https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

RiverGraphNet: Physics-Aware Routing of Gridded Runoff Through Directed River Networks

Mohammad A. Farmani[1], Andrew Bennet[1], Sadaf Moghisi[1], Hoshin Gupta[1], Muhammad Jawad[1], Ali Behrangi[1,2], Ahmad A. Tavakoly[3,4], Guo-Yue Niu[1]

[1]Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ, USA,
[2]Department of Geosciences, University of Arizona, Tucson, AZ, USA,
[3]US Army Engineer Research and Development Center, Coastal and Hydraulics Laboratory, Vicksburg, MS, USA,
[4]Department of Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA,

Corresponding author: Mohammad Farmani, email: farmani@arizona.edu
Guo-Yue Niu, email: niug@arizona.edu

Key Points:

- A physics-aware graph neural network framework (RiverGraphNet) was developed to route physically-generated gridded-runoff through directed river-network topology.
- *RiverGraphNet* substantially improved streamflow-routing performance relative to RAPID across a heterogeneous semi-arid watershed, particularly in headwater and intermediate tributaries.
- Temporal memory representation and training objective strongly shaped routing behavior, highlighting tradeoffs between predictive skill, recession dynamics, and hydrologic realism.

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

Abstract

River routing provides the critical link between runoff generation and downstream streamflow prediction, yet conventional routing models often rely on simplified hydraulic assumptions, fixed parameters, and conservative transport formulations that may limit performance in heterogeneous river systems. Here, we introduce *RiverGraphNet*, a physics-aware graph-based routing framework designed to route physically generated gridded runoff through directed river networks. The framework explicitly isolates routing from runoff generation by coupling *Noah-MP* runoff with a directed river-network graph derived from the *NextGen* hydrofabric for the *Salt–Verde* watershed (Arizona, USA). Gridded runoff is transferred from the Noah-MP domain to graph nodes representing hydrologic routing elements, while streamflow propagation is learned using a *Graph Attention Network* (GAT) informed by physically meaningful node and edge attributes describing drainage structure, terrain, and hydraulic-routing proxies.

RiverGraphNet was evaluated against observed daily streamflow at 23 USGS gauges and compared with RAPID, a widely used physics-based routing benchmark forced with identical *Noah-MP* gridded runoff inputs. Experiments examined the influence of temporal routing memory (3–30 day runoff lags), temporal convolution, and alternative loss functions (MSE, weighted MSE, and JKGE-based objectives). *RiverGraphNet* consistently outperformed RAPID across nearly all gauges and configurations. The best-performing experiment (14-day lag weighted MSE) achieved a median $\mathrm{KG}E_{\{ss\}}$ of 0.72, substantially exceeding RAPID performance (median $\mathrm{KG}E_{\{ss\}}$ of 0.0). Event-scale hydrographs and flow-duration analyses demonstrated improved representation of *peak timing, event magnitude*, and *long-term streamflow distributions*. Attention analysis further revealed that routing improvements were most strongly associated with channel-width and dominant-pathway metrics rather than travel-time proxies alone, suggesting that adaptive representation of network influence provides predictive value beyond conventional travel-time parameterization. These results demonstrate the potential of physics-aware, topology-constrained graph learning as a flexible alternative for routing gridded hydrologic runoff through complex river networks.

Plain Language Summary

Predicting how water moves through river networks is essential for flood forecasting, reservoir management, and water-resource planning. Traditional river-routing models use simplified physical equations to move runoff downstream, but their performance can be limited in complex watersheds where flow behavior varies across space and time. In this study, we developed *RiverGraphNet*, a new routing framework that uses artificial intelligence to learn how water moves through connected river networks while preserving the physical structure of the watershed.

Instead of predicting streamflow directly from weather data, *RiverGraphNet* focuses only on the routing problem. We first generated runoff using a physics-based land surface model (*Noah-MP*), then transferred that runoff into a digital representation of the *Salt–Verde* river network in Arizona. The model uses graph neural networks, a type of machine learning designed for connected systems, to learn how runoff propagates from upstream tributaries to downstream river gauges.

RiverGraphNet consistently outperformed a widely used traditional routing model (RAPID), producing more accurate streamflow predictions across nearly all river gauges. It better reproduced *flood peaks, timing of streamflow events,* and *long-term streamflow behavior*. The results show that combining physically generated runoff with machine learning that respects river-network connectivity can significantly improve routing predictions. This work demonstrates a promising new approach for hydrologic modeling and suggests that future hybrid models combining artificial intelligence with physical conservation laws could further improve realism and reliability.

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

1 Introduction

1.1 Background

Hydrologic and land surface models are widely used to characterize the terrestrial water cycle, and to support applications such as flood forecasting, reservoir operation, climate impact assessment and water resources management (*Cortés-Salazar et al., 2023; Salvati et al., 2024; Farmani et al., 2025; Agnihotri et al., 2023b; Farmani et al., 2024; Tavakoly et al., 2016*). However, realistic streamflow prediction using such models requires not only the partitioning of precipitation into evapotranspiration and runoff, but also the transport routing of runoff through interconnected river networks. River routing models therefore provide the critical link between runoff production and the discharge observed at stream gages, by transforming runoff into spatially and temporally distributed streamflow estimates (*Cortés-Salazar et al., 2023; Salvati et al., 2024; David et al., 2011*).

Physical routing processes govern the timing, attenuation, accumulation, and downstream propagation of runoff signals, strongly influencing hydrograph shape, flood-wave behavior, peak discharge, and baseflow representation (*Salvati et al., 2024; Fenton, 2019; Cortés-Salazar et al., 2023*). Because streamflow depends on both upstream hydrologic conditions and downstream flow propagation through connected river networks, a wide range of routing approaches have been developed for hydrologic application, including Muskingum routing, Muskingum–Cunge routing, kinematic-wave schemes, diffusive-wave formulations, dynamic-wave models, and large-scale vector-based systems such as Routing Application for Parallel Computation of Discharge (RAPID), mizuRoute, CaMa-Flood, and the Model for Scale Adaptive River Transport (MOSART) (*Seo et al., 2026; David et al., 2011; Salvati et al., 2024; Mizukami et al., 2016; Yamazaki et al., 2013; Li et al., 2013*). These approaches are widely used because they provide interpretable representations of flow propagation while preserving mass conservation and maintaining the computational efficiency required for operational hydrologic applications. (*David et al., 2011*). Previous studies have shown that explicitly representing river-network routing can substantially influence streamflow simulations relative to modeling approaches in which runoff transport through river networks is treated implicitly or represented using simplified basin-response functions (*Cortés-Salazar et al., 2023; David et al., 2011*).

Despite these advantages, simplified river-routing approaches such as RAPID and mizuRoute necessarily rely on assumptions and parameterizations that may limit their performance in heterogeneous or highly nonlinear river systems. For example, Muskingum-based and related routing formulations approximate flow translation and attenuation using prescribed routing parameters, which may become less effective under rapidly rising floods, fluctuating hydrographs, highly diffusive flow conditions, or spatially variable channel and floodplain behavior (*Fenton, 2019; Cortés-Salazar et al., 2023*). Many routing frameworks also do not explicitly represent evaporation, seepage, reinfiltration, channel transmission losses, and other runoff–streamflow transformation processes that can be important in semi-arid river systems (*Lima et al., 2025; Nijssen et al., 2001*). In addition, existing routing systems can be difficult to adapt to human regulation, including reservoir operations, diversions, irrigation withdrawals, and managed releases, without additional site-specific parameterization or operational rules (*Cortés-Salazar et al., 2023; Salvati et al., 2024*). These limitations motivate routing approaches that retain physically meaningful river-network structure while allowing more flexible, spatially adaptive representation of runoff propagation through connected river systems.

1.2 The Potential of Graph Neural Networks

Limitations of conventional routing frameworks, and the increasing availability of large hydrologic datasets, have motivated growing interest in artificial intelligence (AI)-based and data-driven approaches for river-network modeling (*Lees et al., 2021; Slater et al., 2025; Moshe et al., 2020; Wang et al., 2025*).

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

Among these approaches, *graph neural networks* (GNNs) have emerged as a particularly promising approach, because river systems naturally form directed spatial graphs in which upstream conditions influence downstream hydrologic responses through structured connectivity pathways. *Sun et al. (2022)* demonstrated that GNNs provide a physically meaningful framework for basin-scale river-network learning by explicitly incorporating river connectivity into streamflow prediction systems. Similarly, *Moshe et al. (2020)* leveraged river-network topology to improve hydrologic prediction across connected basins, while *Xiang and Demir (2021)* showed that distributed runoff pathways and flow-direction relationships can be naturally represented using directed graph structures. Recent studies have further emphasized that river-network structure, hydrological distance, and directional connectivity strongly influence streamflow prediction skill and spatial dependency representation across hydrologic systems (*Seo et al., 2026; Liu et al., 2023*).

Building upon these developments, GNN-based hydrologic frameworks have increasingly been applied to streamflow forecasting and spatiotemporal hydrologic prediction. Existing studies have incorporated graph-based river-network learning for streamflow forecasting (*Sun et al., 2022; Moshe et al., 2020*), graph neural rainfall–runoff models using flow-direction and spatial attributes (*Xiang and Demir, 2021*), graph attention networks for flood propagation and hydrodynamic prediction (*Acosta et al., 2025; Sarkar et al., 2025*), recurrent graph architectures such as GNN–GRU models for flood prediction (*Kazadi et al., 2024*), and spatiotemporal graph neural networks for streamflow prediction across connected river basins (*Akkala et al., 2025*). *Jia et al. (2020)* further proposed *physics-guided recurrent graph networks* that combine graph-based spatial learning with hydrologic process constraints to improve prediction robustness and physical consistency.

Recent STGNN and graph-based flood prediction studies have demonstrated that graph-learning architectures can effectively capture complex spatiotemporal interactions and improve streamflow forecasting performance relative to conventional machine-learning approaches *(Akkala et al., 2025; Roudbari et al., 2024; Sarkar et al., 2025; Wan et al., 2024*). More broadly, recent reviews of deep-learning applications in hydrology have identified GNNs as promising tools for representing non-Euclidean spatial dependencies and distributed hydrologic interactions that cannot be adequately represented using conventional grid-based or independent time-series approaches (*Slater et al., 2025; Zhao et al., 2024*).

More recent studies have begun extending graph-based hydrologic learning beyond streamflow forecasting toward explicit runoff routing and topology-aware flow propagation. *Mosaffa et al. (2026)* introduced a GNN-based routing module designed specifically to represent runoff transport through river networks and argued that many existing deep-learning-based hydrologic models focus primarily on runoff generation while treating routing processes only implicitly. Similarly, *Lima et al. (2025)* and Bindas et al. (2024) explored differentiable routing frameworks for land surface models and highlighted growing interest in machine-learning-based routing emulators capable of approximating traditional routing behavior while reducing computational complexity. In this regard, *Taghizadeh et al. (2025)* and *Acosta et al. (2025)* illustrated the potential of *physics-aware* graph learning for flood propagation and hydrodynamic prediction across spatially-distributed river systems. *Wang et al. (2025)* additionally emphasized the importance of graph topology for improving large-scale flood forecasting and efficient information propagation across river networks while also highlighting challenges associated with long-range graph interactions in large river trees.

Together, these studies demonstrate the growing potential of GNNs for routing-oriented hydrologic modeling and motivate continued development of topology-aware and physics-guided routing frameworks. However, despite recent progress, important scientific gaps remain in graph-based hydrologic routing. Many existing graph-based hydrologic studies focus primarily on direct streamflow forecasting using graph connectivity to improve predictive skill rather than explicitly representing runoff

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

transport through river networks (*Sun et al., 2022; Akkala et al., 2025*). More recent routing-oriented frameworks have combined machine-learning-based runoff generation with graph-based routing modules, such as long short-term memory (LSTM)–GNN architectures in which runoff is generated directly at subbasin or graph-node units prior to routing (*Mosaffa et al., 2026*). Other studies have explored *recurrent neural networks* (RNNs) to route runoff inputs derived from hydrologic models, but typically without explicitly addressing the transfer of physically generated gridded runoff into directed river-network graphs or isolating routing behavior independently from runoff generation processes (*Lima et al., 2025*). Consequently, relatively few studies have investigated how physically generated gridded runoff propagates through topology-aware graph-routing frameworks while evaluating routing realism, upstream–downstream consistency, and agreement with established physics-based routing models.

1.3 Scope of this Investigation

To address these gaps, we developed a physics-aware GNN-based routing framework for hydrologic streamflow prediction and evaluate (for the *Salt-Verde* river system in Arizona) whether GNNs can effectively route physically-generated gridded runoff through a directed river-network topology. The framework is designed as a standalone routing operator that isolates runoff transport from runoff generation. Runoff fields are generated externally by the physically-based Noah-Multi-Parameterization (Noah-MP, citation?) land surface model and transferred to a river-network topology specified using routing elements provided by the NextGen Water Resources Modeling Framework. The runoff is then propagated through a directed graph using topology-aware GNN message passing. The framework incorporates physically meaningful node and edge attributes describing river geometry, drainage structure, terrain characteristics, and hydraulic routing behavior. Together, these components define a physically structured graph-routing system for evaluating graph-based runoff propagation and streamflow prediction across heterogeneous river networks.

Unlike many existing graph-based hydrologic studies that directly predict streamflow or generate runoff at graph nodes, the proposed framework explicitly isolates the routing component by transferring gridded Noah-MP runoff into a physically structured river-network graph prior to routing. GNN-based routing through directed river-connectivity-networks is then used to predict daily streamflow at USGS gauge locations while preserving physically meaningful upstream-downstream flow structure. Model performance is evaluated against both observed and physics-based RAPID-generated streamflow, , with RAPID selected as the physics-based routing benchmark because it is a widely used, computationally efficient river-network routing model that has been extensively developed, evaluated, and applied across regional to global domains, including operational and quasi-operational streamflow forecasting, flood-inundation mapping, water-resources assessment, and hydrologic modeling applications (*Tavakoly et al., 2023*).

By separating runoff generation from routing and explicitly bridging physically generated runoff fields with graph-based river routing, this study evaluates whether topology-aware graph learning can provide an effective and hydrologically meaningful framework for large-scale runoff routing and streamflow prediction.

2 Methods

Figure 1 provides an overview of the RiverGraphNet workflow. The framework consists of four primary components: (1) transfer of gridded Noah-MP runoff to river-network graph nodes, (2) graph-based spatial message passing along the Next Generation Water Model (NGen) river-network topology, (3) learned temporal processing of lagged runoff-response behavior, and (4) extraction of streamflow predictions at graph nodes corresponding to USGS gauge locations.

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

Unlike conventional deep-learning approaches that treat each gauge as an independent time-series prediction problem, the proposed framework explicitly represents the river network as a graph. Nodes represent NGen routing elements associated with flow paths/divides, and directed edges represent hydrologic connectivity between upstream and downstream river elements. This structure constrains information exchange to physically connected routing pathways, allowing the model to learn runoff propagation and streamflow-routing relationships while preserving watershed connectivity.

Gridded Noah-MP runoff is first transferred from the Noah-MP grid to the graph domain. In the divide-based mapping approach, each active runoff grid cell is assigned to the corresponding NGen divide or routing element based on spatial overlap membership, and runoff contributions are aggregated to produce node-aligned runoff inputs. These transferred runoff values provide the local water input to each graph node before graph message passing is applied.

The following sections describe each component of the RiverGraphNet framework in detail. Sections 2.1–2.2 introduce the study area and Noah-MP runoff simulations used to generate gridded runoff forcing. Sections 2.3–2.6 describe the RAPID benchmark, river-network graph construction, grid-to-graph runoff transfer, and graph features. Section 2.7 presents the GAT-based routing model, followed by the training strategy, loss functions, and evaluation procedures in Sections 2.8–2.9.

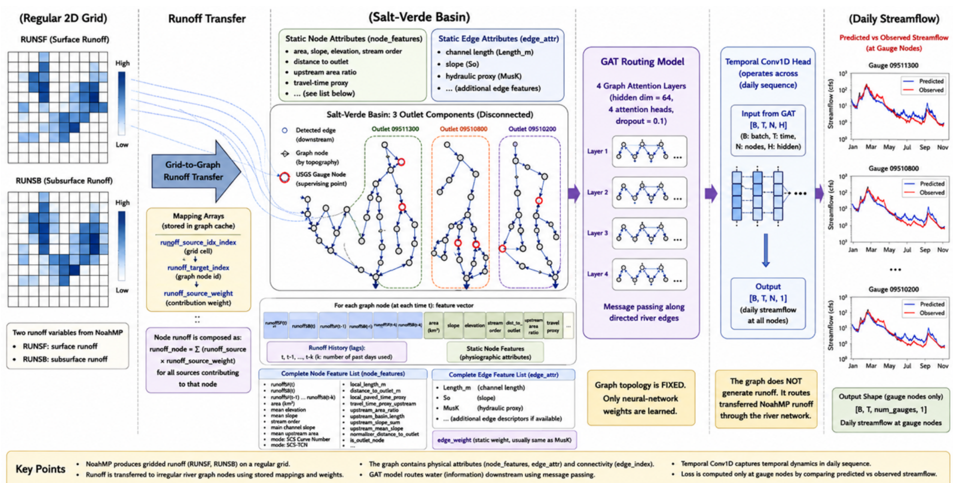


Figure 1. Overview of the proposed graph-based routing framework. Gridded Noah-MP surface and subsurface runoff are transferred to NGen river-network graph nodes using divide-based runoff aggregation. The directed river graph preserves upstream–downstream hydrologic connectivity through graph edges representing river-network flowpaths. A GAT performs topology-aware spatial message passing using graph connectivity and physical node/edge attributes, while lagged runoff features and an optional temporal Conv1D head are used to learn delayed runoff-response behavior. Daily streamflow predictions are extracted at graph nodes corresponding to mapped USGS gauge locations and evaluated against observed streamflow and RAPID routing simulations.

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

2.1 Study Area

The Salt–Verde watershed located in central Arizona, USA (**Figure 2**) is formed by the Salt and Verde River systems that originate in the mountainous regions of northern and eastern Arizona and flow southwest toward the Phoenix metropolitan area. The watershed spans strong topographic and hydroclimatic gradients, with elevations ranging from approximately 300 to 3,450 m above sea level. High-elevation headwater regions are characterized by forested and snow-influenced environments, whereas lower elevations transition into semi-arid and desert landscapes. The watershed contains complex tributary structures and heterogeneous drainage characteristics, providing a challenging setting for streamflow routing applications.

The Salt–Verde watershed plays a critical role in central Arizona water supply and reservoir operations. Streamflow generation is influenced by both winter snow accumulation and summer monsoon precipitation, resulting in highly variable seasonal runoff dynamics. Snowmelt-driven flows from higher elevations contribute substantially to spring runoff, while short-duration convective storms associated with the North American Monsoon can generate rapid hydrologic responses and localized flooding in lower-elevation areas. These mixed runoff-generation mechanisms produce heterogeneous spatial and temporal flow behavior across the watershed. Strong hydroclimatic variability, heterogeneous terrain, and complex river topology make the Salt–Verde watershed a scientifically relevant and challenging testbed for graph-based streamflow routing models.

A total of 26 United States Geological Survey (USGS) streamflow gauges at which 30+ years of observed daily discharge records are available were selected for analysis. These gauges span a wide range of drainage areas, from 36.4 to 6,615 mi^2, with a median drainage area of 499.5 mi^2. Of those, to reduce anthropogenic regulation effects, three gauges that are substantially affected by major reservoir operations were excluded from the primary evaluation. The remaining 23 gauges represent predominantly natural flow conditions and provide broad spatial coverage across the watershed.

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

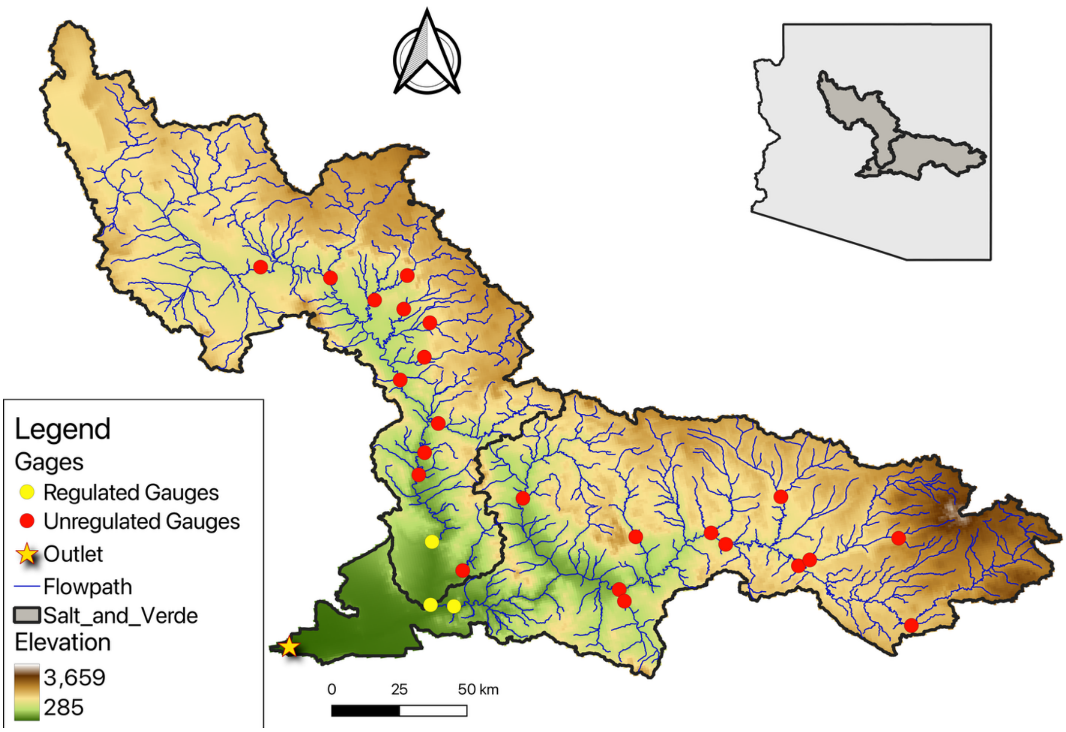


Figure 2. Study area of the Salt–Verde watershed in central Arizona, USA. The figure shows watershed boundaries, elevation distribution, NGen river-network flow paths (Strahler order > 2), and locations of USGS streamflow gauges used in this study. Red markers indicate unregulated gauges used for model training and evaluation, while yellow markers represent gauges affected by reservoir regulation that were excluded from the primary analysis. The inset map shows the location of the watershed within the state of Arizona.

2.2 Noah-MP Runoff Simulations

Runoff inputs for the graph-routing framework were generated using the Noah-MP land surface model (*Niu et al., 2011*). Noah-MP simulations were performed over the Salt–Verde river system driven by hourly *Analysis Of Record for Calibration* (AORC) meteorological forcing data at approximately 1-km spatial resolution for the 1981–2022 period (*Fall et al., 2023*). The AORC dataset provides gridded meteorological forcing variables including precipitation, temperature, radiation, humidity, wind speed, and surface pressure, and has been widely used for hydrologic and land-surface modeling applications.

The Noah-MP configuration used in this study incorporates an advanced soil hydrology representation that includes the mixed-form *Richards'* equation solver, representation of preferential-flow, and the *Van Genuchten* soil hydraulic parameterization (Niu et al., 2020a; Niu et al., 2024). Soil and vegetation characteristics were derived from the STATSGO soil dataset and the USGS 24-category vegetation classification dataset, while bedrock-depth information was obtained from *Shangguan et al. (2017)*. Additional Noah-MP configuration options used in this study are summarized in Table 1. Daily surface

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

runoff (RUNSF) and subsurface runoff (RUNSB) generated by Noah-MP were subsequently used as runoff inputs for both the graph-routing framework and RAPID routing simulations. Identical runoff inputs and simulation periods were used for both routing experiments to ensure a consistent comparison between the physics-based and graph-learning routing frameworks.

Table 1: Noah-MP parameterization schemes used for runoff simulations.

Process	Schemes
Dynamic vegetation	Dynamic vegetation
Canopy stomatal resistance	Ball-Berry type
Moisture factor for stomatal resistance	Plant water stress
Runoff and groundwater	TOPMODEL with groundwater
Surface layer exchange coefficient	Monin-Obukhov similarity theory (MOST)
Radiation transfer	Modified two-stream
Ground snow surface albedo	Two-stream radiation scheme (Wang et al., 2022)
Precipitation partitioning	Wet bulb temperature (Wang et al., 2019)
Lower boundary condition for soil temperature	2-m air temperature climatology at 8m
Snow/soil temperature time scheme	Semi-implicit
Surface evaporation resistance	Sakaguchi and Zeng (2009)
Root profile	Dynamic root (Niu et al., 2020b)
Soil hydrology	Mixed-form Richards equation (Niu et al., 2024)
Soil hydraulics	Van Genuchten parameterization (Niu et al., 2024)
Preferential flow	Dual-permeability representation (Niu et al., 2024)

2.3 RAPID Routing Benchmark

We used a physics-based routing benchmark, namely the *Routing Application for Parallel computation of Discharge* (RAPID; *David et al. (2011)*) method. RAPID is a vector-based river routing model that applies a matrix formulation of the *Muskingum* routing method to efficiently simulate streamflow propagation across large river networks. The model has been widely used in large-scale hydrologic applications and operational streamflow forecasting systems because of its computational efficiency and ability to preserve flow continuity across connected river reaches (*Agnihotri et al., 2023a*).

Daily Noah-MP surface runoff and subsurface runoff generated from the land-surface simulations were used as lateral inflow inputs to the RAPID routing model. A vector-based river network for the *Salt–Verde* watershed was developed using the NHDPlus dataset to maintain consistency between routing experiments and graph-network construction (*Tavakoly, 2017*). RAPID simulations were subsequently used as a benchmark for evaluating graph-based streamflow routing performance and routing realism across the river network.

2.4 River-Network Graph Construction

The graph-routing framework was constructed using the NGen hydrofabric representation of the *Salt–Verde* watershed (*Ogden et al., 2026*). The NGen hydrofabric represents the river system as a connected set of hydrologically linked flow paths, nexus elements, and catchment divides that explicitly preserve upstream–downstream network topology across the watershed. In this study, the hydrofabric was converted into a directed graph structure in which graph nodes represent routing elements associated with river reaches, and graph edges represent directed hydrologic connectivity between upstream and

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

downstream routing elements (**Figure 3**). The resulting graph preserves the spatial organization and topological structure of the physical river system while enabling graph-based message passing for streamflow routing.

The graph was constructed from three outlet systems corresponding to the major downstream outlets within the Salt–Verde watershed: USGS gauges 09511300, 09510000, and 09510200. These outlet systems were merged into a single routing graph containing three disconnected graph components. Directed graph connectivity was represented using upstream-to-downstream edge pairs defining flow propagation through the river network. Gauge locations were mapped onto graph nodes using projected-coordinate matching between USGS gauge locations and NGen routing elements.

The graph cache additionally stores node attributes, edge attributes, and runoff-transfer mapping arrays used during graph-routing simulations. These datasets include static physiographic descriptors, hydraulic edge characteristics, and source-to-target runoff-transfer relationships linking gridded Noah-MP runoff fields to graph-routing elements. The final graph contained 4,118 graph nodes and 4,115 directed edges representing the Salt–Verde routing system used throughout all routing experiments.

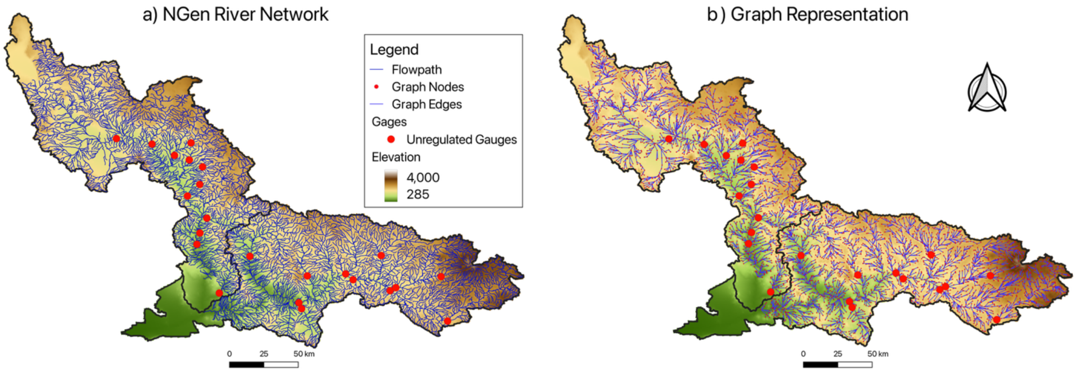


Figure 3. Representation of the Salt–Verde watershed river system and its graph-based routing structure. Panel (a) shows the original NGen hydrofabric flow path network and locations of USGS streamflow gauges used in this study. Panel (b) shows the corresponding directed graph representation used by the routing framework, where hydrologically connected routing elements are represented as graph nodes and directed edges defining upstream–downstream flow connectivity. The graph preserves the spatial organization and topology of the physical river network while enabling graph-based runoff routing and message passing.

2.5 Grid-to-Graph Runoff Transfer

Runoff generated by Noah-MP is produced on a regular gridded domain, whereas the routing framework operates on an irregular river-network graph derived from the NGen hydrofabric. Consequently, a grid-to-graph runoff-transfer procedure is required to transform gridded runoff fields into graph-node runoff inputs prior to routing (**Figure 4**). In this study, runoff transfer was performed using the NGen divide structure, where runoff generated within each divide was aggregated and assigned to the corresponding routing element within the directed graph network.

The runoff-transfer framework uses precomputed source-to-target mapping arrays stored within the graph cache, including $runoff f_{s}ource_{f}lat_{i}ndex$, $runoff f_{t}arget_{i}ndex$, and $runoff f_{s}ource_{w}eight$. These arrays define the relationship between active Noah-MP runoff grid cells and NGen hydrologic divides

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

associated with graph nodes. For each divide, runoff contributions from intersecting runoff grid cells were aggregated using predefined spatial weights, defined by area, to produce node-level runoff inputs consistent with the routing topology. For runoff grid cells located along divide boundaries, runoff contributions were assigned to the divide containing the largest fractional coverage of the corresponding grid cell.

Daily Noah-MP surface runoff (RUNSF) and subsurface runoff (RUNSB) were used as dynamic runoff inputs for the routing framework. For each simulation day, runoff generated on the regular Noah-MP grid was transferred into divide-based graph-node runoff signals prior to graph routing. The resulting runoff tensor therefore represents spatially aggregated runoff inputs aligned with the directed river-network graph used by the graph neural network routing framework.

The divide-based runoff-transfer method provides an efficient mechanism for coupling physically generated gridded runoff fields with graph-based river routing while preserving consistency between hydrologic catchment structure and river-network connectivity. Unlike many graph-based hydrologic studies that directly generate runoff at graph nodes or subbasin units, the proposed framework explicitly bridges gridded land-surface model runoff with topology-aware river-network routing.

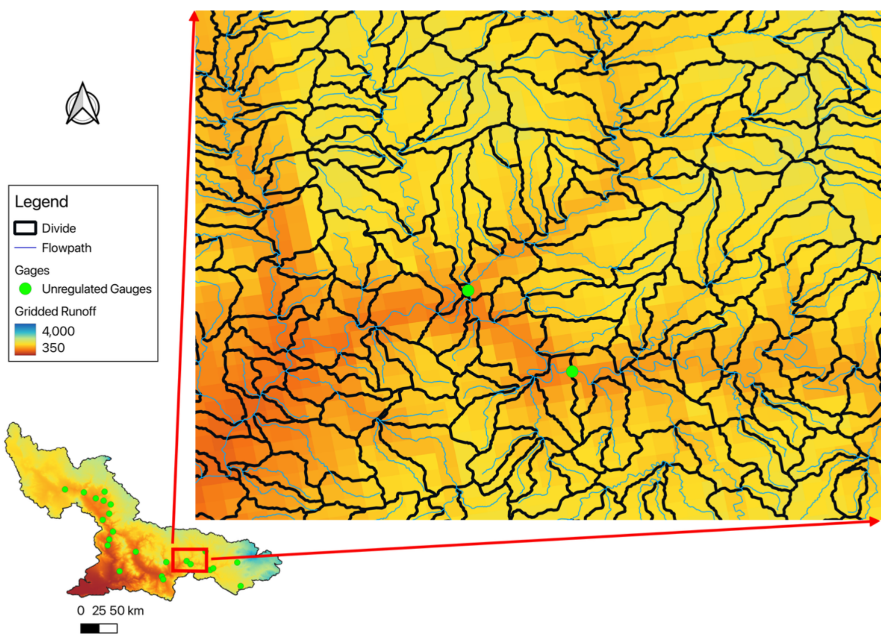


Figure 4 Illustration of the grid-to-graph runoff-transfer procedure used in the routing framework. Black polygons represent NGen hydrologic divides, blue lines indicate river-network flow paths, and the colored raster background represents Noah-MP grids. Runoff generated within each divide is spatially aggregated and transferred to the corresponding graph-routing element prior to graph-based routing. Green markers indicate USGS streamflow gauges used for model training and evaluation.

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

2.6 Graph Features and Topology Attributes

The graph-routing framework incorporates both node and edge attributes derived from the NGen hydrofabric and associated physiographic datasets to provide physically meaningful information for graph-based routing. Node features describe the local hydrologic, geomorphic, and topologic properties of individual routing elements within the river network. These attributes included river-segment length, local and upstream drainage area, stream order, terrain elevation, terrain slope, impervious fraction, dominant soil type, dominant vegetation type, distance to outlet, upstream contributing-area metrics, and travel-time descriptors representing relative routing behavior throughout the watershed.

Several node attributes were specifically designed to provide information related to hydrologic connectivity and flow-transport behavior within the river system. For example, $tot_{d}rainage_{a}reasqkm$, $upstream_{a}rea_{r}atio$, and $upstream_{f}lowpath_{c}ount$ help distinguish headwater routing elements from downstream integrating reaches, while $distance_{t}o_{o}utlet_{m}$, $local_{t}ravel_{t}ime_{p}roxy$, and $travel_{t}ime_{p}roxy_{t}o_{o}utlet$ provide relative measures of routing distance and flow-travel behavior across the network. Additional features describing soil and vegetation properties allow the routing framework to incorporate spatial physiographic variability throughout the watershed.

Edge attributes describe hydraulic and geomorphic relationships between connected routing elements and provide physically meaningful information governing runoff propagation between upstream and downstream graph nodes. Edge features included channel length, slope, Manning-type roughness parameters, channel-width descriptors, Muskingum-style hydraulic proxies, travel-time proxies, storage-related metrics, and conveyance descriptors. These attributes allow graph message passing to depend not only on graph topology but also on hydraulic and geomorphic characteristics influencing runoff transport through the river network.

The combination of directed graph topology, static node attributes, and hydraulic edge descriptors enables the graph-routing framework to represent heterogeneous routing behavior across the watershed. Consequently, graph message passing is informed not only by upstream–downstream connectivity but also by local *terrain*, *drainage structure*, *hydraulic characteristics*, and *network position* within the river system. Table 2 summarizes the primary node and edge features used in the graph-routing framework.

Table 2. Graph topology, runoff-transfer, and physical feature variables stored in the graph cache.

Category	Variables	Description
Graph topology	edge_source, edge_target	Directed upstream-to-downstream graph connectivity
Node location / indexing	flat_index, node_x, node_y	Node location on the runoff/forcing grid
Gauge mapping	gauge_id, gauge_index	USGS gauge identifiers and corresponding graph-node indices
Runoff transfer	runoff_source_flat_index, runoff_target_index, runoff_source_weight, runoff_source_fraction	Mapping between Noah-MP runoff grid cells and NGen divide/graph nodes

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

Runoff-source attributes	elevation, cell_area_m2, distance_to_flowpath_m	Attributes of runoff-contributing grid cells
Node features	node_dem_elevation_m, lengthkm, areasqkm, tot_drainage_areasqkm, order, mean.elevation, mean.slope, mean.impervious, mode.ISLTYP, mode.IVGTYP, local_length_m, distance_to_outlet_m, local_travel_time_proxy, travel_time_proxy_to_outlet, incremental_area_fraction, upstream_area_ratio, upstream_flowpath_count, upstream_area_km2_topologic, normalized_stream_order, is_outlet_node	Static physiographic, topologic, and routing-position descriptors for each graph node
Edge features	Length_m, So, n, nCC, BtmWdth, TopWdth, TopWdthCC, MusX, MusK, travel_time_proxy, velocity_proxy, manning_travel_time_proxy, inverse_manning_travel_time_proxy, storage_proxy, conveyance_proxy, width_ratio_proxy	Hydraulic, geomorphic, and routing-proxy attributes for each directed edge

2.7 Graph Neural Network Routing Framework

The routing model uses a *Graph Attention Network* (GAT) to perform topology-aware message passing over the river graph. During message passing, connected upstream routing elements do not contribute equally. Instead, the attention mechanism learns adaptive weights for graph edges based on node states and edge attributes such as flow path length, slope, hydraulic roughness/channel descriptors, travel-time proxies, upstream drainage attributes, and other derived hydrologic features. This allows the model to learn spatially variable routing influence across headwaters, steep channels, low-gradient reaches, and downstream integrating segments.

For each graph node i, the GAT computes attention coefficients describing the relative influence of neighboring upstream nodes j during graph message passing

$$\alpha_{ij} = \frac{\exp\left(\mathrm{LeakyReLU}\left(a^{T}[Wh_i \| Wh_j]\right)\right)}{\sum_{k\in N(i)} \exp\left(\mathrm{LeakyReLU}\left(a^{T}[Wh_i \| Wh_k]\right)\right)} \qquad \text{(Eq 1)}$$

where h_i and h_j represent node-feature vectors for connected routing elements, W is a learnable linear transformation, a is the attention vector, and $\mathcal{N}(i)$ denotes neighboring upstream nodes connected to node i. The resulting attention coefficient α_{ij} represents the learned routing influence of upstream node j on downstream node i.

The updated node representation is then computed through weighted aggregation of neighboring upstream information:

$$h_i' = \sigma\left(\sum_{j\in\mathcal{N}(i)} \alpha_{ij} W h_j\right) \qquad \text{(Eq 2)}$$

where σ denotes a nonlinear activation function. Through this message-passing procedure, runoff and hydrologic information propagate through the directed river network while allowing the model to learn spatially varying routing behavior across heterogeneous river segments.

Temporal behavior is represented in two complementary ways. First, for each graph node and each prediction day, the model receives lagged runoff inputs from previous days. For example, a configuration may use runoff lags such as 0, 1, 2, 3, 7, and 14 days, meaning that the node feature vector for a given day includes runoff from the current day and selected previous days. The dataset provides a routing-lag context period before each prediction window, so these lagged values come from real runoff data rather than zero padding. Second, when enabled, a temporal Conv1D head is applied after graph message passing to learn short-term temporal smoothing and delayed response patterns across the daily output sequence. Therefore, temporal routing memory is learned from lagged runoff features and temporal convolution, rather than imposed as an explicit mass-conserving hydrologic routing equation

During training, the graph topology, gauge-node mapping, and runoff-transfer mapping remain fixed. The learned parameters are the neural-network weights controlling node-feature transformation, graph attention/message passing, and optional temporal convolution. Streamflow predictions are extracted only at graph nodes mapped to USGS gauge locations using the predefined gauge-index arrays. The resulting predictions are daily routed streamflow estimates informed by Noah-MP runoff, constrained by the NGen river-network structure, and calibrated against observed gauge streamflow.

2.8 Training Strategy and Loss Functions

The graph-routing framework was trained using supervised learning with observed daily USGS streamflow as the prediction target. Training samples were constructed from daily Noah-MP runoff and gauge streamflow using fixed-length temporal windows. In the primary experiments, each sample contained a 180-day streamflow prediction window. An additional routing-lag context period was included before the prediction window so that lagged runoff inputs at the beginning of the sequence were derived from real antecedent runoff values rather than artificial zero padding. The context period was used only to construct lagged runoff features; training losses and reported metrics were computed over the prediction window. In addition to the standard routing configuration, the study evaluated multiple temporal-routing-memory configurations based on different runoff-lag representations. To investigate the influence of antecedent runoff information on graph-routing performance, separate experiments were conducted using lag windows of 3, 7, 14, and 30 days in order. To isolate the contribution of learned temporal sequence processing relative to graph-based spatial routing alone, additional ablation experiments were performed without the temporal convolutional (CNN) head. These experiments allowed systematic evaluation of how temporal-memory representation influences streamflow-routing behavior across the Salt–Verde river network.

Training, validation, and testing datasets were constructed using distribution-balanced temporal block splitting to reduce distributional mismatch between data partitions and provide a robust evaluation of model generalization under heterogeneous hydroclimatic conditions (Figure S1). The continuous streamflow record (1981–2022) was first divided into non-overlapping 180-day temporal blocks, and those blocks were assigned to training, validation, and testing partitions so that observed streamflow distributions were as similar as possible across the splits. This differs from purely chronological splitting and from random day-level sampling. This strategy was adopted because the Salt–Verde watershed exhibits strong interannual hydroclimatic variability, seasonal runoff differences, vegetation and watershed-condition changes over recent decades, and nonstationary streamflow behavior. Consequently, purely random day-level sampling could artificially mix hydrologically distinct periods and introduce unrealistic distribution leakage between training and evaluation datasets. By preserving contiguous temporal blocks while balancing streamflow distributions across the data partitions, the proposed splitting strategy reduces this risk and provides a more robust evaluation of model generalization under heterogeneous hydroclimatic conditions. The resulting training, validation, and testing partitions are summarized in **Figure S1** of the Supplementary Material, which shows the cumulative distribution functions (CDFs) of observed streamflow for each partition.

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

Model training used masked loss functions so that missing streamflow observations did not contribute to optimization. Losses were computed only at graph nodes corresponding to mapped USGS gauge locations. Several loss formulations were evaluated during model development, including *Mean Squared Error* (MSE), *weighted MSE*, and *Jawad Kling–Gupta Efficiency* (JKGE)-based losses (*Jawad et al, 2026*).

For the standard masked MSE loss, missing streamflow observations were excluded using a validity mask. For gauge g, the loss was computed as:

$$\mathrm{MSE_g} = \frac{\sum_{t=1}^{T} m_{g,t}\left(Q_{g,t}^{pred} - Q_{g,t}^{obs}\right)^2}{\sum_{t=1}^{T} m_{g,t}} \qquad \text{(Eq 3)}$$

where $Q_{g,t}^{pred}$ and $Q_{g,t}^{obs}$ denote predicted and observed streamflow at gauge g and time step t, respectively, and $m_{g,t}$ a binary mask equal to 1 for valid observations and 0 for missing values. The final MSE loss was computed by averaging $\mathrm{MSE_g}$ over valid gauges in the batch.

The weighted MSE formulation extends this by including optional gauge or time-dependent weights:

$$\mathrm{WMSE}_g = \frac{\sum_{t=1}^{T} w_{g,t}\left(Q_{g,t}^{pred} - Q_{g,t}^{obs}\right)^2}{\sum_{t=1}^{T} w_{g,t}} \qquad \text{(Eq 4)}$$

where $w_{g,t} = m_{g,t}\, a_g$ combines the missing-data mask and any optional gauge-specific weighting (a_g). Gauges without a specified curriculum weight were assigned $a_g = 1$, while selected gauges received larger weights between 1.25 and 1.5 to increase their contribution during optimization. This formulation prevents gauges with more valid observations from dominating the optimization and allows selected gauges to receive higher or lower training emphasis.

The framework also evaluated JKGE-based losses. For observed discharge y_g and predicted discharge q_g at gauge g, a benchmark operator $B(.)$ defines the time-varying baselines $b_{y,g} = B(y_g)$ and $b_{q,g} = B(q_g)$, from which the anomaly components $\psi_{y,g} = y_g - b_{y,g}$ and $\psi_{q,g} = q_g - b_{q,g}$ are computed. Next the anomaly correlation, anomaly variability and normalized benchmark mismatch components C_g, V_g and M_g are determined as:

$$C_g = \frac{\psi_{q,g}^{T}\psi_{y,g}}{\left\|\psi_{q,g}\right\| \cdot \left\|\psi_{y,g}\right\|}$$

$$V_g = \frac{\left\|\psi_{q,g}\right\|}{\left\|\psi_{y,g}\right\|}$$

$$M_g = \frac{\left\|b_{q,g} - b_{y,g}\right\|}{\left\|b_{y,g} - \overline{y_g}1\right\|}$$

where $\overline{y_g}$ is the mean observed discharge, and 1 is a vector of ones. From these, the JKGE score for gauge g is computed as:

$$JKGE_g = 1 - \sqrt{\left(C_g - 1\right)^2 + \left(V_g - 1\right)^2 + M_g^2} \qquad \text{(Eq 5)}$$

The JKGE loss was then defined as: $L_{JKGE} = 1 - \overline{JKGE}$, where $JKGE$ is the average JKGE over valid gauges. In the section-average JKGE variants, the benchmark operator $B(.)$ was implemented using fixed non-overlapping temporal sections, so that each section had its own local benchmark mean. This allowed

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

the loss to evaluate both low-frequency streamflow behavior and anomaly dynamics within each temporal section.

The routing framework also supports curriculum-style gauge training. The model always operates on the complete fixed river-network graph, but the loss was computed only for gauges active within the current curriculum stage. The curriculum progressively increased routing complexity by beginning with upstream and headwater gauges and gradually incorporating more downstream and hydrologically integrated gauges. This staged strategy was intended to stabilize optimization by first learning simpler local routing behavior before requiring the model to reproduce more complex downstream routing responses.

Model optimization was performed using backpropagation on the fixed directed river-network graph. During training, graph topology, runoff-transfer mappings, gauge-node mappings, and graph feature definitions remained fixed throughout all experiments. The learned parameters consisted of neural-network weights controlling node-feature transformations, graph message passing and attention weighting, and optional temporal convolution operations. Stage-best checkpointing and validation-based model selection were used to preserve stable routing solutions and reduce destructive transitions between curriculum-training stages.

2.9 Evaluation Metrics

Model performance was evaluated using multiple hydrologic metrics designed to assess streamflow *magnitude*, *temporal dynamics*, *variability*, and *routing consistency* across the *Salt–Verde* river system. Performance evaluation focused not only on prediction accuracy at individual gauges, but also on the ability of the routing framework to reproduce physically meaningful upstream–downstream flow behavior throughout the directed river network. Routing performance was evaluated independently at each USGS gauge using observed daily streamflow records and corresponding routed streamflow predictions.

The primary evaluation metrics included the *Kling–Gupta Efficiency* (KGE). The KGE metric is defined as:

$$\mathrm{KGE} = 1 - \sqrt{\{(1-\alpha)^2 + (1-\beta)^2 + (1-\rho)^2\}} \qquad \text{(Eq 6)}$$

where ρ is the Pearson correlation coefficient between observed and predicted streamflow, α is the variability ratio, and β is the bias ratio.

In this study, we report the KGE skill score ($\mathrm{KGE}_{\{ss\}}$), which provides a normalized version of KGE relative to a benchmark model, Following the formulation discussed by *Knoben et al. (2019)*:

$$\mathrm{KGE}_{\{ss\}} = 1 - \frac{1-\mathrm{KGE}}{\sqrt{2}} \qquad \text{(Eq 7)}$$

Using this formulation ensures that $\mathrm{KG}E_{\{ss\}} \leq 0$ means that the model performs no better than a trivial benchmark model that corresponds to simulated values at every time step being set to the observed long-term-mean, while a perfect model yields $\mathrm{KG}E_{\{ss\}} = 1$. All reported evaluation metrics are computed over the independent test period. Additional diagnostics including *Root-Mean-Square Error* (RMSE), and *Percent Bias* (PBIAS) were also computed for each gauge. To quantify improvement relative to the RAPID benchmark, we additionally computed the $\mathrm{KGE}_{\{ss\}}$ improvement as

$$\Delta\mathrm{KGE}_{ss} = \mathrm{KGE}_{ss}^{model} - \mathrm{KGE}_{ss}^{RAPID} \qquad \text{(Eq 8)}$$

where positive values indicate that the evaluated model outperforms RAPID. Percent improvement relative to RAPID was computed

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

$$\%\Delta \text{KGE}_{ss} = 100 \times \frac{\text{KGE}_{ss}^{model} - \text{KGE}_{ss}^{RAPID}}{|\text{KGE}_{ss}^{RAPID}|} \quad \text{(Eq 8)}$$

Because KGE_{ss}^{RAPID} can be zero or negative at some gauges, percent improvements were interpreted cautiously and were not emphasized when the RAPID denominator was near zero.

Since a primary objective of this study was to evaluate graph-based routing behavior relative to a widely used physics-based routing framework, all graph-routing simulations were directly compared against RAPID routing simulations generated using the same Noah-MP runoff forcing inputs. Because both routing systems used identical runoff inputs, differences between routed streamflow predictions primarily reflect differences in routing representation rather than runoff-generation uncertainty. This configuration allowed the study to isolate graph-based routing behavior and evaluate whether topology-aware graph learning can reproduce or improve upon physically based routing behavior.

3 Results

3.1 Overall Routing Performance

Results show that the graph-based routing framework demonstrates strong streamflow-routing skill across the *Salt–Verde* watershed by consistently outperforming the RAPID routing benchmark at most of the evaluation gauge locations. Performance comparisons across routing configurations show that graph-based routing substantially improves streamflow representation relative to RAPID while preserving physically meaningful upstream–downstream routing structure. **Figure 5** summarizes $\text{KGE}_{\{ss\}}$ improvements relative to RAPID for all routing configurations, including different temporal-routing-memory experiments and loss-function formulations.

Among all evaluated experiments, the best-performing configuration corresponds to the 14-day lag weighted-MSE experiment ($lag14_w mse$), which achieved the highest mean $\text{KGE}_{\{ss\}}$ value (0.70) and the highest median $\text{KGE}_{\{ss\}}$ value (0.72) across the evaluation gauges (**Table 3**). Several additional weighted-MSE configurations also ranked among the strongest-performing experiments, indicating that graph-based routing performance remained robust across multiple training configurations. Although the 14-day lag produced the best results for the Salt–Verde watershed, the optimal temporal-memory representation is likely basin dependent and influenced by factors such as watershed size, flow-regime characteristics, snowmelt processes, and subsurface storage dynamics. Consequently, temporal-memory sensitivity analyses similar to those conducted here may be necessary when applying the framework to other river systems.

JKGE-based training configurations produced more variable routing behavior across the watershed. Although several JKGE experiments achieved competitive routing skill, the standard $lag14_j kge$ configuration showed lower overall performance than the *weighted-MSE* and *standard-MSE* experiments. Nevertheless, even the weaker JKGE configurations still produced predominantly positive $\text{KGE}_{\{ss\}}$ improvements relative to RAPID, indicating that the graph-routing framework remained competitive across nearly all gauges regardless of loss formulation.

Figure 5 demonstrates that nearly all routing configurations produce positive $\text{KGE}_{\{ss\}}$ improvements relative to RAPID across essentially the entire gauge network. For most experiments, all gauges show positive improvement over RAPID. Even the $lag14_j kge$ configuration, which exhibits the weakest overall performance among the evaluated experiments, produces only a single gauge with slight negative $\text{KGE}_{\{ss\}}$ change relative to RAPID. The consistent positive improvement distributions indicate that the graph-

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

routing framework robustly improves routing behavior across a wide range of hydrologic conditions, drainage areas, and river-network structures within the Salt–Verde watershed.

For the best-performing configuration (lag14_wmse), the mean $KGE_{\{ss\}}$ increased from −0.29 for RAPID to 0.70 for RiverGraphNet, and all 22 gauges with valid RAPID comparisons exhibited positive improvement. A one-sided paired Wilcoxon signed-rank test comparing gauge-level $KGE_{\{ss\}}$ values between RiverGraphNet and RAPID indicates that the improvement was highly significant ($p = 2.38 \times 10^{-7}$), providing statistical support that the observed performance gains are unlikely to arise from random variation across gauges.

Additional examination of the gauge-level improvement distributions using violin plots (**Figure S2**) reveals patterns consistent with the boxplot results shown in **Figure 5**. The distributions remained predominantly positive across routing configurations and does not exhibit substantial multimodal structure, indicating that the observed performance gains were broadly distributed throughout the gauge network rather than being driven by a small number of highly improved locations.

To provide additional context beyond the RAPID benchmark, RiverGraphNet performance was also compared against the National Water Model (NWM) streamflow simulations (**Figure 6**). RAPID was selected as the primary benchmark throughout this study because it generally achieved higher $KGE_{\{ss\}}$ values and more stable performance across the evaluation network than NWM. Consequently, RAPID represents the more challenging benchmark for assessing routing performance within the Salt–Verde watershed. RiverGraphNet nevertheless outperforms both RAPID and NWM across most evaluation gauges, indicating that the observed performance gains are not simply relative to a weak benchmark but extend to a widely used operational hydrologic modeling framework. This comparison provides additional confidence that the improvements reflect meaningful advances in routing representation rather than advantages relative to a particular benchmark model.

The overall results suggest that graph-based routing provides a flexible, robust, and physically meaningful alternative to traditional routing approaches when physically-generated runoff inputs and physically-structured river-network topology are incorporated into the learning framework. The strongest routing performance emerges from configurations that combined graph-based spatial message passing, antecedent runoff memory, and balanced gauge-aware optimization within the directed routing graph framework.

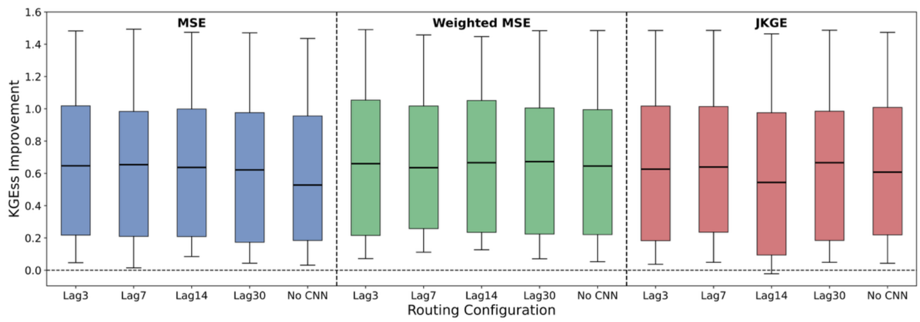
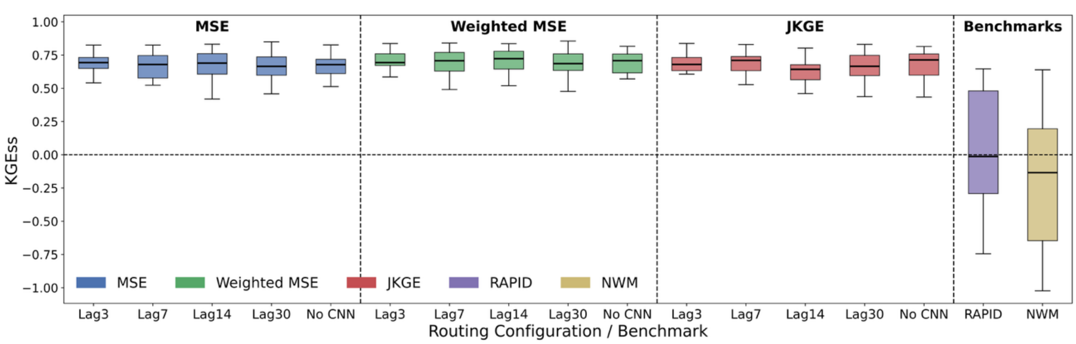


Figure 5. Distribution of $KGE_{\{ss\}}$ improvements relative to the RAPID routing benchmark across all graph-routing configurations evaluated in this study. Boxplots summarize gauge-level KGEss improvement distributions for different temporal-routing-memory configurations (3-, 7-, 14-, and 30-day runoff lags), loss formulations (MSE, weighted MSE, and JKGE), and experiments without the temporal convolutional (CNN) head.

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

Table 3. Summary statistics of gauge-level $KGE_{\{ss\}}$ values for RAPID and all graph-routing configurations across the Salt–Verde watershed. Reported statistics include the mean, median, minimum, maximum, and standard deviation of KGESS computed across all evaluation gauges. Routing configurations include different temporal-routing-memory experiments, loss-function formulations, and ablation experiments without the temporal convolutional (CNN) head. The weighted-MSE configurations generally produced the highest overall routing performance and the most stable gauge-level behavior.

scenario	mean	median	minimum	maximum	standard deviation
RAPID	-0.29	-0.01	-4.88	0.65	1.30
lag14_wmse	0.70	**0.72**	0.34	0.84	0.11
lag14_no_cnn_jkge	0.67	**0.71**	0.30	0.81	0.12
lag7_jkge	**0.68**	**0.71**	0.32	0.83	0.11
lag14_no_cnn_wmse	**0.68**	**0.71**	0.33	0.82	0.12
lag7_wmse	**0.69**	**0.71**	**0.38**	0.84	0.11
lag3_wmse	**0.69**	0.69	0.30	0.84	0.12
lag3_mse	**0.68**	0.69	0.29	0.82	0.11
lag14_mse	0.67	0.69	0.32	0.83	0.13
lag30_wmse	**0.68**	0.69	0.31	**0.86**	0.13
lag3_jkge	0.67	0.68	0.31	0.84	0.12
lag7_mse	0.66	0.68	0.30	0.82	0.12
lag14_no_cnn_mse	0.65	0.68	0.30	0.83	0.12
lag30_jkge	0.65	0.67	0.32	0.83	0.12
lag30_mse	0.66	0.67	0.33	**0.85**	0.11
lag14_jkge	0.61	0.64	0.27	0.80	0.13



https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

Figure 6. Distribution of $\mathrm{KGE}_{\{ss\}}$ across evaluation gauges for all RiverGraphNet routing configurations compared with RAPID and National Water Model (NWM) benchmarks. RiverGraphNet configurations consistently achieve higher median performance and smaller performance variability than both benchmark systems. The comparison demonstrates that performance gains are not limited to improvements over RAPID but also extend to a widely used operational hydrologic modeling framework.

3.2 Spatial Routing Performance Across the Watershed

Spatial evaluation of routing performance reveals substantial variability in streamflow-routing skill across the *Salt–Verde* watershed for both the RAPID and graph-based routing frameworks. Figure 7 compares gauge-level $\mathrm{KGE}_{\{ss\}}$ values for the RAPID benchmark and the best-performing graph-routing configuration ($lag14_w mse$). The spatial distributions demonstrate that routing performance is strongly influenced by *watershed position*, *drainage structure*, and *local hydrologic characteristics*.

The RAPID routing benchmark produces highly variable routing performance across the watershed, with several gauges exhibiting relatively low or even negative $\mathrm{KGE}_{\{ss\}}$ values. The weakest RAPID performance generally occurs in smaller headwater tributaries and intermediate nested basins characterized by steeper terrain and more heterogeneous runoff-generation behavior. These regions often exhibit highly nonlinear hydrologic responses, short runoff concentration times, and stronger sensitivity to localized precipitation variability, which may be difficult to fully represent using fixed-parameter routing approaches.

In contrast, the graph-based routing framework produced consistently positive and spatially stable routing performance across nearly all evaluation gauges. The best-performing graph-routing configuration substantially reduced the occurrence of low-skill gauges and produced improved routing behavior throughout both upstream tributaries and downstream integrating river reaches. The graph-routing framework achieved particularly strong performance in mountainous and hydrologically complex regions where upstream connectivity, local drainage structure, and variable flow-transport behavior play important roles in streamflow propagation.

Figure 8 shows the spatial distribution of $\mathrm{KGE}_{\{ss\}}$ improvements relative to RAPID. Gains are not uniform, with the strongest benefits being concentrated in *headwater* and *intermediate* tributaries. These improvements suggest that topology-aware graph message passing and adaptive attention weighting allows the routing framework to better represent heterogeneous runoff propagation and nonlinear upstream–downstream interactions compared to the fixed routing representation used in RAPID.

The spatial improvement patterns additionally indicate that graph-based routing is able to effectively integrate physically meaningful node and edge attributes describing river geometry, drainage structure, and hydraulic-routing proxies. Because the framework explicitly represents river-network connectivity through directed graph topology, upstream runoff contributions can influence downstream routing behavior through learned spatial interactions rather than fixed routing equations alone. This flexibility likely contributes to the improved routing behavior observed across complex tributary systems and nested watershed structures within the Salt–Verde river network.

Although graph-based routing generally outperforms RAPID throughout the watershed, the magnitude of improvement varies spatially across gauges. Smaller improvements at several downstream gauges may reflect the increased influence of integrated watershed-scale dynamics, routing accumulation uncertainty, and potential anthropogenic flow regulation effects that are not fully represented within the current graph-routing framework. Nevertheless, the consistently positive improvement patterns demonstrate

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

that graph-based routing provides a robust and spatially transferable alternative for routing physically generated runoff through complex river-network systems.

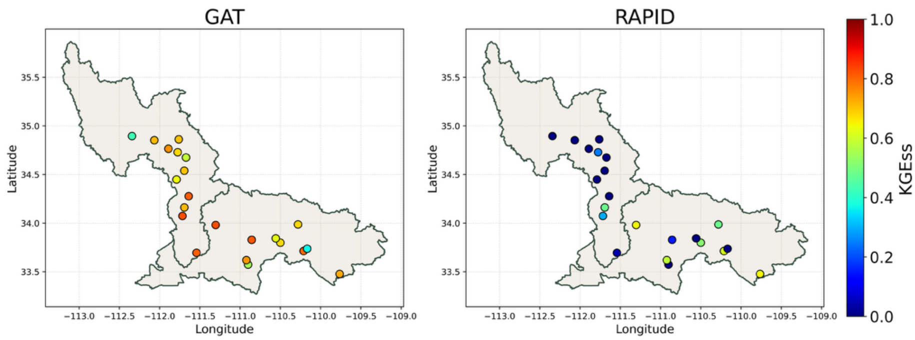
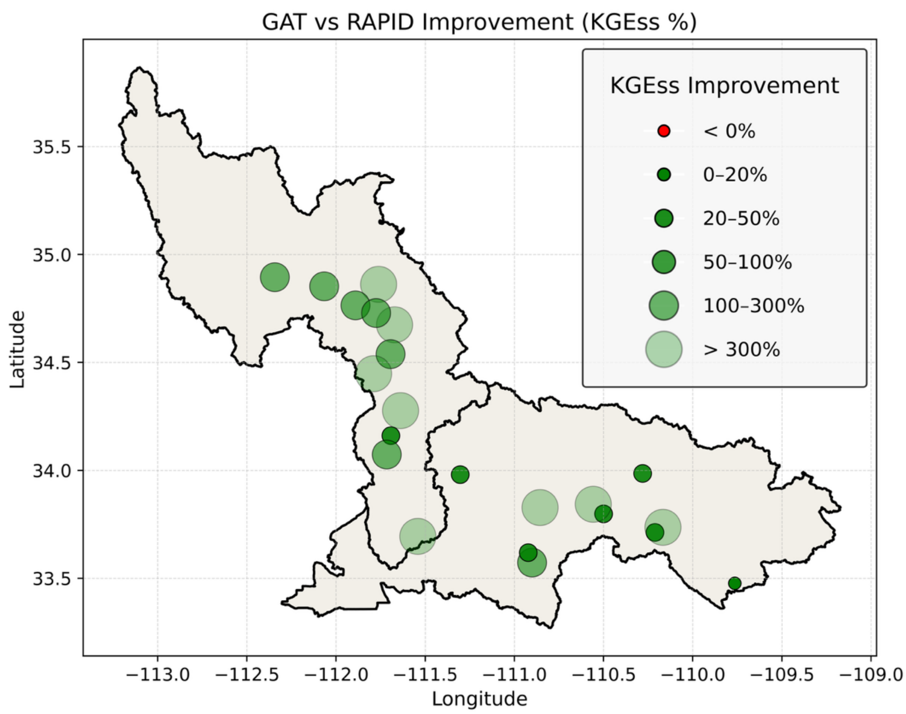


Figure 7. Spatial distribution of gauge-level KGESS values across the Salt–Verde watershed for (a) the RAPID routing benchmark and (b) the best-performing graph-routing configuration (lag14_wmse). Circle colors represent KGESS values at each USGS evaluation gauge. The graph-based routing framework produced more spatially consistent and uniformly positive routing performance across the watershed relative to RAPID, particularly within upstream and intermediate tributary systems.



https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

Figure 8. Spatial distribution of KGEss improvement relative to RAPID for the best-performing graph-routing configuration (lag14_wmse). Positive values indicate improved routing performance of the graph-based framework compared to the RAPID routing benchmark. Circle size and color intensity increase with larger positive improvement magnitude. The strongest improvements generally occurred within upstream and intermediate tributaries characterized by complex river-network topology and heterogeneous hydrologic behavior.

3.3 Influence of Temporal Routing Memory

The influence of temporal-routing-memory representation was evaluated using multiple lag configurations and temporal-processing experiments. To investigate how antecedent runoff information influences graph-based streamflow routing behavior, routing experiments were conducted using runoff-lag windows of 3, 7, 14, and 30 days. Additional ablation experiments were performed without the temporal convolutional (CNN) head to isolate the contribution of learned temporal sequence processing relative to graph-based spatial routing alone. **Figure 5** and **Table 3** summarize routing performance across all temporal-routing-memory configurations.

The results indicate that the routing performance is relatively insensitive to the selected runoff-lag representation over the range of lag windows evaluated. Differences among lag configurations are generally modest, with most experiments producing similar gauge-level $\mathrm{KGE}_{\{ss\}}$ distributions (**Figure 5**). Nevertheless, the 7-day and 14-day lag experiments consistently ranked among the strongest-performing configurations. In particular, the 14-day lag weighted-MSE configuration ($lag14_{w}mse$) achieves the highest mean and median $\mathrm{KGE}_{\{ss\}}$ values among all evaluated experiments.

Although differences among lag configurations are relatively small, the consistently strong performance of the intermediate-lag experiments suggests that antecedent hydrologic information over multi-day timescales may provide useful routing context within the *Salt–Verde* watershed. Streamflow generation in the basin is influenced by the travel time of snowmelt runoff, subsurface-flow persistence, and multi-day runoff accumulation from tributary systems. Lag windows spanning approximately one to two weeks therefore appear sufficient to capture much of the antecedent information relevant for daily routing prediction.

Extending the lag window from 14 to 30 days did not produce additional performance gains despite providing substantially larger antecedent runoff context. This suggests that, for the *Salt–Verde* watershed, incorporating runoff information beyond approximately two weeks provides limited additional benefit for routing performance. However, the optimal temporal-memory representation is likely basin dependent and may vary with watershed size, river-network length, flow-regime characteristics, and dominant hydrologic processes. The similarity of the results across lag configurations further indicates that the graph-routing framework remains robust across a range of temporal-memory representations.

The "no-CNN" ablation experiments demonstrate that the graph-routing framework remains highly competitive even without the temporal convolutional processing head, with several experiments achieving routing performance comparable to, or exceeding, many of the CNN-enabled configurations. These results indicate that much of the routing information is already captured through graph-based spatial message passing together with lagged runoff inputs. The results suggest that the temporal convolutional head serves mainly as an additional refinement mechanism for short-term temporal smoothing and delayed-response representation, rather than as the dominant source of routing memory within the framework.

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

Overall, the temporal-routing-memory experiments indicate that incorporating antecedent runoff information is beneficial for graph-based routing, but that routing performance remains relatively stable across the range of lag windows evaluated. The results further suggest that physically structured graph topology and lagged runoff inputs together capture much of the runoff-propagation behavior required for streamflow prediction within the *Salt–Verde* river network.

3.4 Influence of Loss Functions

The influence of objective functions on graph-routing performance was evaluated by comparing models trained using *standard MSE*, *weighted MSE*, and *JKGE-based loss* formulations. While the previous section focused on temporal-routing-memory representation, this section examines how different training objectives shape routing behavior and hydrologic prediction characteristics. **Figure 5** and **Table 3** summarize watershed-wide performance differences across routing configurations, while representative event-scale hydrograph comparisons in **Figure 9** provide insight into how the loss formulations influence specific hydrologic behaviors.

Weighted-MSE-based configurations consistently produced the strongest overall routing performance across the watershed. Multiple *weighted-MSE* experiments ranked among the top-performing configurations, with the $lag14_{w}mse$ experiment achieving the highest overall mean and median $\mathrm{KGE}_{\{ss\}}$ values (**Table 3**). This behavior suggests that gauge-balanced weighted optimization improves routing robustness across the heterogeneous gauge network by preventing the optimization from becoming dominated by a small subset of gauges with larger flows, longer records, or stronger gradients. Because the *weighted-MSE* formulation combines missing-data masking with optional gauge-specific weighting, the training process remained spatially balanced while allowing additional emphasis on strategically important gauges.

*Event-scale hydrograph comparisons (***Figure 9***) reveal clear behavioral differences among the loss formulations. Models trained using standard MSE and weighted MSE generally reproduces event timing and peak magnitudes more effectively than the JKGE-based configuration, contributing to their superior aggregate performance metrics. In contrast, the JKGE-based configuration often produced closer agreement with observed recession behavior and low-flow recovery following major runoff events. RAPID frequently exhibited delayed response and substantial overestimation during several large runoff events. This suggests that squared-error optimization tends to prioritize reducing errors during high-flow conditions, where absolute deviations are largest, while placing less emphasis on accurately representing recession dynamics and sustained low-flow behavior.*

To further quantify these differences, hydrologic-signature metrics were computed for each routing framework (**Figure S3**). Consistent with the hydrograph comparisons, the loss functions emphasized different aspects of watershed response. The weighted-MSE configuration generally produces smaller errors for mean flow (7.7%), high-flow magnitude (18.7%), and peak flow (6.1%), substantially outperforming RAPID, which exhibits errors of 52.5%, 56.2%, and 293.7%, respectively. In contrast, the JKGE configuration produces the lowest errors for low-flow magnitude (38.8% versus 77.3% for weighted MSE), recession behavior (91.9% versus 99.0%), and baseflow index (18.9% versus 27.0%). These results support the event-scale hydrograph analysis by showing that JKGE training better preserves selected low-flow and recession characteristics, whereas weighted-MSE training tends to reduce errors in mean-flow, high-flow, and peak-flow signatures. The results highlight tradeoffs among loss formulations rather than a universally superior objective function.

In contrast, *JKGE-based* training produces a different routing behavior profile. Although JKGE configurations generally achieves lower overall $\mathrm{KGE}_{\{ss\}}$ performance than the MSE-based formulations, both the hydrograph comparisons and hydrologic-signature analysis indicate that JKGE often produces

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

closer agreement with observed recession behavior, low-flow recovery, and baseflow characteristics following major events (**Figure 9**). Because the JKGE objective explicitly incorporates anomaly correlation, variability consistency, and benchmark mismatch rather than relying solely on squared-error minimization, the resulting routing behavior tends to better preserve selected hydrograph-shape characteristics and temporal variability structure. However, this broader optimization objective may also complicate training across highly heterogeneous routing conditions, contributing to weaker aggregate predictive performance.

These results indicate that loss-function choice influences not only overall routing skill but also the hydrologic behavior emphasized during optimization. *Weighted-MSE* formulations appear most effective for maximizing overall predictive skill and peak-event routing performance, whereas *JKGE-based* objectives may better preserve recession dynamics and low-flow behavior in selected situations. This tradeoff highlights the importance of aligning optimization objectives with the hydrologic behaviors most relevant to the intended routing application.

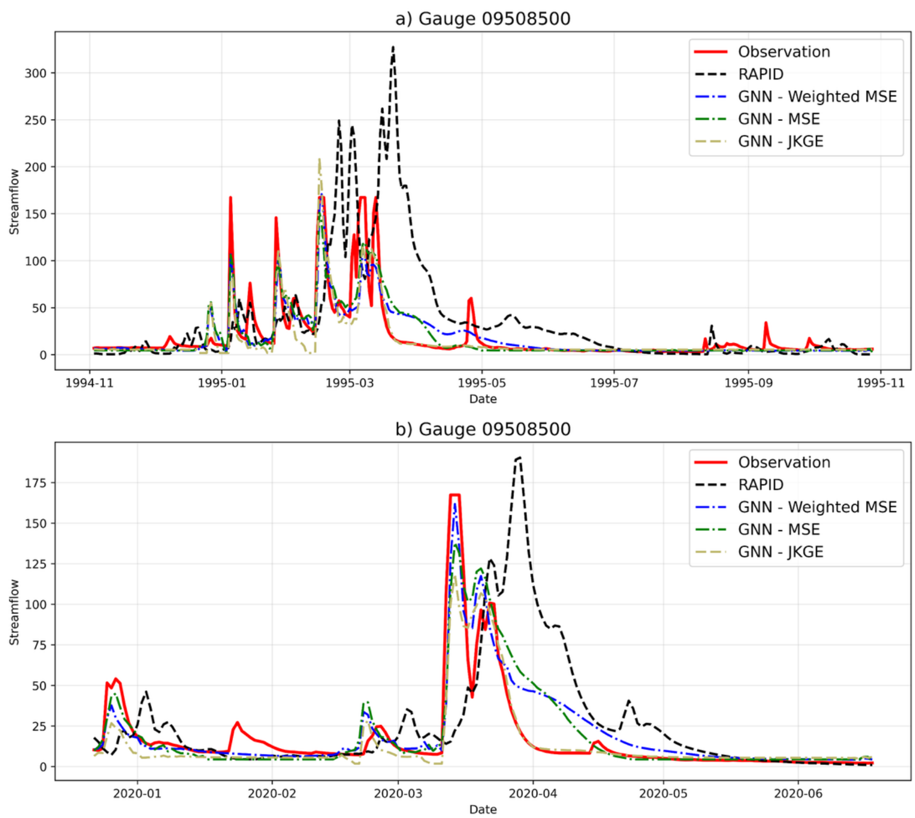


Figure 9. Representative event-scale hydrograph comparisons illustrating the influence of training loss functions on routing behavior at USGS gauge 09508500. Results are shown

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

for observed streamflow, RAPID routing, and graph-routing models trained using weighted MSE, standard MSE, and JKGE objectives.

3.5 Hydrologic Behavior and Event-Scale Routing Evaluation

In addition to aggregate performance metrics, the hydrologic realism of the routing framework was evaluated using representative event-scale hydrographs and flow-duration-curve (FDC) analysis across multiple gauges spanning different portions of the *Salt–Verde* watershed. While $\mathrm{KGE}_{\{ss\}}$-based metrics provide an overall measure of predictive skill, hydrograph and FDC analyses provide complementary insight into the ability of the routing models to reproduce hydrologically meaningful behaviors, including *event timing*, *peak attenuation*, *recession dynamics*, and *long-term streamflow distributions*.

Figure 10 compares representative event-scale hydrographs at four evaluation gauges spanning different hydrologic settings within the watershed. Across all locations, the graph-based routing framework produced substantially more realistic streamflow responses than the RAPID benchmark. In the representative events examined, RAPID exhibits delayed peak timing, exaggerated event magnitudes, and prolonged recession tails following major runoff events. These behaviors are particularly evident at intermediate and downstream gauges, where routing accumulation errors propagate into substantial overestimation of peak discharge and unrealistically persistent post-event flows.

In contrast, the graph-based routing framework reproduced observed hydrograph timing and event magnitude substantially more accurately across the evaluated gauges. The best-performing *weighted-MSE* configuration captured rapid event responses in flashy tributaries while also maintaining hydrologically plausible attenuation behavior in larger integrating basins. Improvements are especially pronounced during multi-peak runoff events, where the graph-routing framework more effectively reproduces the timing and separation of sequential runoff responses. Some smoothing and elevated recession behavior remains in selected cases, but the graph-based routing framework consistently reduces the extreme overestimation and delayed routing behavior observed in RAPID.

Flow-duration-curve analysis further confirms the improved hydrologic realism of the graph-based routing framework (Figure 11). Across representative gauges, RAPID frequently overestimates high-flow magnitudes and exhibited prolonged recession behavior, producing unrealistic streamflow distributions relative to observations. These deviations should be interpreted in the context of the RAPID configuration used in this study, which relied on an existing CONUS-scale routing parameterization and was not recalibrated specifically for the Salt–Verde watershed. Therefore, the comparison should not be interpreted as a general limitation of RAPID, but rather as evidence that fixed large-scale routing parameterizations and standard routing assumptions may require local adaptation or modification to better represent the full range of streamflow conditions in heterogeneous semi-arid river systems.

In contrast, the graph-based routing configurations reproduce observed streamflow distributions much more consistently across high-, intermediate-, and low-flow conditions. The *weighted-MSE* configuration generally provides the closest overall agreement with observed FDC behavior, particularly across the mid-range and high-flow portions of the distribution. The *JKGE-based* configuration occasionally produces improved low-flow agreement, consistent with the recession-behavior patterns discussed in Section 3.4, but generally shows weaker agreement during higher-flow conditions. The improved FDC behavior indicates that the graph-based routing framework better preserves long-term hydrologic variability while maintaining physically meaningful event-scale routing behavior.

Taken together, the hydrograph and FDC analyses demonstrate that the graph-based routing framework improves not only aggregate performance metrics but also hydrologically meaningful streamflow behavior across diverse watershed conditions. By combining physically generated runoff inputs with topology-aware graph routing, the framework more effectively represents *peak timing, event*

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

attenuation, recession dynamics, and *long-term flow distributions* than the conventional RAPID routing benchmark.

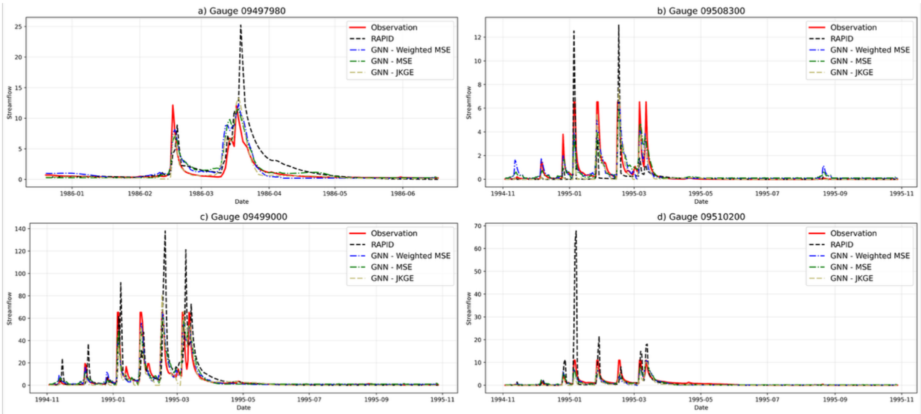
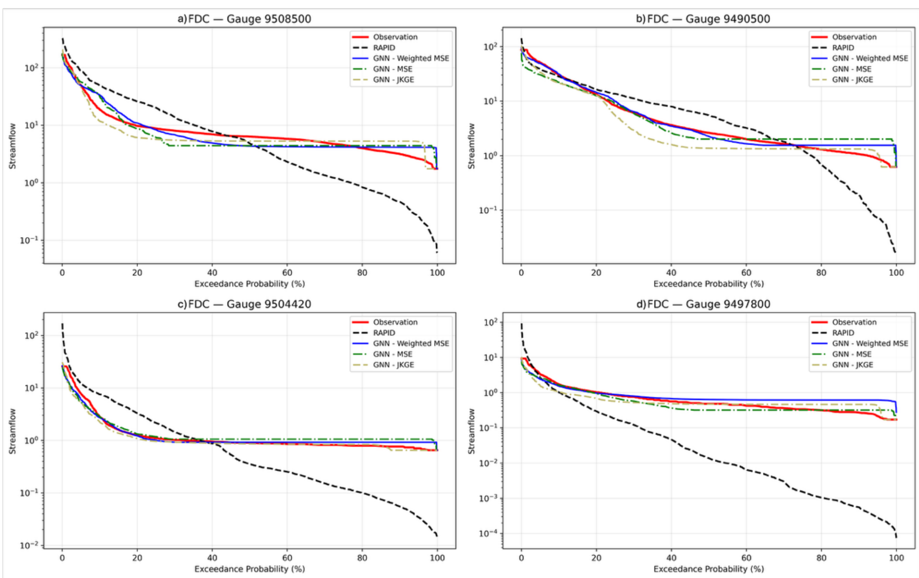


Figure 10. Representative event-scale hydrograph comparisons across four USGS evaluation gauges within the Salt–Verde watershed, showing observed streamflow, RAPID routing, and graph-based routing results using weighted MSE, standard MSE, and JKGE training objectives. The graph-based routing framework generally reproduces event timing, peak magnitude, and recession behavior more realistically than RAPID across diverse hydrologic settings. In the representative cases examined, RAPID frequently exhibits delayed peak response, exaggerated discharge magnitudes, and prolonged recession tails, particularly during multi-peak runoff events and at hydrologically integrated downstream locations.



https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

Figure 11. Flow-duration-curve (FDC) comparisons at four representative USGS gauges within the Salt–Verde watershed for observed streamflow, RAPID routing, and graph-based routing configurations trained using weighted MSE, standard MSE, and JKGE objectives. The graph-based routing framework more consistently reproduces observed streamflow distributions across high-, intermediate-, and low-flow conditions compared to RAPID. Weighted-MSE configurations generally provide the closest overall agreement with observed FDC behavior, while JKGE-based training occasionally improves low-flow representation through sharper recession behavior**.**

3.6 Hydrologic Plausibility and Streamflow–Runoff Consistency

In addition to predictive performance metrics and event-scale hydrograph evaluation, the hydrologic plausibility of the routing frameworks was examined using a watershed-scale streamflow–runoff consistency diagnostic. For each evaluation gauge, the ratio between streamflow and cumulative upstream Noah-MP runoff was computed and compared for RAPID, the graph-based routing framework, and observations (**Figure S4**). This diagnostic provides a relative measure of how well the routed streamflow compares with the runoff volume generated upstream of each gauge and offers insights into whether the routing behavior remains hydrologically consistent across the watershed.

The RAPID routing benchmark produces streamflow-to-upstream-runoff ratios clustered near unity across most evaluation gauges, reflecting the expected conservative behavior of a physics-based routing framework that primarily redistributes runoff through the river network without introducing substantial transformation of the upstream runoff signal. This behavior indicates internally consistent routing relative to the Noah-MP runoff forcing, although agreement with observed watershed response depends on the realism of the underlying runoff-generation inputs.

In contrast, observed streamflow exhibits substantially greater spatial variability in streamflow–runoff ratios across the *Salt–Verde* watershed. Many gauges show ratios below unity, while others exhibit more moderate deviations depending on watershed location and drainage structure. These deviations likely reflect the combined influence of *runoff-generation uncertainty* in Noah-MP, *transmission losses*, *evapotranspiration effects*, *groundwater interactions*, *anthropogenic regulation*, and other *watershed-scale hydrologic processes* that are not explicitly represented in the runoff-routing comparison.

The graph-based routing framework produces streamflow–runoff ratios that more closely resemble the spatial variability observed in the gauge record than the nearly uniform conservative behavior exhibited by RAPID (**Figure S4**). In particular, the graph-based framework captures the heterogeneous spatial structure of runoff–streamflow relationships across both headwater and downstream basins, suggesting that the learned routing transformations better reproduce effective watershed response behavior rather than simply preserving the upstream runoff signal through fixed conservative routing.

These results suggest that graph-based routing improves not only predictive skill but also the realism of watershed-scale runoff–streamflow relationships relative to observations. However, this diagnostic should not be interpreted as a strict mass-conservation evaluation, since discrepancies between observed streamflow and upstream Noah-MP runoff may reflect both routing behavior and biases in the runoff-generation model itself. Nevertheless, the analysis provides an informative hydrologic plausibility check, although it does not by itself distinguish routing-process improvements from compensation for biases in the Noah-MP runoff forcing.

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

3.7 Interpretation of Learned Routing Behavior

To better understand the routing behavior learned by RiverGraphNet, we analyzed the spatial distribution of graph-attention coefficients and their relationships with physical river-network characteristics. **Figure S5a** shows the mean attention coefficients aggregated across the *Salt–Verde* river network. Attention values exhibited substantial spatial variability rather than remaining uniformly distributed throughout the network, indicating that the model learned heterogeneous routing influence among connected river reaches. This behavior suggests that upstream contributions were not treated equally during message passing and that the routing framework selectively emphasized specific flow pathways within the watershed.

Relationships between attention coefficients and physical routing attributes are shown in **Figure S5b**. Mean attention exhibited positive associations with channel width metrics, including both top width (ρ = 0.247) and bottom width (ρ = 0.247), as well as a weaker positive relationship with travel-time proxy (ρ = 0.169). In contrast, attention was negatively associated with channel slope (ρ = −0.190) and Manning roughness (ρ = −0.244), while channel length showed little relationship with attention (ρ = 0.043). These patterns indicate that the model preferentially assigned greater routing influence to wider, lower-gradient, and lower-roughness river reaches, which generally correspond to more hydraulically efficient flow pathways within the network.

Additional insight is provided by the stream-order analysis shown in **Figure S5c**. First-order reaches exhibited the highest median attention values, while attention generally decreased with increasing stream order. This pattern suggests that the routing framework placed greater emphasis on information originating from headwater tributaries and progressively distributed that information through downstream portions of the network. Because headwater reaches represent the locations where runoff signals first enter the river system, the elevated attention assigned to these reaches suggests that accurate propagation of headwater runoff is a critical control on downstream routing performance. The decreasing attention with stream order further indicates that the model progressively integrates information as flow accumulates downstream through the river hierarchy.

To place the stream-order results in the context of the underlying river-network structure, we additionally examined network topology as a function of stream order (**Figure S6**). The number of graph edges does not increase monotonically toward lower stream orders, with second-order reaches representing the most common stream-order class in the network. Similarly, upstream contributing area increases systematically with stream order, indicating that higher-order reaches integrate progressively larger portions of the watershed. Therefore, the elevated attention assigned to first-order reaches cannot be explained solely by network abundance or drainage-area magnitude. Instead, the results suggest that RiverGraphNet preferentially emphasizes headwater runoff signals despite their relatively small contributing areas, highlighting the importance of accurately propagating local runoff generation through the downstream routing hierarchy. This behavior further suggests that routing errors originating in headwater tributaries may accumulate downstream, making accurate representation of upstream runoff signals particularly important for overall routing performance.

To investigate which learned routing behaviors were most strongly associated with performance gains, gauge-level reductions in RMSE relative to RAPID were compared against several attention-derived routing metrics (**Table 4**). The strongest relationships were observed for attention-weighted channel-width metrics, with correlations of ρ = 0.736 for top width and ρ = 0.725 for bottom width. Moderate relationships were also observed for concentration of attention within dominant upstream pathways (ρ = 0.545) and attention-weighted travel-time metrics (ρ = 0.502). These results indicate that the largest improvements over RAPID occurred where the learned routing behavior most strongly emphasized major conveyance pathways and hydraulically efficient portions of the river network.

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

Table 4. Spearman correlations between learned routing metrics and gauge-level RMSE reduction relative to RAPID.

scenario	Spearman ρ
Attention-weighted top channel width shift	0.736
Attention-weighted bottom channel width shift	0.725
Top-10% upstream attention concentration	0.545
Attention-weighted travel-time proxy	0.502

4 Discussion

4.1 Implications for Graph-Based Runoff Routing

The results demonstrate that GNNs can provide an effective framework for routing physically-generated runoff through directed river-network topology. Unlike direct streamflow-forecasting models, the proposed framework isolates the routing component by transferring Noah-MP runoff from gridded model space into NGen river-network graph elements before applying graph-based message passing. This separation is important because it allows the learned model to be evaluated specifically as a routing operator rather than as a combined runoff-generation and streamflow-prediction system.

The strong performance gains relative to RAPID suggest that topology-aware graph learning can capture routing behavior that is difficult to represent using fixed routing assumptions alone. The directed graph structure constrains information propagation to hydrologically-connected pathways, while node and edge attributes allow the model to learn spatially-variable routing relationships across heterogeneous channel and watershed conditions. This provides a flexible routing framework that remains physically structured through the river graph while allowing learned adaptation to local hydrologic behavior.

4.2 Why Graph Routing Improves Performance Relative to RAPID

The largest improvements occurred in headwater and intermediate tributary systems, where runoff responses are often more nonlinear, spatially heterogeneous, and sensitive to local terrain and drainage structure (Bain et al., 2023). In these regions, fixed-parameter routing assumptions may be less able to represent variable travel times, local attenuation behavior, and multi-peak runoff responses. By contrast, the GAT framework learns adaptive upstream–downstream influence weights across the graph, allowing routing behavior to vary across river-network position and physiographic setting.

The hydrograph and FDC results further indicate that the graph-based framework reduces the delayed peaks, exaggerated event magnitudes, and prolonged recession behavior observed in RAPID for representative events. This does not imply that RAPID is generally unsuitable; rather, it suggests that in this application, the combination of Noah-MP runoff forcing and fixed routing assumptions produced systematic mismatch at several gauges. The learned graph-routing framework was able to partially correct these routing-response errors while maintaining consistency with the physical river-network structure.

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

4.3 Temporal Memory and Loss-Function Tradeoffs

The temporal-memory experiments show that intermediate-lag windows, especially the 14-day runoff-lag configuration, provide the strongest overall routing performance. This suggests that graph-based routing benefits from antecedent runoff information over multi-day timescales consistent with snowmelt response, subsurface-flow persistence, and tributary accumulation processes in the *Salt–Verde* watershed. Longer 30-day-lag configurations did not improve performance, indicating that excessive antecedent information may introduce noise or temporal complexity that is less directly relevant to daily routing behavior.

The loss-function experiments revealed an additional tradeoff. *Weighted-MSE* training produced the strongest aggregate routing performance, likely because gauge-balanced optimization stabilized learning across a heterogeneous gauge network. However, *JKGE-based* training occasionally improved recession and low-flow behavior, even though it performed less well overall. This indicates that different loss functions emphasize different hydrologic behaviors. For operational peak-flow prediction in the *Salt–Verde*, *weighted MSE* may be preferable, whereas objectives that better preserve recession dynamics may be useful when low-flow behavior or hydrograph shape is the primary focus.

4.4 Hydrologic Plausibility and Limits of Learned Routing

The streamflow–runoff consistency diagnostic provides an informative hydrologic plausibility check for evaluating how each routing framework transforms upstream runoff into downstream discharge. In this diagnostic, routed streamflow at each gauge is compared with the cumulative Noah-MP-generated runoff originating upstream of that gauge. For a routing framework that primarily redistributes runoff through the river network without substantial transformation, streamflow-to-upstream-runoff ratios would be expected to remain relatively close to unity, subject to numerical losses or routing approximations. This behavior was observed for RAPID, which produced nearly uniform ratios across most gauges, consistent with its physically conservative routing formulation.

In contrast, observed streamflow exhibited substantially greater spatial heterogeneity in these ratios across the watershed. This variability reflects the fact that real watershed response is not governed solely by conservative channel transport of generated runoff. Instead, effective runoff-to-streamflow transformation integrates multiple interacting processes, including runoff-generation uncertainty, transmission losses in ephemeral or semi-arid channels, evapotranspiration losses, infiltration and groundwater recharge, delayed subsurface return flow, channel storage effects, and local anthropogenic influences such as diversions or regulated releases. Consequently, the observed watershed response represents the combined outcome of both runoff production and subsequent hydrologic transformation processes rather than idealized conservative routing alone.

The graph-based routing framework produced streamflow–runoff ratios that more closely resembled the spatial variability exhibited by the observation-based diagnostic than the nearly uniform conservative behavior exhibited by RAPID. This suggests that the learned routing operator does not simply emulate conservative runoff translation, but instead learns effective transformations that better reproduce the integrated watershed-scale response reflected in the observations. From a predictive perspective, this behavior is beneficial because the objective of the routing framework is to reproduce observed discharge rather than merely preserve the raw runoff signal.

However, this interpretation requires caution. Because the diagnostic uses Noah-MP-generated runoff as the upstream forcing reference, discrepancies between routed streamflow and observations cannot be attributed uniquely to routing behavior. Errors in precipitation forcing, snow accumulation and melt timing, soil moisture dynamics, infiltration partitioning, groundwater exchange, or runoff generation within Noah-MP may all propagate into this comparison. As a result, the apparent improvement in

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

hydrologic realism achieved by the graph-based framework may arise from multiple mechanisms: improved representation of routing dynamics, implicit compensation for systematic runoff-generation biases, or a combination of both. Future diagnostic frameworks should seek to separate these effects explicitly, for example by evaluating the routing model using multiple runoff-generation products, synthetic runoff experiments with known routing characteristics, or conservation-constrained routing formulations that limit compensatory adjustments.

This distinction is scientifically important because the current graph-routing framework is designed as a predictive routing emulator rather than a strictly mass-conserving hydrodynamic solver. Although the model remains physically structured through explicit river-network topology, directed connectivity, and physically meaningful node and edge attributes, it does not impose explicit water-balance constraints during optimization. Consequently, the framework is capable of learning effective transformations that improve agreement with observations, but those transformations are not guaranteed to correspond to physically conservative routing behavior.

This highlights an important tradeoff between predictive accuracy and strict physical interpretability in learned hydrologic routing systems. The current framework demonstrates that topology-aware graph learning can substantially improve streamflow prediction relative to conventional routing benchmarks, but future work should investigate hybrid architectures that explicitly incorporate conservation constraints, flow continuity requirements, or differentiable hydrodynamic priors. Such approaches would help distinguish true routing-process improvement from compensatory correction of runoff-generation errors while preserving the predictive flexibility of graph-based learning.

4.5 What Did RiverGraphNet Learn About Routing?

Beyond improving predictive performance, the attention analysis provides insight into the routing behavior learned by RiverGraphNet and offers clues regarding limitations of conventional routing formulations. Unlike RAPID, which routes runoff using prescribed hydraulic relationships and fixed routing parameters, RiverGraphNet learned spatially heterogeneous routing influences throughout the river network. Importantly, the observed relationships should not be interpreted as direct physical flow partitioning. Graph-attention coefficients describe how information is propagated through the learned routing model rather than how water is physically distributed among river reaches. Nevertheless, the attention analysis provides insight into which physical attributes were most influential in the learned routing process. Although hydraulic descriptors such as channel width, slope, roughness, and travel-time metrics were explicitly provided to the model, the learned attention mechanism did not utilize them equally. Instead, systematic preferences emerged across the river network, indicating that certain channel characteristics exerted a stronger influence on information propagation and routing performance than others. This provides a physically interpretable view of how the model routes runoff through the network.

Attention coefficients were positively associated with channel-width metrics and travel-time proxies, while negative relationships were observed with channel slope and Manning roughness. These patterns indicate that the model preferentially emphasized wider, hydraulically efficient flow pathways while reducing the influence of steeper or hydraulically resistant reaches. Such behavior is physically intuitive because wider channels generally provide greater conveyance capacity and also represent locations where runoff contributions from multiple upstream tributaries have already been integrated. Consequently, channel width may serve not only as a hydraulic descriptor but also as an indicator of the relative importance of a pathway within the river-network hierarchy.

More importantly, the routing characteristics most strongly associated with performance improvement were not travel-time metrics alone. Gauge-level reductions in routing error relative to RAPID were most strongly correlated with attention-weighted channel-width metrics, with correlations of $\rho = 0.736$ and $\rho =$

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

0.725 for top and bottom channel width, respectively. Weaker but still substantial relationships were observed for dominant upstream-pathway attention concentration (ρ = 0.545) and attention-weighted travel-time metrics (ρ = 0.502). These results suggest that routing performance depends not only on estimating travel time and attenuation correctly, but also on determining which upstream pathways exert the strongest control on downstream flow. Channel width likely acts as a proxy for major conveyance pathways and accumulated upstream drainage structure, indicating that the model learned where the most influential runoff signals originate and become concentrated within the network, rather than simply how rapidly water propagates downstream. This finding suggests that accurate routing depends not only on representing travel-time and attenuation processes, but also on identifying which upstream tributaries exert the greatest control on downstream discharge. In this sense, the largest performance gains appear to arise from improved representation of network influence and upstream contribution weighting in addition to conventional routing dynamics.

Additional context is provided by the underlying river-network structure (**Figure S6**). Upstream contributing area increases systematically with stream order, indicating that higher-order reaches integrate progressively larger portions of the watershed. However, the attention analysis showed the opposite pattern, with the highest attention values occurring in first-order reaches and progressively lower attention assigned to higher-order channels. Because first-order reaches are not associated with the largest contributing areas, this result suggests that RiverGraphNet is not simply emphasizing the largest rivers or the most aggregated portions of the network. Instead, the model appears to recognize that routing errors originating in headwater tributaries can propagate and accumulate downstream, making accurate representation of headwater runoff signals particularly important for overall routing performance.

An important observation is that RAPID already incorporates substantial spatial variability through reach-specific Muskingum K parameters, which represent travel-time differences throughout the river network (**Figure S7**). Therefore, if travel-time representation alone were the dominant limitation, one would expect the strongest performance gains to be associated with travel-time metrics. Instead, the strongest relationships were observed with attention-weighted channel-width metrics and pathway-concentration measures. This suggests that a key limitation may not be the estimation of travel time itself, but rather the assumption that routing influence is fully determined by predefined hydraulic relationships. Channel width may partially reflect processes not explicitly represented in the routing formulation, including changes in hydraulic conveyance, channel storage, floodplain interaction, transmission losses, and riparian evapotranspiration. In semi-arid systems such as the Salt–Verde watershed, wider channel sections may therefore reflect locations where effective runoff propagation differs from that predicted by travel-time relationships alone. The RiverGraphNet results indicate that allowing routing influence to vary according to channel geometry and network context may provide additional predictive value beyond conventional travel-time parameterization.

This finding provides a useful perspective on why graph-based routing improved performance relative to RAPID. Conventional routing frameworks primarily focus on propagating water through predefined hydraulic relationships, whereas the learned graph-routing framework additionally adapts the relative importance of upstream tributaries according to local channel geometry, network position, and accumulated watershed structure. In other words, the largest improvements were associated with the model's ability to selectively emphasize dominant conveyance pathways rather than treating all upstream contributions according to a fixed routing formulation.

The learned routing behavior suggests a specific limitation of conventional routing formulations: travel-time and attenuation parameters alone may not fully describe how information propagates through complex river networks. In RiverGraphNet, the largest improvements occurred when routing influence

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

was adaptively concentrated along dominant conveyance pathways. This suggests that future versions of RAPID could potentially benefit from routing formulations that allow effective routing influence or attenuation behavior to vary according to channel geometry, stream order, drainage structure, or other network attributes. More generally, hybrid frameworks that retain physically based Muskingum routing while incorporating learned, spatially adaptive weighting of upstream contributions may provide a promising pathway toward improving routing performance without sacrificing physical consistency.

4.6 Limitations and Future Work

Several limitations remain. First, reservoir operations and managed releases are not explicitly represented. Although dam-affected gauges were excluded from the primary training loss, regulated reaches remain an important challenge for applying graph-routing models in operational river systems. Incorporating reservoir storage, release rules, diversions, and irrigation withdrawals would improve applicability in human-influenced basins. One potential approach is to augment the river graph with reservoir-state nodes or self-loop connections that allow information and storage effects to persist locally over time, enabling the model to learn delayed release behavior and operational controls in regulated systems.

Second, the current framework relies on observed streamflow records for supervised training and evaluation. This dependence on gauge observations may limit direct applicability in ungauged or data-sparse regions, where long-term discharge records are unavailable or where the gauge network does not adequately represent spatial variability in runoff-routing behavior. Future work could address this limitation through regional transfer learning, pretraining across multiple well-gauged basins, parameter sharing among hydrologically similar watersheds, and hybrid training strategies that combine limited observations with physics-based routing simulations or conservation-based regularization. These approaches could improve model transferability while reducing dependence on dense local streamflow observations.

Third, the framework depends on the quality of Noah-MP runoff inputs and the accuracy of the grid-to-graph runoff-transfer procedure. Errors in runoff generation, precipitation forcing, soil moisture dynamics, or divide-based runoff aggregation can propagate into the routing model. Future studies should evaluate the framework using multiple runoff forcings, alternative runoff-transfer schemes, and additional basins with contrasting hydroclimatic regimes.

Fourth, the current RiverGraphNet framework does not explicitly enforce mass conservation at individual graph nodes or along river-network edges. Although the model is physically structured through directed river-network topology and physically meaningful node and edge attributes, the learned message-passing operation is not constrained to conserve incoming runoff volume during downstream propagation. As a result, the model may improve agreement with observed streamflow partly by learning effective corrections to biases in Noah-MP runoff generation, precipitation forcing, or grid-to-graph runoff transfer, rather than only by representing conservative river-routing processes. This flexibility can improve predictive skill, but it also limits physical interpretability and makes it difficult to distinguish true routing-process improvements from compensation for runoff-generation errors. Future work should incorporate conservation-aware graph architectures, water-balance regularization, differentiable routing constraints, or hybrid physics–machine learning formulations that preserve mass continuity while retaining the adaptive capacity of graph-based learning.

Fifth, although GAT-attention provides adaptive graph connectivity weighting, attention weights should not automatically be interpreted as physical flow partitioning without additional analysis. Future work should examine whether learned attention patterns correspond to meaningful hydrologic controls such as drainage area, slope, travel time, and upstream contributing structure.

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

Overall, this study demonstrates that topology-aware graph learning can serve as a promising routing framework for converting physically generated gridded runoff into daily streamflow predictions. By explicitly separating runoff generation from routing, the framework provides a useful testbed for evaluating learned river-network routing behavior and for developing future hybrid models that combine physical consistency, mass conservation, and adaptive graph-based learning.

5 Conclusions

This study has developed and tested a standalone graph-based hydrologic routing framework designed to explicitly isolate runoff routing from runoff generation by coupling physically generated Noah-MP runoff with topology-aware graph neural network routing over a directed river-network representation of the *Salt–Verde* watershed. Unlike many existing graph-based hydrologic studies that directly forecast streamflow or generate runoff at graph nodes, the proposed framework focuses specifically on the routing problem by transferring gridded land-surface runoff into a physically-structured river graph derived from the NGen hydrofabric and learning runoff propagation through Graph Attention Network (GAT)-based message passing.

The results demonstrate that graph-based routing can substantially improve streamflow prediction relative to the RAPID physics-based routing benchmark. Across all experimental configurations, the graph-routing framework consistently produced positive performance gains relative to RAPID at almost all gauges, with the strongest configuration achieving a mean $\mathrm{KGE}_{\{ss\}}$ of 0.70 compared with a mean $\mathrm{KGE}_{\{ss\}}$ of −0.29 for RAPID. Spatial analysis showed that improvements were particularly pronounced in headwater and intermediate tributaries, where hydrologic response is highly nonlinear and routing behavior is strongly influenced by heterogeneous drainage structure, terrain complexity, and localized upstream–downstream interactions.

The temporal-routing-memory experiments indicated that routing performance was relatively insensitive to lag length over the range evaluated. Although intermediate runoff-memory representations (7–14 days) consistently produced among the strongest results, differences among lag configurations were generally modest. The no-CNN ablation experiments further indicated that much of the predictive routing skill emerged from the combination of graph topology and lagged runoff inputs, with the temporal convolutional module serving primarily as an additional refinement mechanism rather than the dominant source of temporal routing memory.

Loss-function experiments revealed important tradeoffs between aggregate predictive skill and hydrologic behavior. *Weighted-MSE* optimization produced the strongest overall performance, likely because gauge-balanced optimization improves stability across the heterogeneous evaluation network while preventing dominant gauges from disproportionately influencing training. In contrast, *JKGE-based* objectives often better preserved recession dynamics and low-flow recovery, but generally produced weaker overall predictive performance. These findings highlight that optimization objective selection can meaningfully shape the hydrologic behavior learned by graph-based routing models.

Event-scale hydrograph and flow-duration-curve analyses confirmed that the graph-based routing framework improved not only aggregate performance metrics but also hydrologically meaningful streamflow behavior. Compared with RAPID, the learned routing framework more effectively reproduced *peak timing, event magnitude, hydrograph shape, attenuation behavior*, and *long-term streamflow distributions* across diverse hydrologic settings. These improvements suggest that topology-aware graph message passing provides a flexible mechanism for representing heterogeneous runoff propagation across complex river systems.

Attention analysis provided additional insight into the routing behavior learned by RiverGraphNet. The model preferentially emphasized headwater runoff signals and dominant conveyance pathways, while routing improvements were most strongly associated with channel-width metrics rather than travel-time proxies alone. This suggests that identifying which upstream pathways exert the strongest influence on downstream discharge may be as important as representing travel-time and attenuation processes themselves.

At the same time, the hydrologic plausibility analysis highlighted an important scientific limitation. Although the graph-based routing framework produced watershed-response patterns more consistent with observations than conservative routing alone, the framework does not explicitly enforce mass conservation. Consequently, improved agreement with observed streamflow may reflect both enhanced routing representation and implicit compensation for errors in the Noah-MP runoff-generation forcing. This distinction is important because predictive success does not necessarily imply physically conservative routing realism.

Overall, this study demonstrates that graph neural networks provide a promising framework for hydrologic routing when physically-generated runoff and explicit river-network topology are incorporated into the learning architecture. The results suggest that graph-based routing can outperform conventional routing frameworks, particularly in heterogeneous systems where fixed routing assumptions may be limiting. Future work should focus on hybrid graph-routing architectures that incorporate explicit water-balance constraints, differentiable hydrodynamic priors, reservoir regulation processes, and multi-basin evaluation in order to improve physical interpretability while preserving the predictive flexibility of graph-based learning. More broadly, the results demonstrate that graph neural networks can serve not only as accurate routing models but also as valuable tools for investigating how hydrologic information propagates through complex river networks.

Data and Code Availability

The streamflow observations used in this study were obtained from the U.S. Geological Survey (USGS) National Water Information System (NWIS). River-network topology, hydrofabric attributes, and routing elements were derived from the Next Generation Water Model (NGen) hydrofabric.

Noah-MP runoff simulations used as routing inputs were generated using the modeling framework described in Section 2. The RiverGraphNet source code, training configurations, graph-cache files, evaluation scripts, and supporting analysis workflows are publicly available at [https://github.com/mfarmani95/LSTM_GNN_routing-].

Author Contribution

Mohammad A. Farmani developed the RiverGraphNet framework, implemented the graph-based modeling workflow, performed the graph-routing experiments, analyzed the results, prepared the figures, and wrote the initial manuscript draft. Sadaf Moghisi performed the Noah-MP simulations and RAPID routing experiments, contributed to result analysis, and reviewed the manuscript. Andrew Bennet contributed to the conceptualization of the study, interpretation of the results, and manuscript review. Muhammad Jawad reviewed the manuscript and contributed to interpretation of the results. Hoshin Gupta, Ali Behrangi, and Ahmad A. Tavakoly contributed to manuscript review, interpretation of the results, and refinement of the hydrologic and methodological framing. Guo-Yue Niu contributed to the Noah-MP modeling framework, hydrologic interpretation, supervision, and manuscript review. All authors discussed the results and contributed to the final manuscript.

Competing interest

The contact author has declared that none of the authors has any competing interests.

Acknowledgments

Funding for this project was provided by the National Oceanic and Atmospheric Administration (NOAA), awarded to the Cooperative Institute for Research on Hydrology (CIROH) through the NOAA Cooperative Agreement with The University of Alabama, NA22NWS4320003. This research was supported in part by an appointment to the Department of Defense (DOD) Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and the DOD. ORISE is managed by ORAU under DOE contract number DE-SC0014664. All opinions expressed in this paper are the author's and do not necessarily reflect the policies and views of DOD, DOE, or ORAU/ORISE.

References

Acosta, C. M., Herath, H. M. V. V., Lim, J. Y., Saha, A., Rasnayaka, S., and Marshall, L.: Physics-informed Graph Neural Networks for Operational Flood Modeling, arXiv preprint arXiv:2512.23964, 2025.

Agnihotri, J., Behrangi, A., Tavakoly, A., Geheran, M., Farmani, M. A., and Niu, G. Y.: Higher Frozen Soil Permeability Represented in a Hydrological Model Improves Spring Streamflow Prediction From River Basin to Continental Scales, Water Resources Research, 59, https://doi.org/10.1029/2022WR033075, 2023a.

Agnihotri, J., Behrangi, A., Tavakoly, A., Geheran, M., Farmani, M. A., and Niu, G. Y.: Higher Frozen Soil Permeability Represented in a Hydrological Model Improves Spring Streamflow Prediction From River Basin to Continental Scales, Water Resources Research, 59, e2022WR033075, 10.1029/2022WR033075, 2023b.

Akkala, A., Boubrahimi, S. F., Hamdi, S. M., Hosseinzadeh, P., and Nassar, A.: Spatio-Temporal Graph Neural Networks for Streamflow Prediction in the Upper Colorado Basin, Hydrology, 12, 10.3390/hydrology12030060, 2025.

Bain, R. L., Shaw, M. J., Geheran, M. P., Tavakoly, A. A., Wahl, M. D., and Zsoter, E.: Intercomparison of global ERA reanalysis products for streamflow simulations at the high-resolution continental scale, Journal of Hydrology, 616, 10.1016/j.jhydrol.2022.128624, 2023.

Bindas, T., Tsai, W. P., Liu, J., Rahmani, F., Feng, D., Bian, Y., Lawson, K., and Shen, C.: Improving River Routing Using a Differentiable Muskingum-Cunge Model and Physics-Informed Machine Learning, Water Resources Research, 60, 10.1029/2023wr035337, 2024.

Cortés-Salazar, N., Vásquez, N., Mizukami, N., Mendoza, P. A., and Vargas, X.: To what extent does river routing matter in hydrological modeling?, Hydrology and Earth System Sciences, 27, 3505-3524, 10.5194/hess-27-3505-2023, 2023.

David, C. H., Maidment, D. R., Niu, G.-Y., Yang, Z.-L., Habets, F., and Eijkhout, V.: River Network Routing on the NHDPlus Dataset, Journal of Hydrometeorology, 12, 913-934, 10.1175/2011jhm1345.1, 2011.

Fall, G., Kitzmiller, D., Pavlovic, S., Zhang, Z., Patrick, N., St. Laurent, M., Trypaluk, C., Wu, W., and Miller, D.: The Office of Water Prediction's Analysis of Record for Calibration, version 1.1: Dataset description and precipitation evaluation, JAWRA Journal of the American Water Resources Association, 59, 1246-1272, 10.1111/1752-1688.13143, 2023.

Farmani, M. A., Behrangi, A., Gupta, A., Tavakoly, A., Geheran, M., and Niu, G.-Y.: Do land models miss key soil hydrological processes controlling soil moisture memory?, Hydrology and Earth System Sciences, 29, 547-566, 10.5194/hess-29-547-2025, 2025.

Farmani, M. A., Tavakoly, A. A., Behrangi, A., Qiu, Y., Gupta, A., Jawad, M., Sohi, H. Y., Zhang, X.-Y., Geheran, M. P., and Niu, G.: Improving Streamflow Predictions in the Arid Southwestern United States Through Understanding of Baseflow Generation Mechanisms, 10.22541/essoar.173272456.69006273/v1, 2024.

Fenton, J. D.: Flood routing methods, Journal of Hydrology, 570, 251-264, 10.1016/j.jhydrol.2019.01.006, 2019.

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

Kazadi, A., Doss-Gollin, J., Sebastian, A., and Silva, A.: FloodGNN-GRU: a spatio-temporal graph neural network for flood prediction, Environmental Data Science, 3, 10.1017/eds.2024.19, 2024.

Lees, T., Buechel, M., Anderson, B., Slater, L., Reece, S., Coxon, G., and Dadson, S. J.: Benchmarking data-driven rainfall–runoff models in Great Britain: a comparison of long short-term memory (LSTM)-based models with four lumped conceptual models, Hydrology and Earth System Sciences, 25, 5517-5534, 10.5194/hess-25-5517-2021, 2021.

Li, H., Wigmosta, M. S., Wu, H., Huang, M., Ke, Y., Coleman, A. M., and Leung, L. R.: A Physically Based Runoff Routing Model for Land Surface and Earth System Models, Journal of Hydrometeorology, 14, 808-828, 10.1175/jhm-d-12-015.1, 2013.

Lima, M., Deck, K., Dunbar, O. R. A., and Schneider, T.: Toward routing river water in land surface models with recurrent neural networks, Hydrology and Earth System Sciences, 29, 3145-3164, 10.5194/hess-29-3145-2025, 2025.

Liu, G., Ouyang, S., Qin, H., Liu, S., Shen, Q., Qu, Y., Zheng, Z., Sun, H., and Zhou, J.: Assessing spatial connectivity effects on daily streamflow forecasting using Bayesian-based graph neural network, Sci Total Environ, 855, 158968, 10.1016/j.scitotenv.2022.158968, 2023.

Mizukami, N., Clark, M. P., Sampson, K., Nijssen, B., Mao, Y., McMillan, H., Viger, R. J., Markstrom, S. L., Hay, L. E., Woods, R., Arnold, J. R., and Brekke, L. D.: mizuRoute version 1: a river network routing tool for a continental domain water resources applications, Geoscientific Model Development, 9, 2223-2238, 10.5194/gmd-9-2223-2016, 2016.

Mosaffa, H., Pappenberger, F., Prudhomme, C., Chantry, M., Rüdiger, C., and Cloke, H.: A GNN routing module is all you need for LSTM Rainfall–Runoff models, Hydrology and Earth System Sciences, 30, 2079-2092, 10.5194/hess-30-2079-2026, 2026.

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

Nijssen, B., O'Donnell, G. M., Lettenmaier, D. P., Lohmann, D., and Wood, E. F.: Predicting the Discharge of Global Rivers, Journal of Climate, 14, 3307-3323, 10.1175/1520-0442(2001)014<3307:Ptdogr>2.0.Co;2, 2001.

Niu, G., Fang, Y., Neto, A. A. M., Guo, B., Zhang, X.-Y., Farmani, M. A., Behrangi, A., and Zeng, X.: Representing Preferential Flow through Variably-Saturated Soils with Surface Ponding in a Large-Scale Land Surface Model over the Conterminous US, 10.22541/essoar.172286649.90332939/v1, 2024.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, Journal of Geophysical Research, 116, 10.1029/2010jd015139, 2011.

Niu, G. Y., Fang, Y. H., Chang, L. L., Jin, J., Yuan, H., and Zeng, X.: Enhancing the Noah-MP Ecosystem Response to Droughts With an Explicit Representation of Plant Water Storage Supplied by Dynamic Root Water Uptake, Journal of Advances in Modeling Earth Systems, 12, e2020MS002062, 10.1029/2020MS002062, 2020a.

Niu, G. Y., Fang, Y. H., Chang, L. L., Jin, J., Yuan, H., and Zeng, X.: Enhancing the Noah-MP Ecosystem Response to Droughts With an Explicit Representation of Plant Water Storage Supplied by Dynamic Root Water Uptake, Journal of Advances in Modeling Earth Systems, 12, 10.1029/2020ms002062, 2020b.

Ogden, F. L., Jennings, K., Clark, E. P., Coon, E., Cosgrove, B., da Cunha, L. K., Farthing, M. W., Flowers, T., Frame, J. M., Frazier, N. J., Garrett, J. L., Graziano, T. M., Hughes, J. D., Johnson, J. M., McDaniel, R., Moulton, J. D., Peckham, S. D., Salas, F. R., Savant, G., Viger, R., and Wood, A.: The NextGen Water Resources Modeling Framework: Community Innovation at the Intersection of Hydrologic, Data and Computer Sciences, JAWRA Journal of the American Water Resources Association, 62, 10.1111/1752-1688.70089, 2026.

Roudbari, N. S., Punekar, S. R., Patterson, Z., Eicker, U., and Poullis, C.: From data to action in flood forecasting leveraging graph neural networks and digital twin visualization, Sci Rep, 14, 18571, 10.1038/s41598-024-68857-y, 2024.

Salvati, A., Moghaddam Nia, A., Salajegheh, A., Shirzadi, A., Shahabi, H., Ahmadisharaf, E., Han, D., and Clague, J. J.: A systematic review of Muskingum flood routing techniques, Hydrological Sciences Journal, 69, 810-831, 10.1080/02626667.2024.2324132, 2024.

Sarkar, A., Hakimi, A., Chen, X., Huang, H., Lu, C., Demir, I., and Jannesari, A.: HydroGAT: Distributed Heterogeneous Graph Attention Transformer for Spatiotemporal Flood Prediction, Proceedings of the 33rd ACM International Conference on Advances in Geographic Information Systems, 1019-1030,

Seo, H., Lee, J., Choi, S., Kim, J., and Kim, Y.: Role of river network information in streamflow prediction using graph wavenet and the entity-aware long short-term memory model, Journal of Hydrology, 10.1016/j.jhydrol.2026.135584, 2026.

https://doi.org/10.5194/egusphere-2026-4157
Preprint. Discussion started: 4 August 2026

Shangguan, W., Hengl, T., Mendes de Jesus, J., Yuan, H., and Dai, Y.: Mapping the global depth to bedrock for land surface modeling, Journal of Advances in Modeling Earth Systems, 9, 65-88, 10.1002/2016ms000686, 2017.

Slater, L., Blougouras, G., Deng, L., Deng, Q., Ford, E., Hoek van Dijke, A., Huang, F., Jiang, S., Liu, Y., Moulds, S., Schepen, A., Yin, J., and Zhang, B.: Challenges and opportunities of ML and explainable AI in large-sample hydrology, Philos Trans A Math Phys Eng Sci, 383, 20240287, 10.1098/rsta.2024.0287, 2025.

Sun, A. Y., Jiang, P., Yang, Z.-L., Xie, Y., and Chen, X.: A graph neural network (GNN) approach to basin-scale river network learning: the role of physics-based connectivity and data fusion, Hydrology and Earth System Sciences, 26, 5163-5184, 10.5194/hess-26-5163-2022, 2022.

Taghizadeh, M., Zandsalimi, Z., Nabian, M. A., Shafiee-Jood, M., and Alemazkoor, N.: Interpretable physics-informed graph neural networks for flood forecasting, Computer-Aided Civil and Infrastructure Engineering, 40, 2629-2649, 10.1111/mice.13484, 2025.

Tavakoly, A. A., David, C. H., Gutenson, J. L., Wahl, M. W., and Follum, M.: Development of non-data driven reservoir routing in the routing application for parallel computatIon of discharge (RAPID) model, Environmental Modelling & Software, 161, 10.1016/j.envsoft.2023.105631, 2023.

Tavakoly, A. A., Snow, A. D., David, C. H., Follum, M. L., Maidment, D. R., and Yang, Z. L.: Continental-Scale River Flow Modeling of the Mississippi River Basin Using High-Resolution NHDPlus Dataset, JAWRA Journal of the American Water Resources Association, 53, 258-279, 10.1111/1752-1688.12456, 2016.

Wan, S., Shah, R., Deng, Q., Sabo, J., Liu, H., and Candan, K. S.: Spatio-temporal causal learning for streamflow forecasting, 2024 IEEE International Conference on Big Data (BigData), 6161-6170,

Wang, H., Chen, J., Zheng, Y., and Song, X.: Accelerating flood warnings by 10 hours: the power of river network topology in AI-enhanced flood forecasting, npj Natural Hazards, 2, 10.1038/s44304-025-00083-6, 2025.

Wang, W., He, C., Moore, J., Wang, G., and Niu, G. Y.: Physics-Based Narrowband Optical Parameters for Snow Albedo Simulation in Climate Models, Journal of Advances in Modeling Earth Systems, 14, 10.1029/2020ms002431, 2022.

Wang, Y. H., Broxton, P., Fang, Y., Behrangi, A., Barlage, M., Zeng, X., and Niu, G. Y.: A Wet-Bulb Temperature-Based Rain-Snow Partitioning Scheme Improves Snowpack Prediction Over the Drier Western United States, Geophysical Research Letters, 46, 13825-13835, 10.1029/2019gl085722, 2019.

Xiang, Z. and Demir, I.: High-resolution rainfall-runoff modeling using graph neural network, arXiv preprint arXiv:2110.10833, 2021.

Yamazaki, D., de Almeida, G. A. M., and Bates, P. D.: Improving computational efficiency in global river models by implementing the local inertial flow equation and a vector-based river network map, Water Resources Research, 49, 7221-7235, 10.1002/wrcr.20552, 2013.

Zhao, X., Wang, H., Bai, M., Xu, Y., Dong, S., Rao, H., and Ming, W.: A Comprehensive Review of Methods for Hydrological Forecasting Based on Deep Learning, Water, 16, 10.3390/w16101407, 2024.