Check for updates

Geophysical Research Letters

RESEARCH LETTER
10.1029/2026GL122902

A Pure Data-Driven Fully Distributed Model for Spatially Consistent Streamflow Prediction

Qiang Yu[1,2], Liguang Jiang[1], Shuo Wang[2], and Jun Liu[3]

[1]School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen, China, [2]Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China, [3]Department of Hydrology, Geological Survey of Denmark and Greenland, Copenhagen, Denmark

Key Points:

- We proposed a pure machine learning (ML) model for distributed streamflow prediction without explicit routing modules
- The model can generate spatially consistent and plausible streamflow prediction at grid scale
- The model performs well in ungauged basins, comparable to or outperforming physically-based and ML models with routing modules

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

L. Jiang,
jianglg@sustech.edu.cn

Citation:

Yu, Q., Jiang, L., Wang, S., & Liu, J. (2026). A pure data-driven fully distributed model for spatially consistent streamflow prediction. *Geophysical Research Letters*, *53*, e2026GL122902. https://doi.org/10.1029/2026GL122902

Received 13 MAR 2026
Accepted 7 AUG 2026

Author Contributions:

Conceptualization: Liguang Jiang
Data curation: Qiang Yu
Formal analysis: Qiang Yu, Liguang Jiang
Funding acquisition: Liguang Jiang
Methodology: Qiang Yu, Liguang Jiang
Software: Qiang Yu
Supervision: Liguang Jiang, Shuo Wang
Validation: Qiang Yu
Visualization: Qiang Yu
Writing – original draft: Qiang Yu
Writing – review & editing: Liguang Jiang, Shuo Wang, Jun Liu

© 2026 The Author(s). Geophysical Research Letters published by Wiley Periodicals LLC on behalf of American Geophysical Union.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Abstract Streamflow prediction is essential for water resources management and flood mitigation but remains challenging in large basins due to spatial heterogeneity. Machine learning (ML) models have been widely applied to streamflow prediction, typically relying on explicit routing modules, whereas the potential of pure ML models remains underexplored. This study presents a purely ML-based model (Convolutional Entity-Aware Long Short-Term Memory) for fully distributed streamflow prediction. Spatial cross-validation across 50 gauges in the Amazon Basin yields a median Nash-Sutcliffe Efficiency of 0.73 and an R^2 of 0.84, comparable to or exceeding benchmark models with routing modules. The model reduces annual peak flow bias (APFB = 0.12) and accurately reproduced spatial autocorrelation patterns (Pearson's $r > 0.6$, Moran's $I > 0.8$, and Geary's $C < 0.2$), indicating spatially consistent grid-scale discharge predictions. These findings demonstrate the potential of a pure data-driven ML model for fully distributed streamflow prediction in ungauged basins while ensuring spatial consistency.

Plain Language Summary Predicting streamflow is essential for managing water resources and preventing floods, but it is challenging in large, diverse river basins where landscape and weather conditions vary greatly across regions. Traditional methods rely on detailed physical river maps or complex flow routing rules. This study developed a pure machine learning model to predict streamflow across entire river basins without using pre-programmed flow routing rules. The model was tested in the Amazon Basin, achieving an accuracy comparable to or surpassing that of traditional models. It also produced realistic, spatially consistent streamflow patterns, despite not incorporating explicit flow routing rules. These results suggest that pure data-driven models can provide accurate and spatially consistent streamflow predictions.

1. Introduction

Streamflow prediction is a fundamental task in hydrology and is critical for effective water resources management, natural hazard mitigation, and environmental conservation (Sun et al., 2014; Troin et al., 2021; Yaseen et al., 2016; Yu, Tolson, et al., 2024). Accurate streamflow predictions are essential for forecasting extreme hydrological events, such as floods and droughts, thereby enabling timely interventions to mitigate their adverse social and environmental impacts (Hadi & Tombul, 2018; Meng et al., 2019). Hydrological models, designed to simulate various processes within the water cycle, play a crucial role in streamflow prediction (K. J. Beven, 2012).

Hydrological models are generally classified as lumped or distributed, depending on their representation of spatial heterogeneity in hydrological processes (K. J. Beven, 2012; Singh & Frevert, 2010). Lumped models conceptualize an entire basin as a single homogeneous unit, offering computational efficiency but losing accuracy in large or heterogeneous basins (Getirana et al., 2009; Kling & Gupta, 2009; Savenije, 2010; Sorooshian, 2008). In contrast, distributed models explicitly account for spatial variability in meteorology, topography, geology, and land cover, enabling more detailed and physically realistic simulations of hydrological processes (Nijzink et al., 2016).

Depending on the spatial discretization unit (i.e., grid cells or sub-basins), distributed models can be categorized as fully or semi-distributed (Yu, Tolson, et al., 2024). As one branch of hydrological models, distributed models can also be further divided into physically based, conceptual, and data-driven models (Devia et al., 2015). Physically based models simulate hydrological processes such as evapotranspiration, infiltration, and runoff based on governing physical laws (Zhang et al., 2016). Conceptual models, on the other hand, combine simplified physical representations with empirical formulations (Aghakouchak & Habib, 2010). These two categories are

19448007, 2026, 16, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL122902 by HONG KONG POLYTECHNIC UNIVERSITY HU NG HOM, Wiley Online Library on [01/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

often referred to as physically-informed models, as exemplified by physically-informed distributed models such as MIKE SHE (Abbott et al., 1986), TOPMODEL (K. Beven et al., 1984), and VIC (Liang et al., 1994). Although widely applied over the past few decades, physically-informed models typically require extensive parameterization and calibration, which can limit their practical applicability, particularly at large spatial scales or in data-scarce regions (Francés et al., 2007; Immerzeel & Droogers, 2008).

Machine learning (ML) models, a prominent category of data-driven models, have shown promising performance in streamflow prediction (Kratzert et al., 2018; Kratzert, Klotz, Herrnegger, et al., 2019; Yu, Tolson, et al., 2024). Unlike physically-informed models, ML models can capture complex nonlinear relationships in hydrological systems directly from historical data, without relying on explicit physical representations (Ji et al., 2021; Y. Lin et al., 2021). When trained on large and diverse data sets, regional ML models have been shown to outperform traditional physically-informed models, particularly in streamflow prediction in ungauged basins (PUB) (Feng et al., 2022; Kratzert, Klotz, Herrnegger, et al., 2019; Yu, Jiang, et al., 2024). However, most existing studies employ ML models that focus on lumped predictions at basin outlets, thereby neglecting spatial heterogeneity in hydrological responses (Muhebwa et al., 2024). Although recent studies have extended ML models toward distributed streamflow prediction by incorporating spatially heterogeneous inputs (Anderson & Radić, 2022; Li et al., 2025; Wunsch et al., 2022), many remain limited to prediction at basin outlets or fixed gauge stations, rather than producing spatially consistent, grid-scale streamflow predictions across entire basins (Wang et al., 2024).

Recent studies that apply ML models to fully distributed streamflow prediction generally incorporate explicit routing modules. For example, Yang et al. (2025) coupled a Long Short Term Memory (LSTM) network for runoff generation with the RAPID routing model, while Xiang and Demir (2022) introduced a graph neural network based rainfall-runoff model that integrates routing informed by flow directions derived from elevation data. Longyang et al. (2024) employed a Convolutional Long Short-Term Memory (ConvLSTM) model with an integrated spatial attention mechanism to predict grid-scale discharge, followed by Ridge regression for spatial aggregation instead of a physical routing scheme due to the absence of reliable surface topography information. In contrast, few studies have investigated the potential of pure ML models for fully distributed streamflow prediction. Li et al. (2025) proposed a temporal to spatial framework that applies LSTM models independently at each grid cell and uses convolutional neural networks (CNNs) for spatial aggregation. However, most of these studies evaluate model performance at basin outlets or fixed gauges. Comprehensive assessments at the fully distributed scale remain limited, particularly for ungauged basins and spatially consistent grid-scale predictions.

To address this gap, this study introduces a purely ML distributed hydrological model based on a stacked two-layer ConvLSTM architecture. The proposed model is applied to the Amazon Basin. Model performance is evaluated at two spatial scales. At the sub-basin scale, a spatial cross-validation experiment across observed gauge stations is designed to assess predictive performance in ungauged basins. At the grid scale, reliability and spatial consistency are evaluated for individual grid cells across the entire basin.

2. Methodology

2.1. Study Area and Data Set

This study focuses on the Amazon Basin (Figure S1 in Supporting Information S1), the largest river basin in the world, located in South America and spanning over 6,000,000 km^2. The Amazon Basin is characterized by complex hydrological processes, pronounced spatial heterogeneity, and significant variability in precipitation. The northwestern region receives some of the highest annual rainfall globally, with totals reaching up to 3,500 mm.

The model input consists of three primary components, that is, meteorological forcing data, static attributes, and terrestrial water storage anomalies (TWSA). Eight meteorological variables, including precipitation, temperature, and wind speed, are obtained from the WATCH-Forcing-Data-ERA-Interim (WFDEI) data set (Lange & Büchner, 2020) (see Table S1 in Supporting Information S1). Static attributes, derived from Stacke and Hagemann (2021) and supplemented by Wang et al. (2024), include 25 properties related to topography, soil, and vegetation (see Table S2 in Supporting Information S1). The TWSA data are obtained from the ITSG-Grace2018 data set (Kvas et al., 2019; Mayer-Gürr et al., 2018), which provides Kalman-smoothed daily gravity field solutions to recover sub-monthly gravity variations from GRACE observations. Streamflow observations from 50 hydrological gauge stations, sourced from the Global Runoff Data Center (GRDC), are used for model training

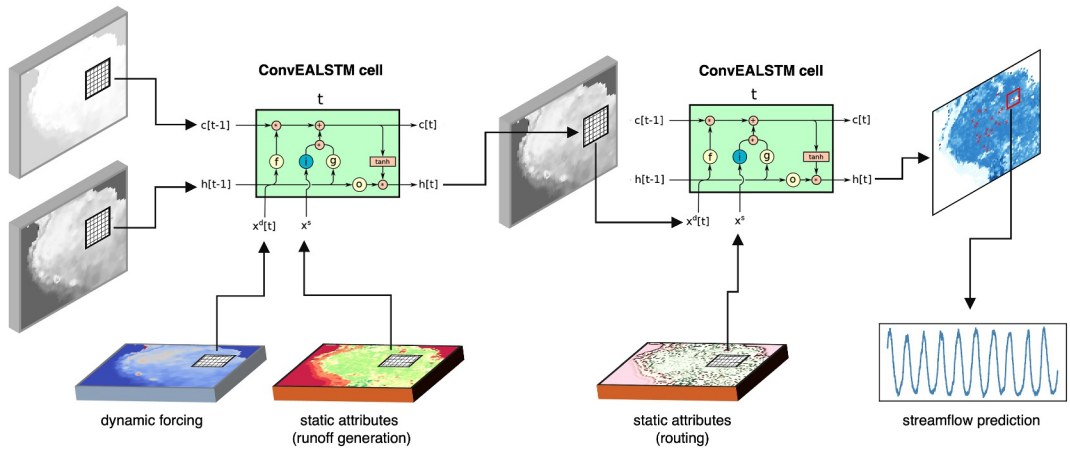


Figure 1. Schematic diagram of the proposed ConvEALSTM model for fully distributed streamflow prediction.

and validation. The distribution of the gauge stations is shown in Figure S1a in Supporting Information S1, and their basin information is provided in Table S3 in Supporting Information S1. These gauge stations cover upstream catchment areas exceeding 4,000 km^2; therefore, the model resolution was set to 0.5° to balance computational efficiency and spatial detail.

2.2. Model Structure

The Convolutional LSTM (ConvLSTM) extends the traditional LSTM by incorporating convolutional structures, effectively substituting traditional matrix multiplications with convolutional operations within both input-to-state and state-to-state transitions (Shi et al., 2015). This design leverages the spatial feature extraction capabilities of CNNs and the temporal dependency modeling strengths of LSTMs, making it particularly well-suited for spatiotemporal prediction tasks (Bassah et al., 2025). ConvLSTM has been successfully applied in various hydrometeorological forecasting tasks, including precipitation (Kim et al., 2017), water level (Guo et al., 2021), and streamflow prediction (Dehghani et al., 2023).

The structure of the proposed stacked two-layer ConvLSTM model is shown in Figure 1. In this architecture, the standard LSTM cell is replaced with the Entity-Aware LSTM (EA-LSTM) developed by Kratzert, Klotz, Shalev, et al. (2019), which incorporates an elegant design for integrating static attributes into the LSTM framework. Incorporating static attributes into LSTM networks has been shown to improve streamflow prediction, particularly in ungauged basins (Feng et al., 2021; Kratzert, Klotz, Herrnegger, et al., 2019). Based on this design, the proposed model is referred to as Convolutional Entity-Aware LSTM (ConvEALSTM). The first layer is designed for runoff prediction and receives dynamic forcing data (meteorological variables and TWSA) together with static attributes related to runoff generation. The second layer is designed for flow routing, using the output from the first layer along with additional routing-related static attributes. The model output represents streamflow predictions at the grid-scale level, which are then compared with streamflow observations from gauge stations according to their geographic locations.

2.3. Experimental Design

The proposed model was evaluated at both the sub-basin and grid scales and compared with benchmark data sets. At the sub-basin scale, a five-fold cross-validation experiment was conducted to assess the model's spatial extrapolation capability for streamflow prediction. In this experiment, the model was trained on daily meteorological forcing and TWSA data from 2003 to 2012 to generate daily streamflow outputs. In each fold, streamflow data from 40 of the 50 gauge stations were used for training, while the remaining 10 stations were reserved for validation, representing a PUB scenario. Before model training, input and output variables were standardized using z-score normalization. Streamflow values were first normalized by catchment area at each gauge station before applying z-score normalization. The model was optimized using the Nash-Sutcliffe Efficiency (NSE). Since the loss function requires minimization, 1-NSE was adopted as the loss function. The hyperparameters (e.g., hidden nodes and learning rate) were tuned using grid search, and their details are provided

19448007, 2026, 16, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL122902 by HONG KONG POLYTECHNIC UNIVERSITY HU NG HOM, Wiley Online Library on [01/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19448007, 2026, 16, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL122902 by HONG KONG POLYTECHNIC UNIVERSITY HU NG HOM, Wiley Online Library on [01/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

in the model code. Since observations were available only at gauge stations, a spatial mask was applied to restrict loss calculations to grid cells containing observations. To reduce uncertainty arising from data set partitioning, the five-fold cross-validation was repeated 10 times, and the median prediction across repetitions was used to represent model performance.

For benchmarking, streamflow predictions were compared with three data sets, that is, the Global Flood Awareness System (GloFAS) v4.0 (Grimaldi et al., 2022), the GRADES data set (P. Lin et al., 2019; Yang et al., 2019) and the GRADES-hydroDL data set (Yang et al., 2025). The GloFAS v4.0 data set is a global, gridded, daily river discharge data set covering the period from 1979 to the near present, produced by the Copernicus Emergency Management Service using the LISFLOOD model forced with ERA5 meteorological reanalysis data, with a spatial resolution of 0.05°. The GRADES and GRADES-hydroDL data sets provide global, reach-level, daily river discharge estimates by utilizing the VIC land surface model and LSTM model, respectively, for runoff generation at a 0.25° grid-scale resolution. This runoff is subsequently routed through the global river network using the Routing Application for Parallel Computation of Discharge (RAPID) model, producing discharge estimates for approximately 2.94 million river reaches. Model performance in spatial cross-validation experiments was assessed using NSE and the coefficient of determination (R^2), with results compared with the three benchmark data sets.

As for the model evaluation at the grid scale, model outputs were compared with two grid-scale benchmark data sets: (a) the Global Runoff Reconstruction (GRUN) data set, which provides monthly runoff estimates at 0.5° spatial resolution for 1902–2014 using a random forest model driven by monthly precipitation and near-surface temperature from atmospheric reanalysis, and (b) the GloFAS gridded daily streamflow data set. Performance at the grid scale was evaluated using the Pearson correlation coefficient between ConvEALSTM predictions and GRUN runoff, as well as between ConvEALSTM predictions and GloFAS streamflow, across all basin grid cells. As GRUN runoff is available at monthly resolution, model outputs were temporally aggregated to the monthly scale for direct comparison.

3. Results and Discussion

3.1. Model Performance at Sub-Basin Scale

Figures 2a and 2b present boxplots of NSE and R^2 values for ConvEALSTM and the three benchmark models across 50 stations in the cross-validation experiment. The ConvEALSTM model achieved a median NSE of 0.73, approximately 7% higher than that of the LISFLOOD model (0.68), 43% higher than that of the VIC-RAPID model (0.51) from the GRADES data set and comparable to the Grid LSTM-RAPID model (0.73) from the GRADES-hydroDL data set. In terms of R^2, the ConvEALSTM model reached a median value of 0.84, higher than VIC-RAPID (0.77) and marginally higher than LISFLOOD (0.80) and Grid LSTM-RAPID (0.83). The incorporation of TWSA as an input substantially enhanced performance: the median NSE increased from 0.46 to 0.73 (a 59% improvement), while the median R^2 improved from 0.58 to 0.84 (a 45% improvement). Moreover, the relatively narrow interquartile range of NSE values for ConvEALSTM indicates stability comparable to Grid LSTM-RAPID and superior to LISFLOOD and VIC-RAPID models. It is important to note that calibration of all LISFLOOD, VIC-RAPID and Grid LSTM-RAPID involved observed streamflow data from GRDC gauge stations within the Amazon Basin, resulting in partial overlap between training and validation stations. In contrast, all validation stations in the ConvEALSTM experiment were excluded from the training set, rendering the evaluation more representative of predictions in ungauged basins.

Figures 2c–2g illustrates the spatial distribution of NSE values for ConvEALSTM and the three benchmark models in the spatial cross-validation experiments. LISFLOOD, VIC-RAPID and Grid LSTM-RAPID perform poorly in several basins (as indicated by orange and red markers), whereas ConvEALSTM consistently outperforms the benchmarks across most regions. Figure 2h compares observed streamflow time series with simulations from the three models at the Óbidos station, the most downstream gauge in the Amazon Basin, with an upstream catchment area of 4,680,000 km^2. ConvEALSTM more accurately captures hydrograph patterns, particularly peak flows, while LISFLOOD, VIC-RAPID and Grid LSTM-RAPID tend to overestimate peak magnitudes. Specifically, ConvEALSTM yields an annual peak flow bias (APFB) of 0.12, which is lower than the values of 0.27, 0.22, and 0.16 obtained by LISFLOOD, VIC-RAPID, and Grid LSTM-RAPID, respectively, thereby demonstrating the superior performance of ConvEALSTM in peak flow prediction.

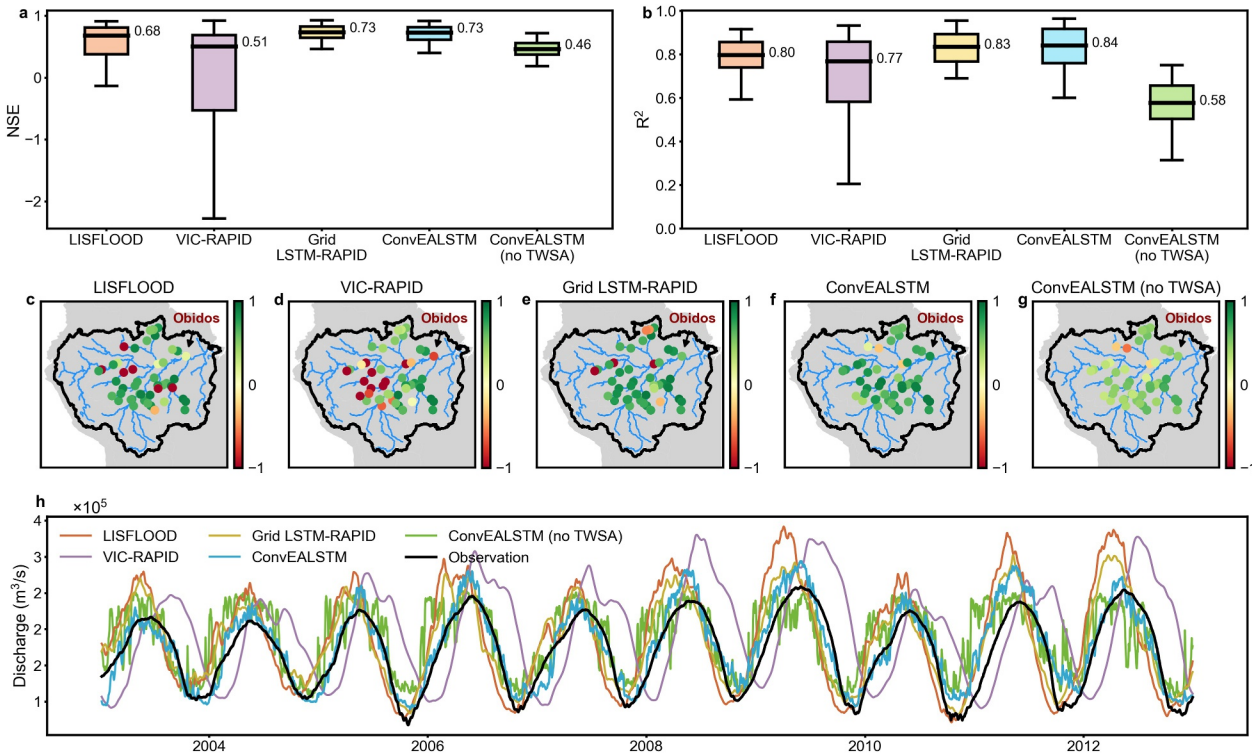


Figure 2. Comparison of streamflow prediction performance among models. (a, b) Boxplot of Nash-Sutcliffe Efficiency (NSE) and R^2 values (Note: Outliers are not marked in this figure). (c–g) Spatial distribution maps of NSE values. (h) Time series of observed and simulated streamflow at the most downstream Óbidos station.

A comparison between ConvEALSTM with and without TWSA input reveals that, although the model without TWSA produces hydrographs of reasonable magnitude, the outputs exhibit numerous unrealistic fluctuations. Incorporating TWSA as an input substantially smooths the predicted hydrograph, thereby improving overall model reliability. Additional hydrographs for other gauge stations are provided in Figure S2 in Supporting Information S1. The model demonstrated strong performance at most stations, with simulated hydrographs closely aligned with observations in both pattern and magnitude. However, model predictions tended to overestimate low flows at certain stations, likely due to the use of 1-NSE as the loss function, which prioritizes peak flow fitting.

3.2. Model Performance at Grid Scale

In Section 3.1, the spatial extrapolation capability of the ConvEALSTM model was evaluated using cross-validation among gauge stations or sub-basins within the Amazon Basin. Building on the trained model from the spatial cross-validation experiment, the model's ability to generate fully distributed streamflow predictions was further assessed by analyzing simulated streamflow across all basin grid cells. Specifically, daily streamflow predictions at each grid cell were compared with the GRUN runoff and GloFAS discharge time series. Figure 3 presents the spatial distribution of correlation coefficients relative to the benchmark GRUN and GloFAS data sets. In both cases, large portions of the grids exhibit high correlations (>0.6, red shades) with clear spatial continuity, indicating strong agreement with the benchmarks and demonstrating the ConvEALSTM model's capability to generate reliable fully distributed streamflow predictions.

Moreover, Moran's I and Geary's C were employed as complementary evaluation metrics to quantify the spatial consistency of the model predictions at the grid scale. Moran's I is a product moment coefficient centered on the mean, typically ranging from −1 to 1, with positive values indicating clustering of similar values and negative values indicating spatial dispersion. In contrast, Geary's C is based on squared differences between neighboring observations and ranges from 0 to 2, where values near 0 indicate strong spatial similarity and values greater than 1 indicate spatial dissimilarity. The Moran's I values for the correlation maps between the model outputs and the GRUN and GloFAS data sets are 0.89 and 0.83, respectively, and the corresponding Geary's C values are 0.10 and

19448007, 2026, 16, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL122902 by HONG KONG POLYTECHNIC UNIVERSITY HU NG HOM, Wiley Online Library on [01/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

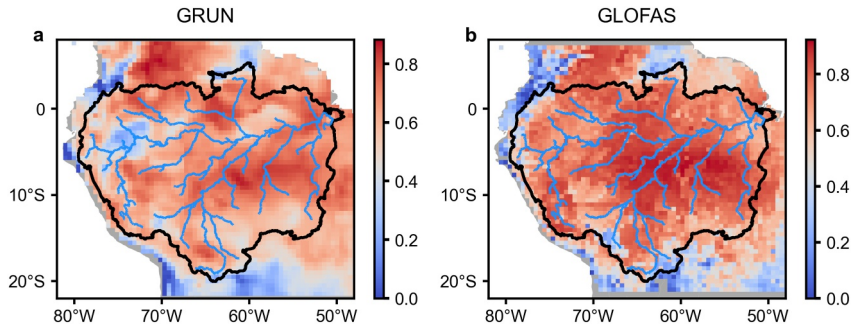


Figure 3. Spatial distribution of Pearson correlation coefficients between ConvEALSTM grid-scale predictions and benchmark data sets across the Amazon Basin. (a) Correlation with monthly GRUN runoff (temporally aggregated). (b) Correlation with daily Global Flood Awareness System streamflow.

0.16. These results indicate that the model produces reliable and spatially consistent grid-scale streamflow predictions.

3.3. Illustrative Case Study of Flow Propagation

In Sections 3.1 and 3.2, the results demonstrate that the proposed model achieves good performance in streamflow prediction at both the sub-basin and grid scale, without incorporating explicit routing modules. As an illustrative case study, the 2007 flood event was selected to demonstrate the model's implicit routing behavior. Three stations (A, B, and C) located along the same river reach were analyzed to examine the simulated flood propagation (Figure 4a). The hydrographs of the 2007 flood event simulated by the ConvEALSTM model at these stations were compared (Figure 4b), and the flood peak discharge and corresponding timing at each station were annotated. The predicted hydrographs clearly capture the flow propagation pattern, with peak discharge increasing and peak timing progressively lagging from upstream to downstream.

For quantitative evaluation of flow propagation in the model predictions, the lag time of peak flow between gauge C and gauges A and B was compared using both modeled and observed streamflow data (Figures 4c and 4e). In the observations, the lag times are 22 and 8 days, respectively, whereas in the model predictions they are 20 and 7 days, which are broadly consistent with the observed values. To further quantify the temporal dependence between upstream and downstream gauges, Pearson correlation coefficients were calculated between gauges A and B and gauge C at time lags of 0, 7, 14, 21, and 28 days to evaluate lag correlations among the three stations (Figures 4d and 4f). In the observations, the highest correlations occur at 21 and 7 days, consistent with the peak lag times. In the model predictions, the highest correlation between gauges B and C also occurs at 7 days, while that between gauges A and C occurs at 14 days, differing by one time step from the observed value. Overall, this case study shows that the ConvEALSTM model is capable of reproducing the observed flow propagation pattern for this representative flood event despite the absence of an explicit routing module. The hydrographs and Pearson correlation coefficients at multiple time lags generated by the other models are provided in Figure S3 in Supporting Information S1. Among all evaluated models, ConvEALSTM demonstrates a peak discharge lag time and flow propagation pattern that align most closely with observation.

3.4. Interpretative Analysis of Flow Propagation

To better understand the underlying mechanism of model behavior in flow propagation during this flood event, the Integrated Gradients (IG) method is applied (Sundararajan et al., 2017). This gradient-based attribution technique measures the influence of each input feature on the model output, quantifying the contributions of variables at three selected stations. Integrated Gradients maps of dynamic forcing variables are generated for the three stations over a 7-day period spanning the peak at the downstream station and the peak at the upstream station. Comparisons are conducted among IG maps derived from models that (a) include TWSA as an input, (b) exclude TWSA, and (c) isolate the individual contribution of TWSA. There is a clear downstream migration of the high IG value zone from upstream to downstream in Figures 5a1–5a5, whereas no distinct spatial pattern is observed in Figures 5b1–5b5. This contrast highlights the contribution of TWSA to flow propagation. The same

19448007, 2026, 16, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL122902 by HONG KONG POLYTECHNIC UNIVERSITY HU NG HOM, Wiley Online Library on [01/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

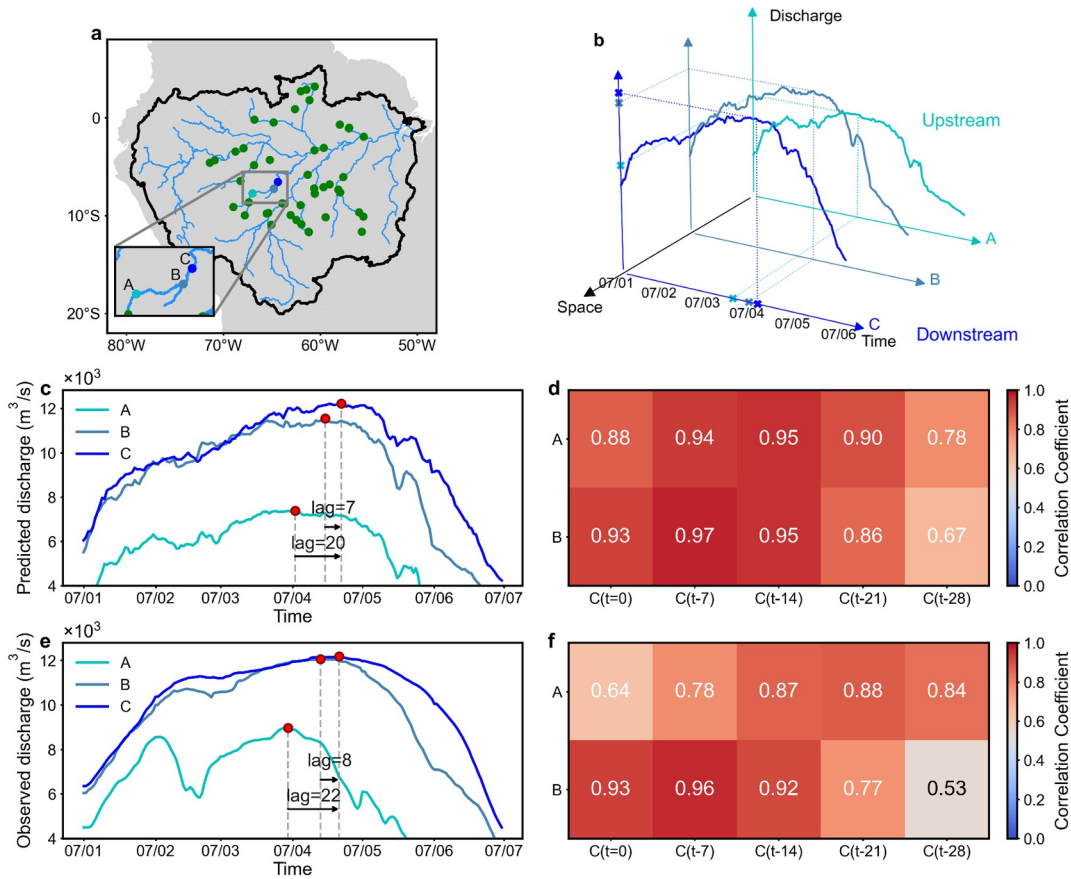


Figure 4. Exemplary spatio-temporal flood propagation. (a) Locations of upstream-to-downstream gauge stations A, B, and C. (b) Modeled hydrographs during the 2007 flood, with increasing peak magnitude and delayed timing downstream. (c, e) Observed versus modeled lag times of peak flows from A to C and B to C. (d, f) Pearson correlation coefficients at multiple time lags (0–28 days) between upstream and downstream stations in observations and model outputs.

pattern is further illustrated in Figure 5c1–5c5 and is consistent with the spatiotemporal evolution observed in Figure 5a1–5a5.

3.5. Limitations and Future Work

The Convolutional LSTM model developed for fully distributed streamflow prediction demonstrates robust performance, yet uncertainty remains a critical aspect of streamflow prediction. To reduce parameter uncertainty, model training was repeated 10 times as an ensemble, and the median values of the 10 runs were used as final results. The ranges across the 10 runs, representing parameter uncertainty at all gauge stations, are shown in Figure S2 in Supporting Information S1. The current spatial resolution of 0.5° is suitable for large basins such as the Amazon but is too coarse for smaller basins. This coarse grid can oversimplify spatial variability in terrain, land use, and soil properties, thereby introducing input-data uncertainty into predictions.

Although the model performs well without explicit physical formulations of runoff generation and routing processes, the absence of physical constraints may also contribute to structural uncertainty (e.g., the infiltration and evapotranspiration processes) in streamflow simulations. This is also supported by the inclusion of terrestrial water storage, which greatly improves model performance. Future work could integrate other physical constraints to enhance the physical realism of predictions. For instance, Wang et al. (2024) successfully embedded the Penman-Monteith equations into ML models, producing plausible estimates of potential evapotranspiration (PET). Incorporating physically-based constraint formulations into the proposed framework may enhance both interpretability and reliability of outputs. Furthermore, future research should aim to optimize the model structure to improve computational efficiency, enabling application at finer spatial resolutions. With the aid of interpretability

19448007, 2026, 16, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL122902 by HONG KONG POLYTECHNIC UNIVERSITY HU NG HOM, Wiley Online Library on [01/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

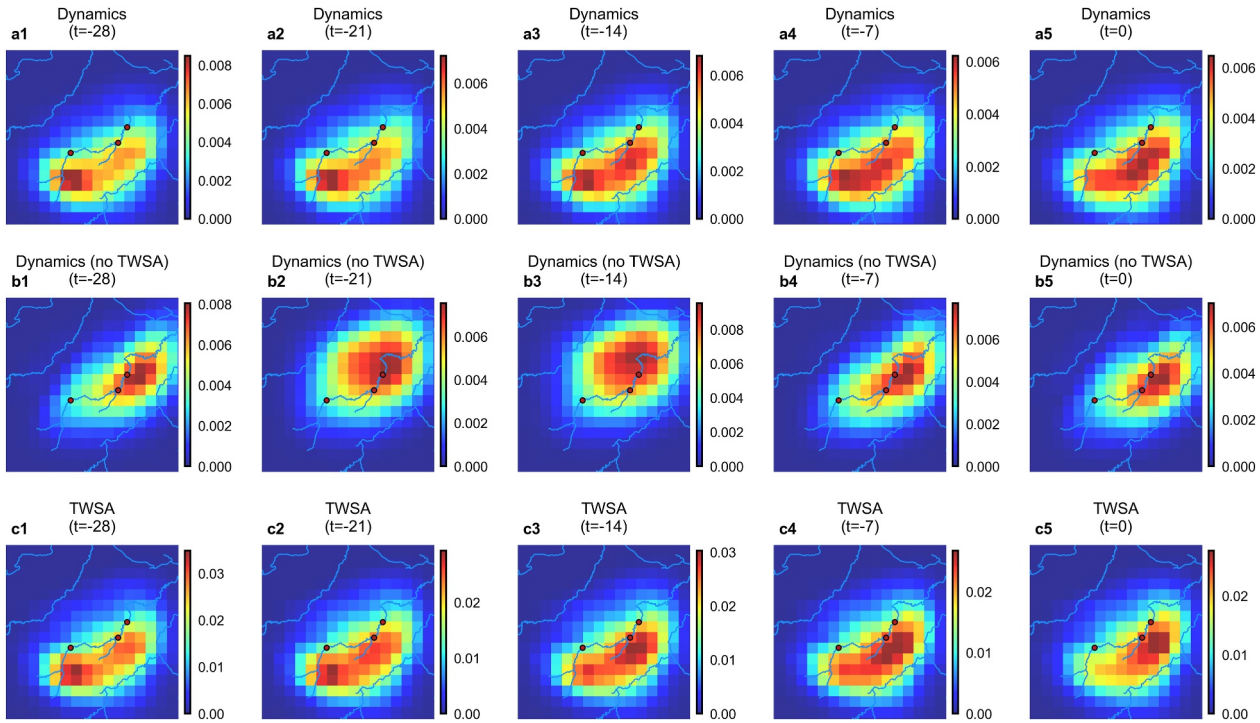


Figure 5. Spatial attribution of gridded inputs over the event period. The Integrated Gradients contribution heatmaps highlight dominant regions contributing to the streamflow prediction across different time steps.

tools such as IG maps, it is possible to gain deeper insights into how ML models operate within rainfall-runoff and flow routing processes, thereby informing model input selection and model structural improvement. These results demonstrate the feasibility of a purely ML-based approach for spatially consistent streamflow prediction, providing a scalable, data-driven alternative to routing-based hydrological models. Beyond the Amazon Basin, the ConvEALSTM framework can be applied to other large river basins worldwide, as it relies solely on commonly available meteorological forcing, globally obtainable catchment attributes, and remotely sensed terrestrial water storage data. Its fully data-driven design is particularly advantageous in regions with limited gauging infrastructure, where traditional hydrological models face substantial challenges. Future research could further evaluate the transferability of the model to other large basins.

4. Conclusion

This study developed and tested the ConvEALSTM model, a purely ML-based framework designed to generate fully distributed streamflow predictions without relying on explicit routing modules. Applied to the Amazon Basin, the model demonstrated strong predictive skill at the sub-basin scale, achieving a median NSE of 0.73 and R^2 of 0.84 in spatial cross-validation, confirming its robust performance in ungauged basins. In simulating extreme hydrological events, the model successfully constrained peak flow magnitudes, achieving a low APFB of 0.12 at the basin outlet. At the grid scale, the model produced reliable and spatially consistent outputs that correlated strongly with GRUN and GloFAS benchmark data sets (Pearson correlation >0.6 across major grid regions). The spatial autocorrelation analysis yielded high Moran's *I* values (0.89 and 0.83) and low Geary's C values (0.10 and 0.16) relative to benchmark data sets, respectively.

Although the model does not include an explicit routing module, it effectively captures flow propagation patterns by reliably predicting lag times and lag-correlations between upstream and downstream gauge stations. Further interpretative analysis indicates that the incorporation of TWSA as an additional model input makes a key contribution to the model's implicit representation of flow propagation. These findings demonstrate the capability of a pure ML approach for fully distributed streamflow prediction without physical routing modules and provide a comprehensive, quantified assessment of its performance in ungauged basins while ensuring spatial consistency.

19448007, 2026, 16, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL122902 by HONG KONG POLYTECHNIC UNIVERSITY HU NG HOM, Wiley Online Library on [01/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

The WFDEI meteorological data set is available from the ISIMIP portal (Lange & Büchner, 2020). The static attribute dataset can be accessed at Zenodo (Stacke & Hagemann, 2021). TWSA data from ITSG-Grace2018 are available at TU Graz (Mayer-Gürr et al., 2018). The source code for the ConvEALSTM model developed in this study is archived at Zenodo (Yu, 2025).

Acknowledgments
This work was supported by the National Key R&D Program of China (No. 2024YFF0808803), National Natural Science Foundation of China (No. 42471348), High-level University Special Fund (G03050K001), and Shenzhen Science and Technology Program (KCXFZ20240903093659003). Computational resources were supported by the Center for Computational Science and Engineering at the Southern University of Science and Technology. We thank Chao Wang for useful discussions at the early stage.

References

Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E., & Rasmussen, J. (1986). An introduction to the European Hydrological System—Systeme Hydrologique Europeen,"SHE", 1: History and philosophy of a physically-based, distributed modelling system. *Journal of Hydrology*, *87*(1–2), 45–59. https://doi.org/10.1016/0022-1694(86)90114-9

Aghakouchak, A., & Habib, E. (2010). Application of a conceptual hydrologic model in teaching hydrologic processes. *International Journal of Engineering Education*, *26*(4), 963–973.

Anderson, S., & Radić, V. (2022). Evaluation and interpretation of convolutional long short-term memory networks for regional hydrological modelling. *Hydrology and Earth System Sciences*, *26*(3), 795–825. https://doi.org/10.5194/hess-26-795-2022

Bassah, R., Corzo, G., Bhattacharya, B., Haider, S. M., Swain, E. D., & Aumen, N. (2025). Forecasting water levels using the ConvLSTM algorithm in the Everglades, USA. *Journal of Hydrology*, *652*, 132195. https://doi.org/10.1016/j.jhydrol.2024.132195

Beven, K., Kirkby, M., Schofield, N., & Tagg, A. (1984). Testing a physically-based flood forecasting model (TOPMODEL) for three UK catchments. *Journal of Hydrology*, *69*(1–4), 119–143. https://doi.org/10.1016/0022-1694(84)90159-8

Beven, K. J. (2012). *Rainfall-runoff modelling: The primer*. John Wiley & Sons.

Dehghani, A., Moazam, H. M. Z. H., Mortazavizadeh, F., Ranjbar, V., Mirzaei, M., Mortezavi, S., et al. (2023). Comparative evaluation of LSTM, CNN, and ConvLSTM for hourly short-term streamflow forecasting using deep learning approaches. *Ecological Informatics*, *75*, 102119. https://doi.org/10.1016/j.ecoinf.2023.102119

Devia, G. K., Ganasri, B. P., & Dwarakish, G. S. (2015). A review on hydrological models. *Aquatic Procedia*, *4*, 1001–1007. https://doi.org/10.1016/j.aqpro.2015.02.126

Feng, D., Lawson, K., & Shen, C. (2021). Mitigating prediction error of deep learning streamflow models in large data-sparse regions with ensemble modeling and soft data. *Geophysical Research Letters*, *48*(14), e2021GL092999. https://doi.org/10.1029/2021GL092999

Feng, D., Liu, J., Lawson, K., & Shen, C. (2022). Differentiable, learnable, regionalized process-based models with multiphysical outputs can approach state-of-the-art hydrologic prediction accuracy. *Water Resources Research*, *58*(10), e2022WR032404. https://doi.org/10.1029/2022WR032404

Francés, F., Vélez, J. I., & Vélez, J. J. (2007). Split-parameter structure for the automatic calibration of distributed hydrological models. *Journal of Hydrology*, *332*(1–2), 226–240. https://doi.org/10.1016/j.jhydrol.2006.06.032

Getirana, A. C., Bonnet, M.-P., Rotunno Filho, O. C., & Mansur, W. J. (2009). Improving hydrological information acquisition from DEM processing in floodplains. *Hydrological Processes: International Journal*, *23*(3), 502–514. https://doi.org/10.1002/hyp.7167

Grimaldi, S., Salamon, P., Disperati, J., Zsoter, E., Russo, C., Ramos, A., et al., & others. (2022). *River discharge and related historical data from the Global Flood Awareness System. V4.0*. European Commission, Joint Research Centre (JRC). Retrieved from https://cds.climate.copernicus.eu/

Guo, F., Yang, J., Li, H., Li, G., & Zhang, Z. (2021). A convLSTM conjunction model for groundwater level forecasting in a karst aquifer considering connectivity characteristics. *Water*, *13*(19), 2759. https://doi.org/10.3390/w13192759

Hadi, S. J., & Tombul, M. (2018). Forecasting daily streamflow for basins with different physical characteristics through data-driven methods. *Water Resources Management*, *32*(10), 3405–3422. https://doi.org/10.1007/s11269-018-1998-1

Immerzeel, W., & Droogers, P. (2008). Calibration of a distributed hydrological model based on satellite evapotranspiration. *Journal of Hydrology*, *349*(3–4), 411–424. https://doi.org/10.1016/j.jhydrol.2007.11.017

Ji, H., Chen, Y., Fang, G., Li, Z., Duan, W., & Zhang, Q. (2021). Adaptability of machine learning methods and hydrological models to discharge simulations in data-sparse glaciated watersheds. *Journal of Arid Land*, *13*(6), 549–567. https://doi.org/10.1007/s40333-021-0066-5

Kim, S., Hong, S., Joh, M., & Song, S. (2017). Deeprain: Convlstm network for precipitation prediction using multichannel radar data. *arXiv Preprint arXiv:1711.02316*. https://doi.org/10.48550/arXiv.1711.02316

Kling, H., & Gupta, H. (2009). On the development of regionalization relationships for lumped watershed models: The impact of ignoring sub-basin scale variability. *Journal of Hydrology*, *373*(3–4), 337–351. https://doi.org/10.1016/j.jhydrol.2009.04.031

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences*, *22*(11), 6005–6022. https://doi.org/10.5194/hess-22-6005-2018

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., & Nearing, G. S. (2019). Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, *55*(12), 11344–11354. https://doi.org/10.1029/2019WR026065

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences*, *23*(12), 5089–5110. https://doi.org/10.5194/hess-23-5089-2019

Kvas, A., Behzadpour, S., Ellmer, M., Klinger, B., Strasser, S., Zehentner, N., & Mayer-Gürr, T. (2019). ITSG-Grace2018: Overview and evaluation of a new GRACE-only gravity field time series. *Journal of Geophysical Research: Solid Earth*, *124*(8), 9332–9344. https://doi.org/10.1029/2019jb017415

Lange, S., & Büchner, M. (2020). ISIMIP2a atmospheric climate input data [Dataset]. *ISIMIP Repository*. https://doi.org/10.48364/ISIMIP.886955

Li, J., Hsu, K., Sorooshian, S., & Lu, D. (2025). From lumped to spatially distributed hydrologic models: A generalizable data-driven framework across North America. *Authorea Preprints*. https://doi.org/10.22541/essoar.176556008.89599411/v1

19448007, 2026, 16, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL122902 by HONG KONG POLYTECHNIC UNIVERSITY HU NG HOM, Wiley Online Library on [01/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, *99*(D7), 14415–14428. https://doi.org/10.1029/94jd00483

Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., et al. (2019). Global reconstruction of naturalized river flows at 2.94 million reaches. *Water Resources Research*, *55*(8), 6499–6516. https://doi.org/10.1029/2019WR025287

Lin, Y., Wang, D., Wang, G., Qiu, J., Long, K., Du, Y., et al. (2021). A hybrid deep learning algorithm and its application to streamflow prediction. *Journal of Hydrology*, *601*, 126636. https://doi.org/10.1016/j.jhydrol.2021.126636

Longyang, Q., Choi, S., Tennant, H., Hill, D., Ashmead, N., Neilson, B. T., et al. (2024). An attention-based explainable deep learning approach to spatially distributed hydrologic modeling of a snow dominated mountainous karst watershed. *Water Resources Research*, *60*(11), e2024WR037878. https://doi.org/10.1029/2024wr037878

Mayer-Gürr, T., Behzadpur, S., Ellmer, M., Kvas, A., Klinger, B., Strasser, S., & Zehentner, N. (2018). ITSG-grace2018—Monthly, daily and static gravity field solutions from GRACE [Dataset]. *GFZ Data Services*. https://doi.org/10.5880/ICGEM.2018.003

Meng, E., Huang, S., Huang, Q., Fang, W., Wu, L., & Wang, L. (2019). A robust method for non-stationary streamflow prediction based on improved EMD-SVM model. *Journal of Hydrology*, *568*, 462–478. https://doi.org/10.1016/j.jhydrol.2018.11.015

Muhebwa, A., Gleason, C. J., Feng, D., & Taneja, J. (2024). Improving discharge predictions in ungauged basins: Harnessing the power of disaggregated data modeling and machine learning. *Water Resources Research*, *60*(9), e2024WR037122. https://doi.org/10.1029/2024wr037122

Nijzink, R. C., Samaniego, L., Mai, J., Kumar, R., Thober, S., Zink, M., et al. (2016). The importance of topography-controlled sub-grid process heterogeneity and semi-quantitative prior constraints in distributed hydrological models. *Hydrology and Earth System Sciences*, *20*(3), 1151–1176. https://doi.org/10.5194/hess-20-1151-2016

Savenije, H. H. (2010). HESS opinions "Topography driven conceptual modelling (FLEX-Topo)". *Hydrology and Earth System Sciences*, *14*(12), 2681–2692. https://doi.org/10.5194/hess-14-2681-2010

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., & Woo, W. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems* (Vol. 28).

Singh, V. P., & Frevert, D. K. (2010). *Watershed models*. CRC Press.

Sorooshian, S. (Ed.) (2008). *Hydrological modelling and the water cycle: Coupling the atmospheric and hydrological models*. Springer.

Stacke, T., & Hagemann, S. (2021). Land surface parameter fields at 0.5 deg resolution for use with the HydroPy model (Version 1.0.0) [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.4541239

Sun, A. Y., Wang, D., & Xu, X. (2014). Monthly streamflow forecasting using Gaussian process regression. *Journal of Hydrology*, *511*, 72–81. https://doi.org/10.1016/j.jhydrol.2014.01.023

Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In *International conference on machine learning* (pp. 3319–3328).

Troin, M., Arsenault, R., Wood, A. W., Brissette, F., & Martel, J. (2021). Generating ensemble streamflow forecasts: A review of methods and approaches over the past 40 years. *Water Resources Research*, *57*(7), e2020WR028392. https://doi.org/10.1029/2020WR028392

Wang, C., Jiang, S., Zheng, Y., Han, F., Kumar, R., Rakovec, O., & Li, S. (2024). Distributed hydrological modeling with physics-encoded deep learning: A general framework and its application in the Amazon. *Water Resources Research*, *60*(4), e2023WR036170. https://doi.org/10.1029/2023WR036170

Wunsch, A., Liesch, T., Cinkus, G., Ravbar, N., Chen, Z., Mazzilli, N., et al. (2022). Karst spring discharge modeling based on deep learning using spatially distributed input data. *Hydrology and Earth System Sciences*, *26*(9), 2405–2430. https://doi.org/10.5194/hess-26-2405-2022

Xiang, Z., & Demir, I. (2022). Fully distributed rainfall-runoff modeling using spatial-temporal graph neural network. *Civil and Environmental Engineering*. https://doi.org/10.31223/X57P74

Yang, Y., Feng, D., Beck, H. E., Hu, W., Abbas, A., Sengupta, A., et al. (2025). Global daily discharge estimation based on grid long short-term memory (LSTM) model and river routing. *Water Resources Research*, *61*(6), e2024WR039764. https://doi.org/10.1029/2024WR039764

Yang, Y., Pan, M., Beck, H. E., Fisher, C. K., Beighley, R. E., Kao, S., et al. (2019). Quest of calibration density and consistency in hydrologic modeling: Distributed parameter calibration against streamflow characteristics. *Water Resources Research*, *55*(9), 7784–7803. https://doi.org/10.1029/2018WR024178

Yaseen, Z. M., Kisi, O., & Demir, V. (2016). Enhancing long-term streamflow forecasting and predicting using periodicity data component: Application of artificial intelligence. *Water Resources Management*, *30*(12), 4125–4151. https://doi.org/10.1007/s11269-016-1408-5

Yu, Q. (2025). Convolutional entity-aware LSTM model for fully distributed streamflow prediction (Version 1.0) [Computer software]. *Zenodo*. https://doi.org/10.5281/zenodo.17061805

Yu, Q., Jiang, L., Schneider, R., Zheng, Y., & Liu, J. (2024). Deciphering the mechanism of better predictions of regional LSTM models in ungauged basins. *Water Resources Research*, *60*(7), e2023WR035876. https://doi.org/10.1029/2023WR035876

Yu, Q., Tolson, B. A., Shen, H., Han, M., Mai, J., & Lin, J. (2024). Enhancing long short-term memory (LSTM)-based streamflow prediction with a spatially distributed approach. *Hydrology and Earth System Sciences*, *28*(9), 2107–2122. https://doi.org/10.5194/hess-28-2107-2024

Zhang, Y., Zheng, H., Chiew, F. H., Peña-Arancibia, J., & Zhou, X. (2016). Evaluating regional and global hydrological models against streamflow and evapotranspiration measurements. *Journal of Hydrometeorology*, *17*(3), 995–1010. https://doi.org/10.1175/jhm-d-15-0107.1

19448007, 2026, 16, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL122902 by HONG KONG POLYTECHNIC UNIVERSITY HU NG HOM, Wiley Online Library on [01/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License