Fair Graph Learning Using Constraint-Aware Priority Adjustment and Graph Masking in River Networks

Erhu He[1], Yiqun Xie[2], Alexander Sun[3], Jacob Zwart[4], Jie Yang[5], Zhenong Jin[5], Yang Wang[1], Hassan Karimi[1], Xiaowei Jia[1]

[1]University of Pittsburgh
[2]University of Maryland
[3]University of Texas at Austin
[4]U.S. Geological Survey
[5]University of Minnesota
[1]{erh108, yaw70, hkarimi, xiaowei}@pitt.edu, [2]xie@umd.edu, [3]alex.sun@beg.utexas.edu, [4]jzwart@usgs.gov,
[5]{yang8884, jinzn}@umn.edu

Abstract

Accurate prediction of water quality and quantity is crucial for sustainable development and human well-being. However, existing data-driven methods often suffer from spatial biases in model performance due to heterogeneous data, limited observations, and noisy sensor data. To overcome these challenges, we propose Fair-Graph, a novel graph-based recurrent neural network that leverages interrelated knowledge from multiple rivers to predict water flow and temperature within large-scale stream networks. Additionally, we introduce node-specific graph masks for information aggregation and adaptation to enhance prediction over heterogeneous river segments. To reduce performance disparities across river segments, we introduce a centralized coordination strategy that adjusts training priorities for segments. We evaluate the prediction of water temperature within the Delaware River Basin, and the prediction of streamflow using simulated data from U.S. National Water Model in the Houston River network. The results showcase improvements in predictive performance and highlight the proposed model's ability to maintain spatial fairness over different river segments.

Introduction

Freshwater plays a critical role at the intersection of global economic, food, and energy services (Konar et al. 2011; Hoekstra and Mekonnen 2012; Carr et al. 2013), but continues to suffer from increasing demands for water-based ecosystem services and a shifting climate. The assessment of water quality and water quantity in freshwater ecosystems is critical for the future sustainability of the planet and human well-being. Accurate predictions of seasonal changes, temperature trends, and flow variations can help support decision and policy making on water management (e.g., reservoir operations), maintain the desired habitat for aquatic life, and detect disasters (e.g., floods and droughts) at an early stage. This work is focused on predicting water temperature and streamflow in stream networks.

Given the importance of this problem, researchers from multiple domains, including hydrology, meteorology, and environmental engineering, have been developing physics-based models to simulate water dynamics at all scales (Markstrom 2012; Regan et al. 2018), but these models are only approximations of reality due to incomplete or missing knowledge of certain processes or excessive complexity in modeling these processes (Gupta and Nearing 2014; Lall 2014). Recently, graph neural networks (GNNs) have been widely adopted as a data-driven solution to model stream networks (Jia et al. 2021b; Moshe et al. 2020; Sun et al. 2021; Topp et al. 2023; Chen, Zwart, and Jia 2022; Jia et al. 2021a; Chen et al. 2021; Jia et al. 2023) as they can learn to capture complex interactions amongst stream segments (e.g., through mass advection and diffusion).

Despite promising results demonstrated by initial tests at small scales, existing GNN methods are often spatially biased and limited in their accuracy when applied to modeling large-scale stream networks due to several reasons. First, data are highly heterogeneous due to spatial variation of stream characteristics (e.g., soil properties, channel geometry, and roughness), which are difficult to measure and cannot be included in the features. Thus, the mapping from input to the target variables can vary across different stream segments. Moreover, the effect of water flows from upstream to downstream segments can also be different. Second, water temperature and streamflow observations are available for a small subset of well-observed stream segments while other segments have much less or no observations. Third, the data can be noisy due to sensor-induced measurement errors and limited spatial resolution of weather data. Due to these reasons, existing models often compromise the performance of certain sites in exchange of better performance at other sites. Such biases may unintentionally induce unfair distribution of social resources (e.g., subsidies and assistance) and treatment with respect to environmental policies, especially for low-income and remote regions where no gauging stations have been built or fewer field studies have been conducted to collect high-quality data. Thus, addressing this disparity in model performance is essential, as it substantially affects strategic decisions in water resource management and ecological policy-making, particularly in enhancing support for regions with limited data.

Prior works have investigated different approaches to enforce the fairness for general machine learning (ML) models (Kamishima, Akaho, and Sakuma 2011; Alasadi, Al Hilli, and Singh 2019; Zhang and Davidson 2021) and GNNs (Bose and Hamilton 2019; Kang et al. 2020; Tang et al. 2020). Many of these works focus on ensuring the balance of output distribution (e.g., equal opportunity (Hardt, Price, and Srebro 2016)) over different protected groups, which is not suitable for our target problem. Some other works use regularization-based methods to reduce the performance disparity (Kamishima, Akaho, and Sakuma 2011; Kang et al. 2020), but the regularization may intentionally degrade the performance for certain groups to pursue better overall balance. Recently, researchers have also investigated preserving fairness on GNNs by considering the connections amongst different nodes (Tang et al. 2020), but these approaches do not fully capture the heterogeneity in (i) the mapping from input to the target variable, and (ii) the effect each stream segment receives from its upstream segments.

To address these issues, we propose Fair-Graph, a new fairness-preserving method on graph models that mitigates model bias over space and enhances performance on individual nodes using two learning strategies. First, we introduce a centralized coordination strategy to adjust training priorities over different stream segments, as inspired by previous works on spatial fairness (Xie et al. 2022; He et al. 2022, 2023). However, increasing the priority for some stream segments may not necessarily improve their performance due to the limited data quantity and data quality, and may degrade performance for other regions. Hence, we create a constraint-aware priority adjustment strategy by incorporating a performance upper-bound for each stream segment. To further address the data heterogeneity and improve the performance on individual nodes, the Fair-Graph method will extract node-specific patterns in aggregating the information from neighboring nodes and making predictions from graph-based node embeddings. Specifically, we propose to learn two graph masks in the aggregation phase and the adaptation phase of the graph learning. The aggregation mask determines how each stream segment is affected by other nearby segments while the adaptation mask selects the observations from other segments and uses them to jointly adapt the graph model to each individual stream segment. The constraint-aware priority adjustment strategy and the graph masking strategy will be coupled together in the training process. As a result, the GNN model will learn node embeddings that maximize the fairness after learning customized aggregation and adaptation using graph masking.

Our experiments demonstrate the superiority of the proposed method over a diverse set of baselines in enhancing the prediction and preserving the spatial fairness on two large-scale heterogeneous river basins, the Delaware River Basin and the Houston River network. We also show that the aggregation mask and adaptation mask can extract meaningful patterns on the graph of stream networks.

Problem Formulation and Preliminaries

Problem definition: The objective of this work is to develop a robust model for capturing the dynamics of water temperature and streamflow in river networks. Specifically, we consider N river segments within a connected river network. Each river segment, indexed by i, is associated with a set of input features $\mathbf{X}_i$ observed at various time steps (e.g., dates), represented as $\{\mathbf{x}_i^1, \mathbf{x}_i^2, ..., \mathbf{x}_i^T\}$. These input features include variables describing the environmental conditions, such as daily average solar radiation, air temperature, precipitation, and wind speed specific to each segment. Additionally, observed target variables $\mathbf{Y}_i^t$ (i.e., water temperature or streamflow) are available for certain time steps $t \in \{1, ..., T\}$ and certain segments $i \in \{1, ..., N\}$.

Graph representation of stream networks: Graphs have been commonly used to represent multiple stream segments and their interactions in a stream network (Sun et al. 2021). Formally, a graph is denoted by $G = \{\mathcal{V}, \mathcal{E}\}$ where the node set V contains the set of river segments, the edge set E represents the connections between stream segments, e.g., from upstream to downstream segments. The edges can also be weighted, e.g., based on the inverse of stream distance between two river segments. The edge weights are stored in a node adjacency matrix $\mathbf{A}$. The use of GNNs on this graph structure enables information propagation to facilitate predictive learning in two ways: (i) when predicting each stream segment, GNN aggregates useful information from other segments that contributes to the prediction of target variables, e.g., streamflow can be affected by rainfall at upstream river segments. (ii) GNN models can facilitate the information sharing among well-observed stream segments and poorly observed stream segments.

Spatial fairness: Here we introduce the spatial fairness measure M_{fair}, which is defined on a spatial partitioning $\mathcal{P}$ (Xie et al. 2022). The partitioning $\mathcal{P}$ splits a study region into multiple partitions, i.e., different stream segments, as $\mathcal{P} = \{p | \forall p \in \mathcal{P}\}$. The fairness over a spatial partitioning $\mathcal{P}$ aims to ensure the balance of model performance over all the space partitions p that are contained in $\mathcal{P}$. First, we consider a metric $M_{\mathcal{F}}$ used to evaluate the performance of a model $\mathcal{F}$, e.g., root mean squared error (RMSE). Another key variable needed for the fairness definition is $E_{\mathcal{P}}$, which measures the mean model performance over all the partitions. This is implemented as the overall performance of a base model $\mathcal{F}_{\mathbf{\Theta}_0}$ over all the partitions, as $E_{\mathcal{P}} = M_{\mathcal{F}}(\mathcal{F}_{\mathbf{\Theta}_0}, \{\cup p | \forall p \in \mathcal{P}\})$, where parameters $\mathbf{\Theta}_0$ are trained without any consideration of spatial fairness. Intuitively, if the model performance on a specific partition p, i.e., $M_{\mathcal{F}}(\mathcal{F}_{\mathbf{\Theta}}, p)$, has a large deviation from the overall mean performance $E_{\mathcal{P}}$, the model $\mathcal{F}_{\mathbf{\Theta}}$ is potentially unfair across partitions. This can be formally defined as $M_{fair}(\mathcal{F}_{\mathbf{\Theta}}, M_{\mathcal{F}}, \mathcal{P}) = \sum_{p \in \mathcal{P}} \frac{d(M_{\mathcal{F}}(\mathcal{F}_{\mathbf{\Theta}}, p), E_{\mathcal{P}})}{|\mathcal{P}|}$, where $d(\cdot, \cdot)$ is the absolute distance in our test.

Proposed Method

Model Architecture

We build an ML model architecture $\mathcal{F}$ for modeling the water dynamics in river segments by capturing their spatial and temporal dependencies (Fig. 1). For each stream segment, its thermal status and water quantity change gradually based on the current weather input and its historical state. We capture this temporal dependency by using a long short-term

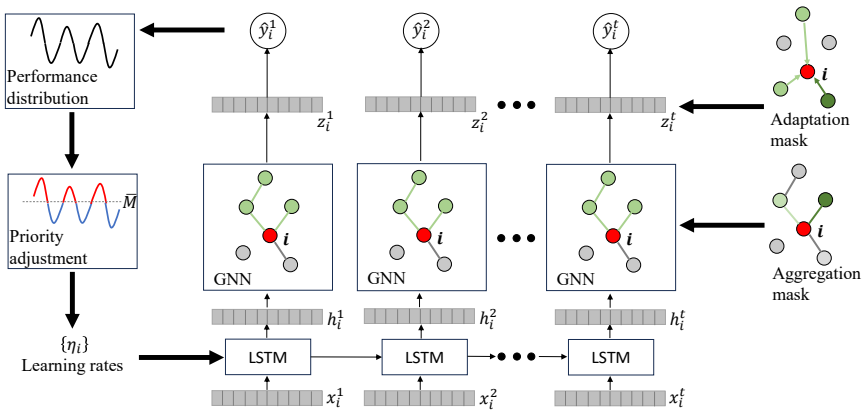


Figure 1: A diagram of the proposed model. For each node i and each time t, the long short-term memory (LSTM) network extracts an embedding $\mathbf{h}_{i,t}$. Then we apply a GNN to refine each time's embedding by aggregating information from neighboring nodes (highlighted in green), producing a new embedding $\mathbf{z}_{i,t}$. Finally, the fully connected layers output the prediction $\hat{\mathbf{y}}_i^t$.

memory (LSTM) network layer, which generates the hidden presentation $\mathbf{h}_{i,t}$ for each segment i at each time step t by combining the current input $\mathbf{x}_{i,t}$ and the previous LSTM state and hidden representation $\mathbf{h}_{i,t-1}$.

The water temperature and streamflow at each river segment can also be affected by the water advected from upstream river segments. Hence, after gathering the hidden representations for all the segments, we use L graph convolutional layers to capture the interactions between nearby stream segments. Formally, the l-th layer of graph convolution can be expressed as follows:

$$\begin{aligned}\mathbf{z}_{i,t}^{(l)} &= g_a^{(l)}([\mathbf{z}_{i,t}^{(l-1)}, \mathbf{a}_{i,t}^{(l)}]; \theta_a^{(l)}), \text{for } l \in \{1, ..., L\}, \\ \mathbf{a}_{i,t}^{(l)} &= \text{Pooling}(\{\mathbf{z}_{j,t}^{(l-1)}, \forall j \in \mathcal{N}(i)\}, \mathbf{A}_{i.}),\end{aligned} \tag{1}$$

where $\mathbf{z}_{i,t}^{(0)} = \mathbf{h}_{i,t}$, $\mathcal{N}(i)$ is the set of neighboring river segments for i, and $\theta_a^{(l)}$ represents the parameters in the l-th graph convolution layer. The latent representation $\mathbf{a}_{i,t}^{(l)}$ embeds the information from neighboring river segments, and is obtained through a weighted pooling of embeddings from the neighbors based on the weights from the adjacency matrix $\mathbf{A}$. We concatenate $\mathbf{a}_{i,t}^{(l)}$ with the last GNN layer's embedding of the target node $\mathbf{z}_{i,t}^{(l-1)}$ before the transformation using the function $g_a(\cdot)$. In this work, we adopt the GraphSAGE method (Hamilton, Ying, and Leskovec 2017) to implement the function $g_a(\cdot)$. Finally, we stack fully connected output layers to convert the aggregated embeddings $\mathbf{z}_{i,t}$ to the predicted output $\hat{\mathbf{y}}_{i,t} = g_o(\mathbf{z}_{i,t}^{(l)}; \theta_o)$, where θ_o represents the parameters in the output layers.

In the proposed method, we allow the output layers g_o to be fine-tuned separately to each node while keeping the LSTM and graph convolutional layers to be shared across nodes. This adaptation process enables the model to differentiate the behavior of different nodes. To fully capture the data heterogeneity over stream segments, we discuss ways to modify the function g_a and g_o for different nodes using graph masking.

Graph Masking for Addressing Data Heterogeneity

Here we introduce two graph masks to modify the aggregation process and the adaptation process, which aim to tackle the performance disparity issue from (1) the data perspective, and (2) the model training perspective, respectively.

Aggregation neighborhood refinement: Edge weights in aggregation inherently determine the influence between connected river segments. However, they are often set uniformly (e.g., 0 or 1) or based on common distance metrics but do not fully reflect physical characteristics that affect the interactions amongst stream segments. For example, the streamflow in one river segment could be affected by the rainfall from multiple upstream segments, and their contributions depend on both their distance to the target segment and their stream characteristics (e.g., soil and groundwater properties), which determines water flow velocity and the conversion from rainfall to surface runoff. Due to the heterogeneous nature of stream characteristics, one challenge is to determine which subset of neighbors contributes most to the prediction on each stream segment.

Existing works on graph attention networks (Velickovic et al. 2017) are based on the similar idea of learning node-specific neighborhood but they rely on input data to predict attention weights on neighbors through a global function. This method remains limited in many spatial datasets as many important physical characteristics that account for heterogeneity (e.g., groundwater, soil properties) can be missing from input data. To address this limitation, we introduce a trainable aggregation mask $\mathcal{M}^a$, which learns the level of contribution of other nodes in the graph convolution process. The aggregation process (Eq. 1) then becomes $\mathbf{z}_{i,t}^{(l)} = g_a^{(l)}([\mathbf{z}_{i,t}^{(l-1)}, \sum_{j \in \mathcal{N}(i)} \mathbf{A}_{ij} \mathcal{M}_{ij}^a \mathbf{z}_{j,t}^{(l-1)}]; \theta_a^{(l)})$. This approach is based on the assumption that the aggregation mask is specific to each spatial location and remains static over time as most stream characteristics are static or change very slowly over time. Directly training the mask $\mathcal{M}^a$ can redefine the neighborhood for the aggregation operation on each entity separately based on its historical data and bet-

ter capture the spatial heterogeneity.

Sample selection in adaptation: We consider adapting the GNN model to each node by fine-tuning the parameters in the output layers θ_o. However, this approach has limits in adapting the model to many stream segments with limited observations. To tackle this challenge, we propose to learn an adaptation mask $\mathcal{M}^s$ to measure the degree of contribution made by different nodes to the adaptation process to a target node i. The loss function for the adaptation process to the node i thus becomes $\mathcal{L}_i^{adp} = \sum_j \mathcal{M}_{ij}^s \mathcal{L}^{sup}(X_j, Y_j)$, where $\mathcal{L}^{sup}(X_j, Y_j)$ denotes the standard supervised loss on the labeled samples from the node j. Different from the standard adaptation process that uses only the data from each node for model fine-tuning, the adaptation mask helps collect data from multiple nodes and use them jointly in the adaptation process. We hypothesize that combining data from multiple entities sharing similar mapping relations $\mathbf{X} \rightarrow \mathbf{Y}$ can help mitigate the data scarcity issue on certain nodes and improve the adaptation performance.

Training graph masks: In the training process, we first initialize the model via global training, i.e., tuning the parameters θ_a and θ_o using the observations from all the nodes and without using the adaptation (i.e., node-specific θ_o) and graph masks. Then we split the observation data for each node into training and validation sets, and update the graph masks through a bi-level optimization process.

- **Inner loop:** In this stage, we fine-tune the model parameters (i.e., shared θ_a and node-specific θ_o) using training data and the current graph masks. Specifically, for each node, the model parameters are updated to minimize the training loss based on the node's training data and the current graph masks. This helps the model refine its internal representations to each node's characteristics and better grasp the complex relationships in the data.
- **Outer loop:** The outer loop evaluates the fine-tuned model using the separate validation data. The loss on the validation data is used to guide the update of the graph masks. Specifically, the graph masks are adjusted in a way that minimizes the validation loss, encouraging the model to focus on the most relevant connections between nodes.

This separation of inner and outer loops enables an iterative refinement process, ultimately leading to the estimation of adaptive graph weights that accurately reflect the underlying relationships amongst the nodes in the graph. We use the graph masks to fine-tune the model parameters, and for nodes with no observations, we use the θ_o parameters learned from the initial model.

Constraint-aware Priority Adjustment

Amongst existing fairness-enforcing methods, the most common strategy is to incorporate additional fairness losses as the term in the loss function (Kamishima, Akaho, and Sakuma 2011; Zafar et al. 2017), e.g., $\mathcal{L} = \sum_i \mathcal{L}_i^{sup}/N + \lambda \cdot \mathcal{L}_{fair}$, where $\sum_i \mathcal{L}_i^{sup}/N$ is the prediction loss (e.g., MSE loss) and λ is a scaling factor. Another popular strategy involves incorporating additional discriminators during training to penalize learned representations that may reveal the identity of a group (e.g., gender) in an adversarial manner (Alasadi, Al Hilli, and Singh 2019; Sweeney and Najafian 2020; Zhang and Davidson 2021). However, these fairness-preserving methods often lead to competition between the predictive performance $\mathcal{L}_{pred}$ and fairness $\mathcal{L}_{fair}$. As a result, the model often intentionally degrades the performance for certain regions to pursue better overall balance. Moreover, existing methods can still be affected by sparse and imbalanced training samples.

Priority adjustment: To mitigate these concerns, we propose a centralized coordination algorithm for enforcing the fairness of the predictive performance over different stream segments. The objective is to elevate the priority for regions with relatively poor predictive performance while considering the performance constraint due to the data quality and quantity in each segment. As inspired by prior works (Xie et al. 2022; He et al. 2022, 2023), we introduce a global referee to regularly evaluate the performance disparity during the training process and identify the stream segments that are under-represented by the current predictive model $\mathcal{F}$. Then the referee will adjust the learning rate for different stream segments based on their relative performance. In each iteration, the referee evaluates the performance (e.g., RMSE) $M_{\mathcal{F}}(i)$ on each river segment $i \in \{1, ..., N\}$, and measures its deviation with the overall performance $\bar{M}$. We then change the learning rate η_i for the segment i as

$$\begin{aligned} \eta_i &= \frac{\eta_i' - \eta_{min}'}{\eta_{max}' - \eta_{min}'} \cdot \eta_{init}, \\ \eta_i' &= \max(M_{\mathcal{F}}(i) - \bar{M}, 0), \end{aligned} \tag{2}$$

where η_{init} is the learning rate used to train model $\mathcal{F}$, $\eta_{min}' = \arg\min_{\eta_i'}\{\eta_i' \,|\, \eta_i', \forall i \in \{1, ..., N\}\}$, and $\eta_{max}' = \arg\max_{\eta_i'}\{\eta_i' \,|\, \forall i \in \{1, ..., N\}\}$. In the method description we use RMSE as an example, for which the lower $M_{\mathcal{F}}(i)$ indicates better performance, but the proposed method can be applied to other metrics (e.g., R-squared or Nash–Sutcliffe model efficiency coefficient (NSE)) by slight modifications.

According to Eq. 2, if the performance $M_{\mathcal{F}}(i)$ is worse than the overall performance, its learning rate η_i will increase relatively to other segments. As a result, the samples in this segment will have a higher impact for training the predictive model $\mathcal{F}$. Moreover, all the learning rates after the update are normalized back to the range $[0, \eta_{init}]$ to keep the optimization process stable.

Constraint-aware adjustment: We assume there is a performance upper-bound $M_{\mathcal{F}}^*(i)$ that can be achieved in each segment by using any subset of training data from this segment. Such a performance upper-bound exists because of the data paucity or data quality issues in the segment. When optimizing the spatial fairness, the referee needs to consider both the current validation loss and the performance upper-bound for each segment as the ignorance of the performance upper-bound may lead to negative model training. For example, for two segments i and j with their current performance $M_{\mathcal{F}}^*(j) < \bar{M} < M_{\mathcal{F}}(j) < M_{\mathcal{F}}(i) \approx M_{\mathcal{F}}^*(i)$, the fairness only driven algorithm may lower the priority on the segment j or keep elevating the training priority on the segment i,

even though this will not improve the performance on the segment i. As a result, this may degrade the performance on the segment j, i.e., increase $M_{\mathcal{F}}(j)$.

To address this limitation, we propose to incorporate the performance upper-bound in the coordination process, as

$$\eta_i = \frac{\eta'_i - \eta'_{min}}{\eta'_{max} - \eta'_{min}} \cdot \eta_{init}, \\ \eta'_i = \max(M_{\mathcal{F}}(i) - \max(\bar{M}, M^*_{\mathcal{F}}(i)), 0). \tag{3}$$

This helps ensure that the referee will not elevate training priority for segments that have already achieved their performance upper-bound as this may degrade the performance for other segments.

In this work, we approximate the value of $M^*_{\mathcal{F}}(i)$ by the maximum validation performance obtained from multiple rounds of training with randomly selected subsets of training data. Specifically, we adopt a process similar to cross-validation, dividing observed data across varying times and locations into training and validation sets. Then we train and evaluate standard GNNs, including segment-specific fine-tuning, to establish the performance upper-bound. In the test, we observe that most of the segments are unable to reach the performance upper-bound. To bridge the performance gap between these segments and the corresponding upper-bound, we introduce a strategy to align the upper-bound to be more reliable for the distribution of model performance by $\tilde{M}^*_{\mathcal{F}}(i) = M^*_{\mathcal{F}}(i) + (\bar{M} - \bar{M}^*_{\mathcal{F}})$, where $\bar{M}^*_{\mathcal{F}}$ is the mean of performance upper-bound.

Coupled graph masking and priority adjustment: The priority adjustment will be used together with the graph masking, to enable the model to pursue better performance balance while allowing individual nodes to enhance its performance. In the process of priority adjustment, we freeze the parameters specific to each node's output layers and only update the shared parameters (LSTM+GNN). The goal is to learn graph embeddings $\mathbf{z}$ so as to optimize the fairness after the model is fine-tuned on each individual node. Specifically, we perform regular training using learning rates $\{\eta_i\}$ assigned by the referee over data in all individual river segments. Note that when we iterate over each river segment in updating the model parameters, we also combine the data samples from other segments, as guided by the corresponding adaptation mask.

Experiments

Dataset

Stream water temperature prediction in the Delaware River Basin (DRB): The DRB is an ecologically diverse region and a watershed along the east coast of the United States that provides drinking water to over 15 million people (Williamson et al. 2015). The dataset used in our evaluation is from the U.S. Geological Survey's National Water Information System (USGS 2016) and the Water Quality Portal (Read et al. 2017). Observations at a specific latitude and longitude were matched to river segments that vary in length from 48 to 23,120 m. The river segments were defined by the geospatial fabric used for the National Hydrologic Model (Regan et al. 2018), and the river segments are split up to have roughly a 1-day water travel time. Refer to (Oliver et al. 2021) for the full observational dataset. Specifically, DRB contains 456 stream segments with input features at the daily scale from Jan 01, 1980, to Jul 31, 2020 (14,823 dates). In the following experiments, we use data from the first 27 years (Jan 01, 1980, to Jan 20, 2007) for training and then test in the next 13 years (Jan 21, 2007, to Jul 31, 2020).

Water flow prediction in Houston River network: Houston (Harris County) Texas is subject to frequent flood hazards. A major flood occurs somewhere in Harris County about every two years, resulting from fluvial, pluvial, tropical cyclone related storm surge, or most likely a compounded impact from all three flooding mechanisms. The Houston River network was extracted from the National Hydrography Dataset Plus (NHDPlus) database (U.S. Geological Survey 2023). The streamflow data were taken from the National Water Model (NWM) reanalysis dataset (v2.0) (NOAA 2023) and aggregated to 3-hr intervals. Houston River network consists of 412 river segments with input features at the 3-hr interval from Jan 01, 2000, to Dec 31, 2018 (55,520 time steps). In the following experiments, we use data from the first 15 years (Jan 01, 2000, to Aug 04, 2015) for training and then test in the next 4 years (Aug 05, 2015, to Dec 31, 2018)

Implementation Details

We implement the proposed method using Tensorflow 2 under the environment of Windows 10, CPU i9 11900F, and GeForce RTX 3080 GPU. Our implementation is released[1].

For the DRB dataset, we generate the adjacency matrix $\mathbf{A}$ based on the river distance between each pair of river segment outlets, represented as $dist(i, j)$. We standardize the stream distance and then compute the graph edge weights as $\mathbf{A}_{ij} = 1/(1 + exp(dist(i, j)))$. For the Houston River network dataset, the adjacency matrix includes two-hop neighboring rivers and is unweighted.

We first train a base model with the initial adjacency matrix for 200 epochs (converged) without considering the fairness, using Adam ($\alpha = 0.002$) as the optimizer. From this base model, we implement different candidate approaches to improve the predictive and fairness performance. We use the first 2/3 of the dataset for training and the remaining 1/3 for testing. Within the training data, the last 1/3 of time steps are further separated as validation data for graph masks learning and priority adjustment.

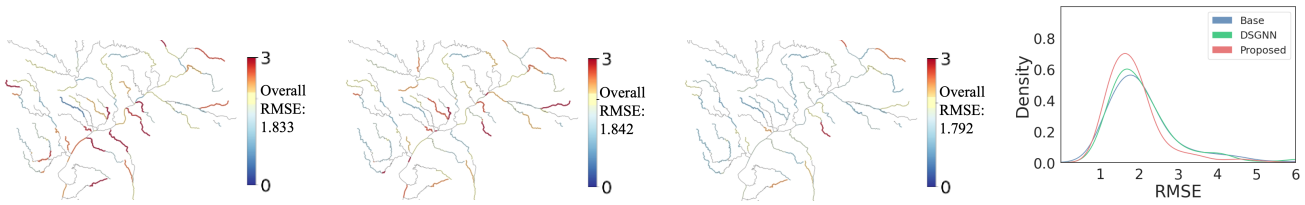
Candidate Methods

We compare our method with the following baseline methods: the base LSTM+GNN model using the initial adjacency matrix without consideration of spatial fairness (Base), the regularization-based fairness enforcement method (REG) (Kamishima et al. 2012), the adversarial discriminator-based fairness enforcement approach (ADL) (Alasadi, Al Hilli, and Singh 2019), and the graph-based fairness-preserving algorithm using degree-specific parameters (DSGNN) (Tang et al. 2020). For a fair comparison, the proposed method and all the baselines are built upon the base model (Base).

[1] https://github.com/ai-spatial/Fair-Graph

Method	Delaware River Basin		Houston River network	
	RMSE	Fairness	NSE	Fairness
Base	1.833	0.616	0.721	0.113
REG	1.824	0.581	0.726	0.086
ADL	1.843	0.591	0.723	0.095
DSGNN	1.842	0.584	0.729	0.094
Fair-Graph$_{ma}$	1.807	0.605	0.732	0.123
Fair-Graph$_{ma+ft}$	1.804	0.601	0.737	0.114
Fair-Graph$_{ma+mo}$	1.795	0.596	0.740	0.111
Fair-Graph$_{pa}$	1.830	0.556	0.723	**0.075**
Fair-Graph$_{cpa}$	1.834	**0.532**	0.737	0.088
Fair-Graph$_{pa+ma+mo}$	**1.786**	0.552	0.738	0.078
Fair-Graph$_{cpa+ma+mo}$	1.792	0.537	**0.758**	0.087

Table 1: The fairness and predictive performance. For root mean squared error (RMSE) and Fairness, lower is better; for Nash–Sutcliffe model efficiency coefficient (NSE), higher is better. Bold indicates the best performing model for a given metric and region.



(a) (b) (c) (d)

Figure 2: The distributions of predictive root mean squared error (RMSE) by (a) Base, (b) the graph-based fairness-preserving algorithm using degree-specific parameters (DSGNN), and (c) the proposed model in the Delaware River Basin. (d) The kernel density plot for the predictive performance across stream segments in the Delaware River Basin.

Additionally, we include variants of the proposed methods to show the effectiveness of each component in the proposed method. For the graph masking approach, we implement the proposed method using only the aggregation mask and shared output layers (Fair-Graph$_{ma}$), using the aggregation mask and node-specific output layers tuned using only the observations from each node (Fair-Graph$_{ma+ft}$), and using both the aggregation mask and the adaptation mask (Fair-Graph$_{ma+mo}$). For the constraint-aware priority adjustment approach, we implement the proposed method using only the priority adjustment following Eq. 2 (Fair-Graph$_{pa}$), using the constraint-aware priority adjustment with performance upper-bound following Eq. 3 (Fair-Graph$_{cpa}$), using graph masking and priority adjustment (Fair-Graph$_{pa+ma+mo}$), and the complete version using graph masking and constraint-aware priority adjustment (Fair-Graph$_{cpa+ma+mo}$).

Results

Overall accuracy and fairness evaluation: Table 1 presents a comprehensive overview of the performance of our proposed methods and other baseline models. For water temperature prediction, we use the RMSE as the evaluation metric (average error magnitude between predicted and observed). For water flow prediction, we use NSE (Nash et al. 1970), which is a widely used metric for the evaluation of streamflow in hydrology. The value of NSE ranges in $(-\infty,1]$ and the higher value indicates better performance. The fairness performance is measured as the mean absolute distance between the overall performance and the performance of individual segments.

The proposed method outperforms the baselines (Base, REG, ADL, DSGNN) in both accuracy and fairness (Table 1). Moreover, the results show the effectiveness of using both graph masking and priority adjustment. The proposed Fair-Graph$_{ma+mo}$ method outperforms the base model, which shows the benefit of incorporating aggregation and adaptation masks. The improvement from Base to Fair-Graph$_{ma}$ confirms the benefit of refining graph neighborhood in better capturing the effect of weather input and water flows across different river segments. Moreover, Fair-Graph$_{ma+mo}$ performs better than Fair-Graph$_{ma+ft}$. This is because the integration of the adaptation mask can mitigate the data paucity issue by identifying and leveraging beneficial training samples from other river segments.

In addition, we observe that Fair-Graph$_{pa}$ and Fair-Graph$_{cpa}$ improve the fairness compared to the base model and other baselines (REG, ADL, DSGNN). The methods using both priority adjustment and graph masking (Fair-Graph$_{pa+ma+mo}$ and Fair-Graph$_{cpa+ma+mo}$) generally have similar fairness performance as the methods using only priority adjustment (Fair-Graph$_{pa}$ and Fair-Graph$_{cpa}$),

while exhibiting improved accuracy. This demonstrates that the performance enhancement brought by graph masking does not compromise the overall fairness. The incorporation of the upper-bound constraint into priority adjustment presents a trade-off between enhancing fairness and predictive performance. Consistently training under-performing river segments that have already reached the performance upper-bound can negatively affect the performance on other segments and thus limit overall accuracy. One potential reason why Fair-Graph$_{pa+ma+mo}$ and Fair-Graph$_{pa}$ have better fairness performance than their counterparts using performance upper-bound (Fair-Graph$_{cpa+ma+mo}$ and Fair-Graph$_{cpa}$) is that the performance of well-performing segments get degraded due to consistent training on under-performing segments as performed by the priority adjustment without considering performance upper-bound. Thus, the performance of these well-performing segments gets closer to the overall performance, leading to a fairness improvement.

Detailed fairness evaluation: Fig. 2 (a)-(c) shows the distributions of RMSE for a subset of segments by Base, DSGNN, and the proposed Fair-Graph$_{cpa+ma+mo}$ model. The proposed method effectively reduces the RMSE for segments that are poorly modeled by Base and DSGNN. Fig. 2 (d) presents the density distributions for the RMSE for river temperature predictions by the base model, the DSGNN model, and the proposed model (Fair-Graph$_{cpa+ma+mo}$). The distribution of RMSE for the proposed model is narrower, which indicates an improvement in fairness. Moreover, our proposed model improves on poorly performing segments (right section of RMSE distribution) while maintaining the competitive performance in well-performing segments.

Also, it is important to point out the real-world impact of these improvements in predictive performance and fairness. The changes in performance values for stream network predictions are small but environmentally significant. For example, subtle changes in water temperature (e.g., less than 0.1 degree) can greatly affect the aquatic environment for fish growth (Letcher et al. 2015). Such small changes can also lead to strong dynamics of concentration of nutrients in the water (Hanson et al. 2020).

Graph masking analysis: Fig. 3 illustrates the learned graph masks for a specific node (highlighted in red) in the context of predicting stream water temperature. Fig. 3 (b) shows that certain river features have been excluded in the aggregation phase due to their limited effect on improving predictions for the target node. The reason is two-fold: (1) the information from these rivers can be redundant to that of other rivers, and (2) certain characteristics of these streams such as groundwater and reservoir impact may diminish their influence on downstream areas. Fig. 3 (c) shows that the river segments selected by the adaptation mask can be located far from the target river. This is because these distant segments might share similar weather conditions and stream characteristics with the target river, and thus contribute to enhancing the prediction for the target river.

Evaluation under perturbed data: To test the proposed method in scenarios with sparser and low-quality data, we randomly select 30 out of the 74 river segments in the DRB that originally possess over 500 observations, and then randomly drop 90% of water temperature observations on these segments (Table 2). The graph masking method improves overall RMSE, which indicates the ability to capture the underlying relationships between different segments. Additionally, the proposed methods achieve comparable fairness measures compared to the test using complete data (Table 1).

(a) (b) (c)

Figure 3: Visualizations of graph masks in the DRB for a specific node (highlighted in red), with green indicating a relatively higher weight. (a): Initial adjacency matrix. (b): The aggregation mask. (c): The adaptation mask.

Method	Delaware River Basin	
	RMSE	Fairness
Base	2.037	0.635
REG	2.011	0.611
ADL	2.052	0.620
DSGNN	2.090	0.615
Fair-Graph$_{ma}$	1.999	0.630
Fair-Graph$_{ma+ft}$	1.985	0.620
Fair-Graph$_{ma+mo}$	1.974	0.615
Fair-Graph$_{pa}$	2.014	0.594
Fair-Graph$_{cpa}$	1.983	0.588
Fair-Graph$_{pa+ma+mo}$	1.970	0.594
Fair-Graph$_{cpa+ma+mo}$	**1.953**	**0.586**

Table 2: The fairness and predictive performance on perturbed data. For root mean squared error (RMSE) and Fairness, lower is better. Bold indicates the best performing model for a given metric and region.

Conclusions

We introduce a new method for enhancing prediction and spatial fairness on graphs using graph masking and constraint-aware priority adjustment. Our experiments on two large-scale heterogeneous river basins have demonstrated the effectiveness of the proposed method. Moreover, the proposed method is shown to learn meaningful stream relationships and benefit graph learning using sparser data. The proposed method is widely applicable to many other applications (e.g., agriculture and traffic management) in which graphs can be used to represent the spatial relationships amongst entities. Future work is planned on graph masking for unmonitored streams and creating new fairness metrics that consider both fairness preservation and predictive accuracy. Additionally, future work could focus on developing a generalized performance upper-bound estimation applicable across diverse models and scenarios.

Acknowledgements

This work was supported by the USGS awards G21AC10564 and G22AC00266, the NSF awards 2147195, 2239175, 2316305, 2105133, and 2126474, the NASA award 80NSSC22K1164, the Momentum award at the University of Pittsburgh, the DRI award at the University of Maryland, and the University of Pittsburgh Center for Research Computing. We thank Galen Gorski and anonymous reviewers for reviewing earlier draft of this manuscript. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Alasadi, J.; Al Hilli, A.; and Singh, V. K. 2019. Toward fairness in face matching algorithms. In *Proceedings of the 1st International Workshop on Fairness, Accountability, and Transparency in MultiMedia*, 19–25.

Bose, A.; and Hamilton, W. 2019. Compositional fairness constraints for graph embeddings. In *International Conference on Machine Learning*, 715–724. PMLR.

Carr, J. A.; D'Odorico, P.; Laio, F.; and Ridolfi, L. 2013. Recent history and geography of virtual water trade. *PloS one*, 8(2): e55825.

Chen, S.; Appling, A.; Oliver, S.; Corson-Dosch, H.; Read, J.; Sadler, J.; Zwart, J.; and Jia, X. 2021. Heterogeneous stream-reservoir graph networks with data assimilation. In *2021 IEEE International Conference on Data Mining (ICDM)*, 1024–1029. IEEE.

Chen, S.; Zwart, J. A.; and Jia, X. 2022. Physics-guided graph meta learning for predicting water temperature and streamflow in stream networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2752–2761.

Gupta, H. V.; and Nearing, G. S. 2014. Debates—The future of hydrological sciences: A (common) path forward? Using models and data to learn: A systems theoretic perspective on the future of hydrological science. *Water Resources Research*.

Hamilton, W.; Ying, Z.; and Leskovec, J. 2017. Inductive representation learning on large graphs. In *Advances in neural information processing systems*, 1024–1034.

Hanson, P. C.; Stillman, A. B.; Jia, X.; Karpatne, A.; Dugan, H. A.; Carey, C. C.; Stachelek, J.; Ward, N. K.; Zhang, Y.; Read, J. S.; et al. 2020. Predicting lake surface water phosphorus dynamics using process-guided machine learning. *Ecological Modelling*, 430: 109136.

Hardt, M.; Price, E.; and Srebro, N. 2016. Equality of opportunity in supervised learning. *Advances in neural information processing systems*, 29.

He, E.; Xie, Y.; Jia, X.; Chen, W.; Bao, H.; Zhou, X.; Jiang, Z.; Ghosh, R.; and Ravirathinam, P. 2022. Sailing in the location-based fairness-bias sphere. In *Proceedings of the 30th International Conference on Advances in Geographic Information Systems*, 1–10.

He, E.; Xie, Y.; Liu, L.; Chen, W.; Jin, Z.; and Jia, X. 2023. Physics Guided Neural Networks for Time-Aware Fairness: An Application in Crop Yield Prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 14223–14231.

Hoekstra, A. Y.; and Mekonnen, M. M. 2012. The water footprint of humanity. *Proceedings of the national academy of sciences*, 109(9): 3232–3237.

Jia, X.; Chen, S.; Zheng, C.; Xie, Y.; Jiang, Z.; and Kalanat, N. 2023. Physics-guided Graph Diffusion Network for Combining Heterogeneous Simulated Data: An Application in Predicting Stream Water Temperature. In *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*, 361–369. SIAM.

Jia, X.; Xie, Y.; Li, S.; Chen, S.; Zwart, J.; Sadler, J.; Appling, A.; Oliver, S.; and Read, J. 2021a. Physics-Guided Machine Learning from Simulation Data: An Application in Modeling Lake and River Systems. In *2021 IEEE International Conference on Data Mining (ICDM)*, 270–279. IEEE.

Jia, X.; Zwart, J.; Sadler, J.; Appling, A.; Oliver, S.; Markstrom, S.; Willard, J.; Xu, S.; Steinbach, M.; Read, J.; et al. 2021b. Physics-guided recurrent graph model for predicting flow and temperature in river networks. In *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)*, 612–620. SIAM.

Kamishima, T.; Akaho, S.; Asoh, H.; and Sakuma, J. 2012. Fairness-aware classifier with prejudice remover regularizer. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2012, Bristol, UK, September 24-28, 2012. Proceedings, Part II 23*, 35–50. Springer.

Kamishima, T.; Akaho, S.; and Sakuma, J. 2011. Fairness-aware learning through regularization approach. In *2011 IEEE 11th International Conference on Data Mining Workshops*, 643–650. IEEE.

Kang, J.; He, J.; Maciejewski, R.; and Tong, H. 2020. Inform: Individual fairness on graph mining. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 379–389.

Konar, M.; Dalin, C.; Suweis, S.; Hanasaki, N.; Rinaldo, A.; and Rodriguez-Iturbe, I. 2011. Water for food: The global virtual water trade network. *Water Resources Research*, 47(5).

Lall, U. 2014. Debates—The future of hydrological sciences: A (common) path forward? One water. One world. Many climes. Many souls. *Water Resources Research*.

Letcher, B. H.; Schueller, P.; Bassar, R. D.; Nislow, K. H.; Coombs, J. A.; Sakrejda, K.; Morrissey, M.; Sigourney, D. B.; Whiteley, A. R.; O'Donnell, M. J.; et al. 2015. Robust estimates of environmental effects on population vital rates: an integrated capture–recapture model of seasonal brook trout growth, survival and movement in a stream network. *Journal of Animal Ecology*, 84(2): 337–352.

Markstrom, S. L. 2012. *P2S–coupled Simulation with the Precipitation-Runoff Modeling System (PRMS) and the Stream Temperature Network (SNTemp) Models*. US Department of the Interior, US Geological Survey.

Moshe, Z.; Metzger, A.; Elidan, G.; Kratzert, F.; Nevo, S.; and El-Yaniv, R. 2020. Hydronets: Leveraging river structure for hydrologic modeling. *arXiv preprint arXiv:2007.00595*.

Nash, J. E.; et al. 1970. River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*.

NOAA. 2023. NOAA: National Water Model CONUS Retrospective Dataset. https://registry.opendata.aws/nwm-archive. Last accessed on 4 May 2023.

Oliver, S. K.; et al. 2021. Predicting water temperature in the Delaware River Basin. U.S. Geological Survey Data Release.

Read, E. K.; et al. 2017. Water quality data for national-scale aquatic research: The Water Quality Portal. *Water Resources Research*.

Regan, R. S.; et al. 2018. Description of the national hydrologic model for use with the precipitation-runoff modeling system (PRMS). Technical report, US Geological Survey.

Sun, A. Y.; Jiang, P.; Mudunuru, M. K.; and Chen, X. 2021. Explore spatio-temporal learning of large sample hydrology using graph neural networks. *Water Resources Research*, 57(12): e2021WR030394.

Sweeney, C.; and Najafian, M. 2020. Reducing sentiment polarity for demographic attributes in word embeddings using adversarial learning. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 359–368.

Tang, X.; Yao, H.; Sun, Y.; Wang, Y.; Tang, J.; Aggarwal, C.; Mitra, P.; and Wang, S. 2020. Investigating and mitigating degree-related biases in graph convoltuional networks. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 1435–1444.

Topp, S. N.; Barclay, J.; Diaz, J.; Sun, A. Y.; Jia, X.; Lu, D.; Sadler, J. M.; and Appling, A. P. 2023. Stream temperature prediction in a shifting environment: Explaining the influence of deep learning architecture. *Water Resources Research*, 59(4): e2022WR033880.

U.S. Geological Survey. 2023. National Hydrography Dataset (ver. USGS National Hydrography Dataset Best Resolution (NHD) for Hydrologic Unit (HU) 4 - 2001). Published 20230101. Accessed April 20, 2023 at https://www.usgs.gov/national-hydrography/access-national-hydrography-products.

USGS. 2016. US Geological Survey. National water information system data available on the world wide web (USGS water data for the nation). doi: 10.5066/F7P55KJN.

Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y.; et al. 2017. Graph attention networks. *stat*, 1050(20): 10–48550.

Williamson, T. N.; et al. 2015. Summary of hydrologic modeling for the Delaware River Basin using the Water Availability Tool for Environmental Resources (WATER). Technical report, U.S. Geological Survey Scientific Investigations Report.

Xie, Y.; He, E.; Jia, X.; Chen, W.; Skakun, S.; Bao, H.; Jiang, Z.; Ghosh, R.; and Ravirathinam, P. 2022. Fairness by "Where": A Statistically-Robust and Model-Agnostic Bi-Level Learning Framework. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Zafar, M. B.; Valera, I.; Gomez Rodriguez, M.; and Gummadi, K. P. 2017. Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In *Proceedings of the 26th international conference on world wide web*, 1171–1180.

Zhang, H.; and Davidson, I. 2021. Towards Fair Deep Anomaly Detection. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 138–148.