
- [Hu *et al.*, 2020] Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. Strategies for pre-training graph neural networks, 2020.
- [Jiang *et al.*, 2019] Xiaodong Jiang, Pengsheng Ji, and Sheng Li. Censnet: Convolution with edge-node switching in graph neural networks. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 2656–2662. International Joint Conferences on Artificial Intelligence Organization, 7 2019.
- [Jiang *et al.*, 2024] Jiange Jiang, Chen Chen, Yang Zhou, Stefano Berretti, Lei Liu, Qingqi Pei, Jianming Zhou, and Shaohua Wan. Heterogeneous dynamic graph convolutional networks for enhanced spatiotemporal flood forecasting by remote sensing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:3108–3122, 2024.
- [Kazadi *et al.*, 2024] Arnold Kazadi, James Doss-Gollin, Antonia Sebastian, and Arlei Silva. Floodgnn-gru: a spatio-temporal graph neural network for flood prediction. *Environmental Data Science*, 3:e21, 2024.
- [Kipf and Welling, 2017] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks, 2017.

- [Kumar *et al.*, 2023] Vijendra Kumar, Kul Vaibhav Sharma, Tommaso Caloiero, Darshan J. Mehta, and Karan Singh. Comprehensive overview of flood modeling approaches: A review of recent advances. *Hydrology*, 10(7), 2023.
- [Luo *et al.*, 2024] Yuxuan Luo, Yanlai Zhou, Hua Chen, Lihua Xiong, Shenglian Guo, and Fi-John Chang. Exploring a spatiotemporal hetero graph-based long short-term memory model for multi-step-ahead flood forecasting. *Journal of Hydrology*, 633:130937, 2024.
- [Nash and Sutcliffe, 1970] J.E. Nash and J.V. Sutcliffe. River flow forecasting through conceptual models part i — a discussion of principles. *Journal of Hydrology*, 10(3):282–290, 1970.
- [Neo *et al.*, 2025] Sun Han Neo, Sachith Seneviratne, Herath Mudiyanseilage Viraj Vidura Herath, Abhishek Saha, Sanka Rasnayaka, and Lucy Amanda Marshall. Flood-ldm: Generalizable latent diffusion models for rapid and accurate zero-shot high-resolution flood mapping, 2025.
- [Oliveira Santos *et al.*, 2023] Victor Oliveira Santos, Paulo Alexandre Costa Rocha, John Scott, Jesse Van Griensven Thé, and Bahram Gharabaghi. A new graph-based deep learning model to predict flooding with validation on a case study on the humber river. *Water*, 15(10), 2023.
- [Rahimzad *et al.*, 2021] M. Rahimzad, A. Moghaddam Nia, H. Zolfonoon, et al. Performance comparison of an lstm-based deep learning model versus conventional machine learning algorithms for streamflow forecasting. *Water Resources Management*, 35(12):4167–4187, 2021.
- [Rentschler *et al.*, 2022] Jun Rentschler, Melda Salhab, and Bramka Arga Jafino. Flood exposure and poverty in 188 countries. *Nature Communications*, 13(1):3527, 6 2022.
- [Roudbari *et al.*, 2024] Naghmeh Shafiee Roudbari, Shubham Rajeev Punekar, Zachary Patterson, Ursula Eicker, and Charalambos Poullis. From data to action in flood forecasting leveraging graph neural networks and digital twin visualization. *Scientific Reports*, 14(1):18571, 8 2024.
- [Sharma *et al.*, 2023] Pushan Sharma, Wai Tong Chung, Bassem Akoush, and Matthias Ihme. A review of physics-informed machine learning in fluid mechanics. *Energies*, 16(5), 2023.
- [Sun *et al.*, 2023] Alexander Y. Sun, Zhi Li, Wonhyun Lee, Qixing Huang, Bridget R. Scanlon, and Clint Dawson. Rapid flood inundation forecast using fourier neural operator, 2023.
- [Tabari, 2020] Hossein Tabari. Climate change impact on flood and extreme precipitation increases with water availability. *Scientific Reports*, 10(1):13768, 2020.
- [Taghizadeh *et al.*, 2025] Mohammad Taghizadeh, Zahra Zandsalimi, Mohammad Amin Nabian, Majid Shafiee-Jood, and Negin Alemazkoor. Interpretable physics-informed graph neural networks for flood forecasting. *Computer-Aided Civil and Infrastructure Engineering*, pages 1–21, 2025.
- [Teutsch and Mäder, 2022] Philipp Teutsch and Patrick Mäder. Flipped classroom: Effective teaching for time series forecasting, 2022.
- [Thangamuthu *et al.*, 2023] Abishek Thangamuthu, Gunjan Kumar, Suresh Bishnoi, Ravinder Bhattoo, N M Anoop Krishnan, and Sayan Ranu. Unravelling the performance of physics-informed graph neural networks for dynamical systems, 2023.
- [UNDRR and CRED, 2020] UNDRR and CRED. The Human Cost of Disasters: An Overview of the Last 20 Years (2000–2019), 10 2020.
- [Veličković *et al.*, 2018] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks, 2018.
- [Xie *et al.*, 2021] Shuai Xie, Wenyan Wu, Sebastian Mooser, Q.J. Wang, Rory Nathan, and Yuefei Huang. Artificial neural network based hybrid modeling approach for flood inundation modeling. *Journal of Hydrology*, 592:125605, 2021.
- [Xu *et al.*, 2019] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks?, 2019.
- [Yang and Li, 2020] Yulei Yang and Dongsheng Li. Nenn: Incorporate node and edge features in graph neural networks. In Sinno Jialin Pan and Masashi Sugiyama, editors, *Proceedings of The 12th Asian Conference on Machine Learning*, volume 129 of *Proceedings of Machine Learning Research*, pages 593–608. PMLR, 11 2020.
- [Zhang *et al.*, 2024] Zhiyu Zhang, Wenchong Tian, Chenkaixiang Lu, Zhenliang Liao, and Zhiguo Yuan. Graph neural network-based surrogate modelling for real-time hydraulic prediction of urban drainage networks. *Water Research*, 263:122142, 2024.
- [Zhao *et al.*, 2020] Qun Zhao, Yuelong Zhu, Kai Shu, Ding-sheng Wan, Yufeng Yu, Xudong Zhou, and Huan Liu. Joint spatial and temporal modeling for hydrological prediction. *IEEE Access*, 8:78492–78503, 2020.