https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

Bakaano-Hydro (v1.1). A distributed hydrology-guided deep learning model for streamflow prediction

Confidence Duku[1,2]

[1]Climate Resilience Team, Wageningen Environmental Research, Wageningen University & Research, Wageningen, The Netherlands
[2]Earth Systems and Global Change, Wageningen University & Research, P.O. Box 47, 6700AA, Wageningen, The Netherlands

https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

Abstract

Reliable streamflow prediction is fundamental to hydrological forecasting, water resources planning, and climate adaptation. However, existing data-driven approaches often lack physical interpretability and struggle to incorporate spatial heterogeneity and hydrological connectivity. Conversely, traditional process-based models are limited by high calibration demands and structural uncertainty, especially in data-scarce regions. These challenges underscore the need for hybrid frameworks that combine the strengths of physically based modeling with the predictive capacity of machine learning. Here, I present Bakaano-Hydro, a distributed hydrology-guided deep learning model for streamflow prediction. The model integrates a gridded runoff generation method, a topographic flow routing scheme, and a temporal convolutional network to capture both spatial and temporal hydrological dynamics. This architecture enables incorporation of spatial heterogeneity and explicitly represents hydrological connectivity, while using neural networks to learn streamflow dynamics and enhance predictive performance. Bakaano-Hydro's performance is evaluated across six river basins spanning four continents, encompassing diverse climate zones, land-use patterns, and hydrological regimes. Results indicate that Bakaano-Hydro demonstrates robust performance in humid and snow-fed basins where saturation-excess runoff dominates, while revealing key limitations in arid and semi-arid regions characterized by infiltration-excess processes. Bakaano-Hydro advances the state of the art in data-driven hydrological modeling by integrating physical realism with deep learning. Its modular and fully automated pipeline enables rapid deployment in data-scarce regions, while maintaining high reliability and interpretability. These features make Bakaano-Hydro a promising tool for operational forecasting, climate risk assessment, and adaptation planning across diverse hydrological and socio-environmental contexts. The model code is publicly available at https://github.com/confidence-duku/bakaano-hydro to facilitate reproducibility and community-driven development.

1. Introduction

Streamflow prediction is pivotal for meeting societal needs and underpins various critical areas including water resources management, hydropower, flood management, infrastructure planning and climate adaptation (Palmer and Ruhi, 2019; Depetris, 2021; Herrera et al., 2017; Cassagnole et al., 2021; Golden et al., 2025). Traditionally, physically based hydrological models have been the primary tool for simulating streamflow, relying on process-based equations to describe hydrological processes (e.g. Arnold et al., 1998; Hamman et al., 2018; Schaphoff et al., 2018; De Roo et al., 2000). However, these models require extensive calibration, struggle with parameter uncertainty in data-scarce regions, and often exhibit biases due to

https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

structural limitations (Herrera et al., 2022; Moges et al., 2021). Recently, deep learning models have emerged as powerful alternatives, offering data-driven solutions that can learn complex streamflow-generation relationships directly from observational data. In particular, Long Short-Term Memory (LSTM) networks have demonstrated superior predictive performance compared to many conceptual and physically based models (Hunt et al., 2022; Kratzert et al., 2019; Kratzert et al., 2018; Gauch et al., 2021; Nearing et al., 2024; Arsenault et al., 2023). These LSTM-based models employ lumped modelling approaches where an entire watershed is treated as a single computational unit. Meteorological forcings, soil, topographic and land use data are aggregated or area-weighted over a catchment, and the model outputs streamflow at the outlet. Architectural constraints of available sequence-based neural network algorithms including LSTMs drive the adoption of lumped modelling approach. While these LSTMs are effective for capturing temporal dependencies in hydrological time series, they are not inherently designed to process gridded or spatially distributed data. As a result, these LSTM-based lumped modelling approaches, which are the state-of-the art in data-driven hydrological modelling face critical limitations, despite their demonstrated predictive performance. A primary concern is their inability to capture spatial heterogeneity, particularly in large or complex basins where sub-basin variability in precipitation, land cover, soil moisture, and terrain strongly influences runoff generation. Consequently, the predictive skill of lumped data-driven models typically declines with increasing basin size (Hunt et al., 2022). Studies have shown that incorporating spatial variability, particularly in rainfall inputs, can enhance the predictive performance of lumped data-driven approaches (Wang and Karimi, 2022). Another major limitation is the inability to represent upstream-downstream hydrological connectivity, a key driver of streamflow, flood dynamics, water availability, and ecosystem stability. Without explicit representation of hydrological connectivity, these models struggle to simulate localized hydrological responses, flow propagation, and the impacts of land-use change on water resources, making them less applicable for flood forecasting, water resource management, and ecosystem conservation.

In recent years, two types of approaches have emerged to integrate spatial heterogeneity into data-driven streamflow prediction frameworks: hybrid process-based and data-driven models, and the use of convolutional neural network. For example, Yu et al. (2024) employed a spatially recursive hybrid approach that first trains lumped regional LSTM model to predict local streamflow at a subbasin outlet and subsequently uses the predicted streamflow as input to a physics-based hydrological routing model to predict streamflow at the larger basin outlet. This approach, while incorporating elements of spatial variability, cannot explicitly model spatially varying hydrological processes such as runoff generation across the basin. Anderson and Radić (2022), on the other hand, employed a purely data-driven approach

https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

involving the use of convolutional long-short term memory (ConvLSTM) architectures to predict streamflow. ConvLSTMs are essentially spatiotemporal neural networks that improve upon standard LSTMs by incorporating spatial dependencies through convolutional operations (Shi et al., 2015). While this architecture is well-suited for spatiotemporal sequence modeling, its application in hydrological contexts remains limited. In particular, ConvLSTMs do not inherently respect the directional and hierarchical nature of river networks governed by topography, and thus cannot enforce hydrologically consistent flow accumulation and routing.

Consequently, despite progress in developing methods to better capture spatial heterogeneity in data-driven hydrological modeling, existing approaches remain inadequate for fully distributed applications that require representation of upstream–downstream connectivity and physically consistent flow propagation. To address these limitations, I introduce Bakaano-Hydro, a distributed hydrology-guided deep learning model for streamflow simulation. The term 'Bakaano' originates from a native Ghanaian language and refers to river-side communities, who are seasonally at risk of fluvial flooding, a primary motivation for the development of this tool. Bakaano-Hydro employs a serial hybridization approach and integrates a gridded process-based rainfall-runoff method that captures spatial heterogeneity and dynamic interactions of meteorological forcings and physiographic attributes generating spatially distributed runoff estimates; a flow routing method propagating runoff through the river network based on topographic constraints to preserve hydrological connectivity.; and a sequential neural network that uses routed flow sequences extracted at any point along a river network to predict streamflow. This approach ensures that primary hydrological responses to climate, soil, topography, and vegetation interactions and changes are captured by process-based components, enhancing interpretability while leveraging deep learning for pattern recognition. I demonstrate the applicability, robustness and generalizability of Bakaano-Hydro in six case-study basins of varying sizes, spanning four continents, hydroclimatic gradients and land-use patterns.

2. Bakaano-Hydro model structure

Bakaano-Hydro integrates a gridded runoff generation method, a topographic flow routing scheme, and a temporal convolutional network to capture both spatial heterogeneity, hydrological connectivity and temporal dynamics (Fig. 1). Bakaano-Hydro is scalable and operates at a daily time-step and at the spatial resolution of the elevation data.

https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

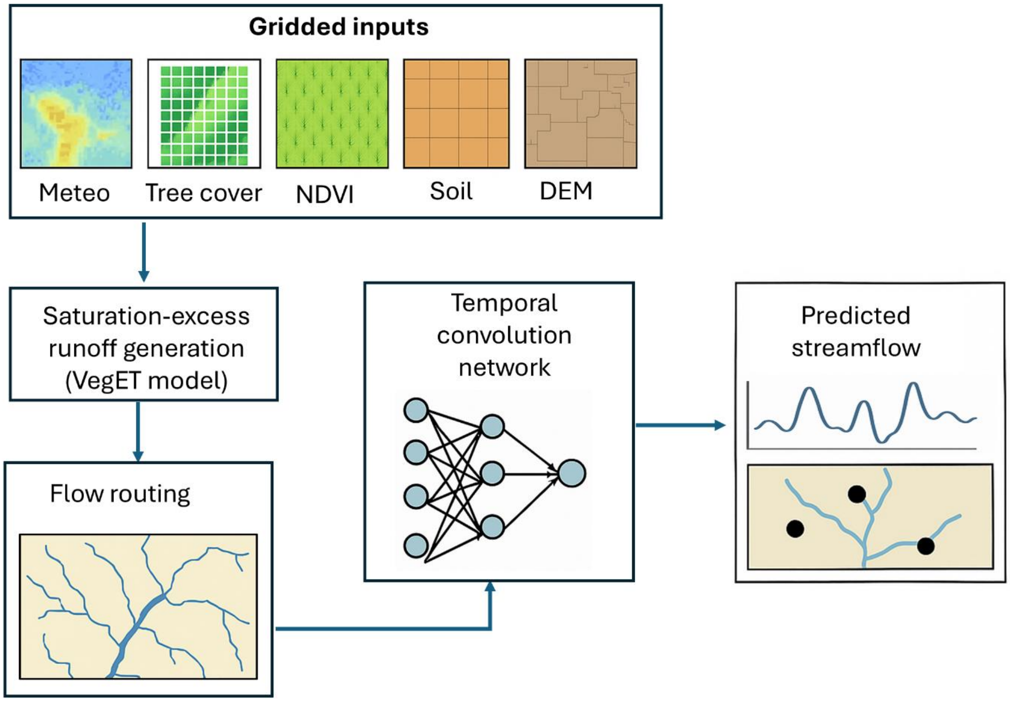

Figure 1. Conceptual diagram of the Bakaano-Hydro model. Meteo is meteorological variables; NDVI is Normalised Difference Vegetation Index and DEM is Digital Elevation Model.

2.1. Runoff generation with a process-based method

The runoff generation phase in Bakaano-Hydro is designed to simulate spatiotemporal total runoff as a function of climate, topography, vegetation, and soil interactions. Notably, runoff serves as an intermediate variable in the model, with the primary focus on accurately representing its spatiotemporal variability rather than capturing absolute runoff values. Runoff generation is based on the Vegetation ET (VegET) method (Senay, 2008; Senay et al., 2023). VegET operates on gridded data and follows a one-dimensional soil water balance framework, where each grid cell is modeled independently without explicit lateral water flow. This gridded structure allows for seamless integration with meteorological datasets and remote sensing products. Precipitation is partitioned into canopy interception, evapotranspiration, soil moisture storage, and total runoff. A detailed description of VegET can be found in (Senay et al., 2023).

Interception is explicitly parameterized based on tree, herbaceous, and bare cover fractions. The interception storage capacity varies dynamically with vegetation cover, ensuring that dense forested areas

https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

intercept more rainfall than sparsely vegetated or bare soil regions. Actual evapotranspiration is estimated as a function of the Normalized Difference Vegetation Index (NDVI) and reference evapotranspiration (PET). NDVI, acts as a proxy for vegetation density and health, allowing the model to dynamically adjust transpiration rates based on vegetation cover. In areas with abundant vegetation and sufficient soil moisture, actual evapotranspiration approaches potential evapotranspiration, while in water-limited regions, actual evapotranspiration is constrained by soil water availability. In Bakaano-Hydro, the Hargreaves equation is used to estimate potential evapotranspiration (Hargreaves and Samani, 1985). In VegET, soil water content is estimated as a function of soil properties such as wilting point, field capacity, and saturation. Finally, runoff generation is estimated based on saturation-excess mechanism, where precipitation that exceeds the soil's storage capacity contributes to runoff.

2.2. Topographic runoff routing through river network

The flow routing phase is pivotal to Bakaano-Hydro, bridging process-based runoff generation with data-driven streamflow prediction. Daily total runoff generated by VegET method is routed to the river channel network using the weighted flow accumulation method based on the multiple flow direction (MFD) approach (Quinn et al., 1991). The MFD routing scheme distributes flow from each cell to up to eight neighboring cells, with partitioning proportional to the elevation gradient between cells. For each study domain or delineated basin, the resulting routed runoff is used to construct time-series inputs corresponding to the spatial locations of hydrological observation stations. To account for potential misalignments between observed station coordinates and the digital river network, a **coordinate snapping** procedure is implemented. This step ensures that station coordinates are matched to the nearest river segment, thereby minimizing spatial discrepancies and enhancing the reliability of extracted runoff signals.

It is important to note that this initial routing implementation does not explicitly account for hydrological processes such as transmission losses, overbank flow, or in-channel travel time delays, which can influence both the magnitude and timing of streamflow, particularly in arid and semi-arid regions or large river systems. These processes are intentionally omitted at this stage, as the primary objective is to simulate maximum potential daily flow at each hydrological point. These routed runoff time-series serve as inputs to the subsequent deep learning stage, where temporal dependencies and additional hydrological dynamics are learned directly from data.

https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

2.3 Streamflow simulation with neural network

Bakaano-Hydro employs Temporal Convolutional Networks (TCN) (Bai et al., 2018), attention mechanisms (Luong et al., 2015), and Feature-wise Linear Modulation (FiLM) conditioning (Perez et al., 2018), to capture the complex hydro-climatic dynamics governing streamflow (Fig.2). Importantly, Bakaano-Hydro is designed as a single multitask model that can be trained on observations across a wide range of spatial domains—from individual river basins to continental or even global extents. This enables the model to learn shared representations across regions, improving its capacity for regional generalization and zero-shot inference in ungauged or poorly monitored catchments. Such a design supports scalable streamflow prediction and offers a unified approach for simulating hydrological dynamics across data-rich and data-scarce regions alike.

2.3.1 Input representation, preprocessing and feature engineering

Bakaano-Hydro provides two architectural options for the neural network component. First, a two-input branch neural network design that receives dynamic and static input data. In this design, the dynamic input branch receives three features and the static branch, ten features. The dynamic input branch receives routed runoff sequences extracted from a specified hydrological station or point along a river network; the routed runoff sequences scaled by the upstream contributing area (i.e. number of upstream grid cells); and also scaled by the depth-to-water index, a topographic metric reflecting subsurface hydrological storage. The last two features facilitate generalization across catchments of varying scales and response times. To ensure numerical stability during training, a quantile transformation is applied to the dynamic input features, mapping them into percentile space using the empirical cumulative distribution function. This transformation reduces sensitivity to extreme values and ensures the model learns from the full range of hydrometeorological variability. The resulting normalized dynamic predictors are structured as a time series over a fixed lookback period (i.e. 365 days), where streamflow for day t depends on this lookback period allowing the model to learn temporal dependencies relevant to streamflow generation. The quantile transformation is implemented globally within a basin i.e. on concatenated data from all stations together.

The static input branch comprises physiographic attributes representing watershed-scale controls on runoff and routing. These variables, including slope, soil water holding capacity, soil saturation point, fractional tree and herbaceous cover, are computed using a weighted flow accumulation approach and normalized by the number of contributing grid cells at a hydrological station or specified point along a river network. Additionally, sine and cosine transformations of latitude and longitude serve as spatial encodings.

https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

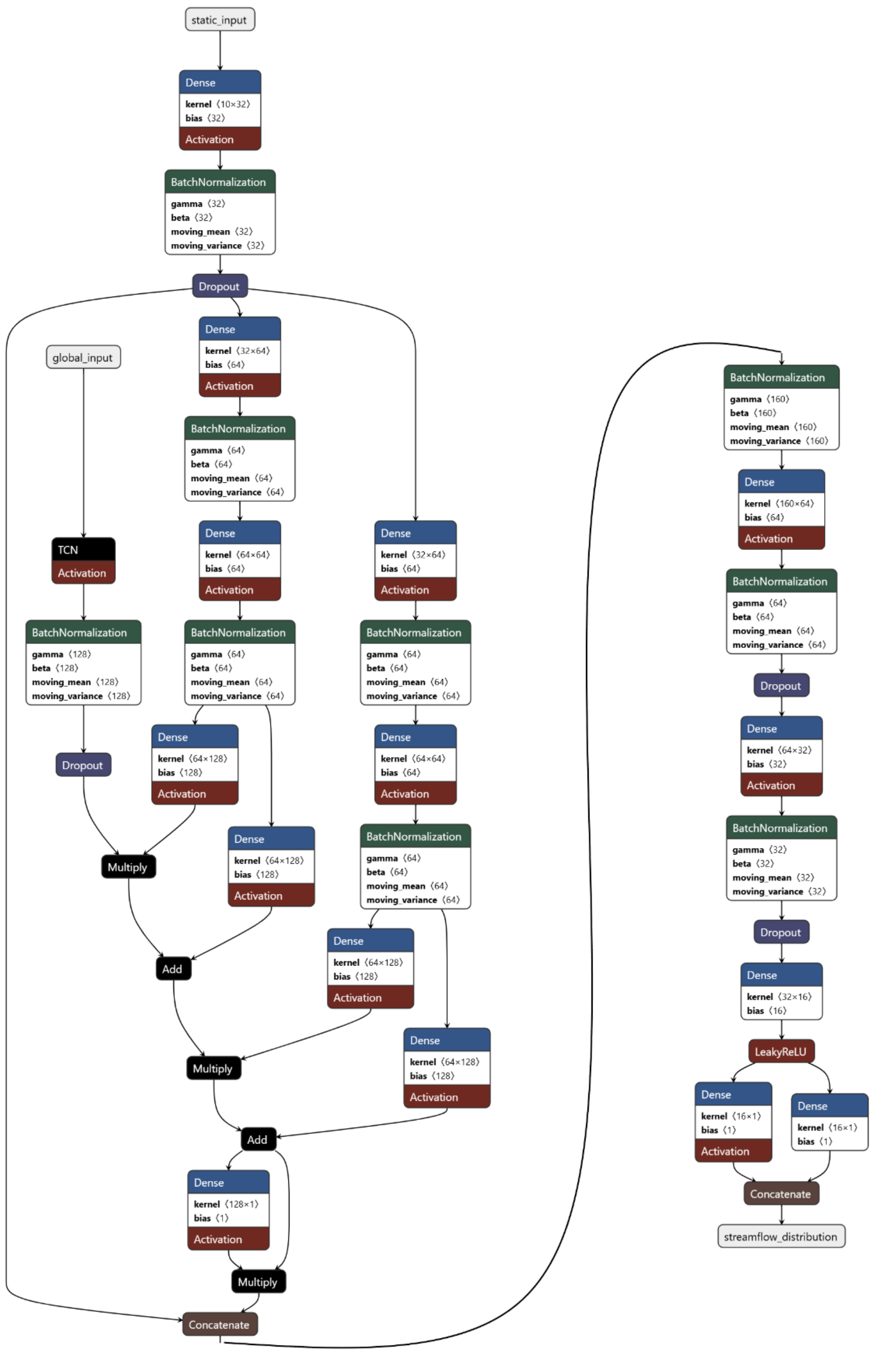

Figure 2. Architecture of Bakaano-Hydro neural network component illustrating the two input-branch design.

https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

Features in the static input branch are scaled using the min-max scaling method. This normalization method transforms each features into a range of [0,1], preserving the relative relationships between different catchment attributes while ensuring that all features contribute proportionally to the learning process.

The second option for the neural network component is the three-input branch design, which includes a second dynamic input branch in addition to the aforementioned two. This second dynamic input branch comprises only 1 feature, the routed runoff sequences scaled by the upstream contributing area. Unlike the first dynamic input branch, quantile transformation is applied independently for each hydrological station.

2.3.2 Capturing temporal dependencies

Unlike traditional recurrent architectures such as LSTMs, which process sequences sequentially, TCNs leverage dilated causal convolutions, which allows for efficient parallel computation and stable gradient propagation. TCNs also capture multi-scale temporal dependencies, e.g. short-term precipitation-driven runoff and longer-term hydrological memory effects, preserving seasonal and interannual variability in streamflow (Bai et al., 2018). In Bakaano-Hydro, a single TCN layer with dilation rates of 2,4,8,16,32,64,128 and 256 are employed. Batch normalization is used after TCN layer and mitigates the impact of skewed streamflow distributions, while a dropout rate of 0.4 promotes generalization by reducing overfitting to the training data (Ioffe and Szegedy, 2015; Srivastava et al., 2014).

In addition, attention mechanisms (Luong et al., 2015) are integrated in the architecture to dynamically assign importance to the TCN outputs, ensuring that the most hydrologically relevant information is emphasized in the prediction process. Standard convolutional and recurrent architectures process all time steps with equal weight, which can dilute critical signals. The attention mechanism addresses this issue by adaptively prioritizing key temporal features based on their contribution to streamflow response. Attention scores serve as dynamic weights that scale the original TCN outputs, effectively enhancing influential hydrological signals while suppressing less relevant temporal features.

2.3.3 Feature-wise linear modulation for contextualization

Alongside the temporal dependencies captured by the TCN, the model incorporates a dense network to process static catchment attributes, ensuring physiographic and geospatial characteristics inform streamflow predictions. To effectively integrate static catchment attributes into the temporal modeling of

https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

streamflow, the model employs FiLM. FiLM allows the network to adapt its learned representations of hydrological dynamics in response to catchment-specific properties (Perez et al., 2018). FiLM applies two modulation parameters to the TCN's output feature representations: a scaling factor and a shifting factor. The scaling factor adjusts the strength of the extracted temporal features, while the shifting factor modifies the baseline streamflow response, ensuring that the model accounts for regional hydrological differences. FiLM learns these modulation parameters through a multi-layer perceptron that maps static catchment descriptors to the learned transformation space. The scaling and shifting transformations allow the model to condition its temporal feature extraction on catchment attributes, ensuring that the learned representations capture regional differences in hydrological behavior.

3. Bakaano-Hydro code architecture and workflow

Bakaano-Hydro follows an object-oriented architecture designed for modularity, scalability, and flexibility. The implementation is structured across multiple Python modules, each responsible for specific tasks within the hydrological simulation pipeline (Fig.3). The main module (*runner.py*) serves as the central workforce, managing gridded runoff simulation, topographic flow routing, sequential neural network training, and streamflow simulation. Data acquisition and preprocessing are executed separately from the main hydrological modeling components to ensure data reusability. Modules responsible for data retrieval and preprocessing include *meteo.py*, *tree_cover.py*, *dem.py*, *soil.py*, *ndvi.py,* and *utils.py*. These modules handle the downloading and processing of environmental variables such as meteorological data, vegetation indices, topography, and soil properties. The *meteo.py* module handles automated retrieval and preprocessing of gridded daily precipitation, maximum, minimum and mean temperature. Bakaano-Hydro allows for automated retrieval of these meteorological variables from three sources; ERA5-land (Muñoz-Sabater et al., 2021), CHIRPS (Funk et al., 2015) and the CHELSA-W5E5 database (Karger et al., 2023; Karger, 2021) in addition to the option for users to provide their own meteorological data. CHIRPS provides only daily precipitation data, as a result, temperature data are downloaded from ERA5-land if this option is selected. Unlike ERA5-land and CHIRPS, data availability from CHELSA is limited to the period 1981–2016. The *ndvi.py* module handles automated retrieval and preprocessing of NDVI data from MODIS (Didan, 2021; Dimiceli et al., 2015). The NDVI data are provided at 16-day intervals.

Following the VegET procedure, a daily mean climatology of NDVI for a specified period is established with linear interpolation (Senay et al., 2023). Soil properties, including wilting point, field capacity, and saturation, are retrieved from International Soil Reference and Information Center's SoilGrid database

https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

(Poggio et al., 2021) using the *soil.py* module. DEM is sourced from HydroSHEDS (Lehner et al., 2008), providing hydrologically corrected topographic data. Users can also provide their own elevation data. Fractional tree and herbaceous cover at annual time-step are retrieved from the MODIS (Dimiceli et al., 2015) using the *tree_cover.py* module. Because Bakaano-Hydro operates at the spatial resolution of the elevation data, all input data are regridded to match the DEM data. The separation of data handling from model execution allows flexibility in using the processed datasets for multiple modeling scenarios without redundant computations. Hydrological simulations are managed through distinct process-based components.

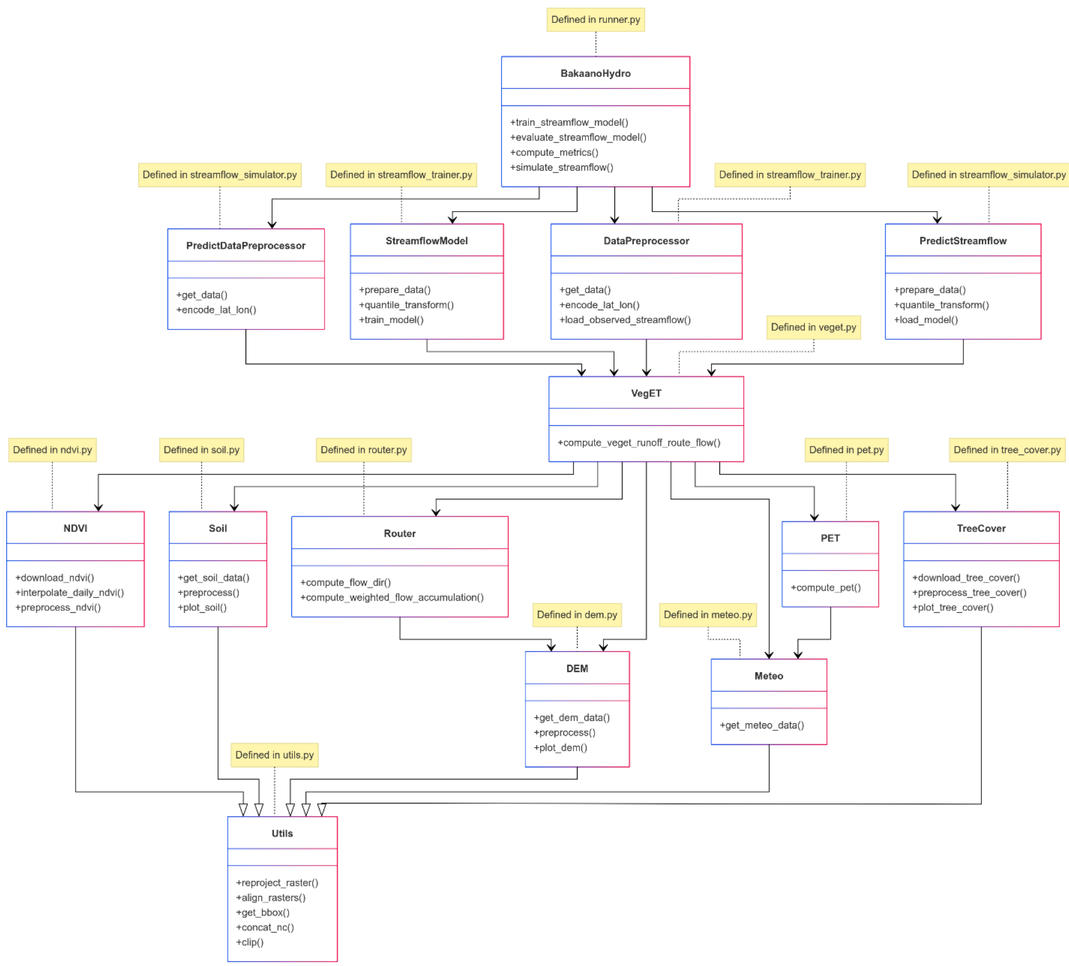

Figure 3. Bakaano-Hydro code structure, showing the modular design and interdependencies across components. Modules are grouped by functionality, with each class defined in a separate Python script.

https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

The *veget.py* module implements the VegET method for runoff generation. Flow routing is handled by *router.py*, ensuring the representation of hydrological connectivity across upstream and downstream regions. Training and simulation are executed within *streamflow_trainer.py* for model optimization and *streamflow_simulator.py* for hydrological predictions. The modular design of Bakaano-Hydro ensures that different components can be updated or replaced independently, promoting extensibility and interoperability with various modeling approaches.

Deploying Bakaano-Hydro requires three primary data or inputs from a user. First, a shapefile of the river basin or study area. Second, registration at Google Earth Engine (https://code.earthengine.google.com/register). Bakaano-Hydro retrieves, NDVI, tree cover, herbaceous cover and meteorological variables from ERA5-land or CHIRPS from Google Earth Engine Data Catalog. This platform requires prior registration for subsequent authentication during execution of the model. Finally, observed streamflow data in NetCDF format from Global Runoff Data Center (GRDC) (GRDC, 2025) . Because Bakaano-Hydro aims to use only open-source data, it currently accepts observed streamflow data only from GRDC. Subsequently, training and using the model for simulations consists of five steps. First, data retrieval and preprocessing. This involves regridding and reprojection where necessary. Second, runoff generation with VegET and subsequent routing through a river network. Third, training a model instance based on specified training period. Fourth, evaluating the trained model on out-of-sample data. Finally, applying the trained and evaluated model for subsequent simulation, scenario analysis etc.

4. Application examples and diagnostic evaluation

To evaluate the robustness and generalizability of the model, Bakaano-Hydro was deployed across six river basins of varying sizes spanning four continents and representing a broad range of hydroclimatic, ecological and land use conditions. They include the Amazon, Niger, Orange, Missouri, Upper Mississippi, and Rhine basins (Fig. 4). These basins encompass highly contrasting environmental regimes, providing a rigorous testbed for model performance.

4.1. Study areas

Figure 4 shows the location of the river basins and the hydrological stations used in training and evaluating the model in each basin. The Amazon basin, with an area of approximately 7,000,000 km^2, is characterized by humid tropical rainforest with year-round precipitation and high runoff. It represents a perennial, energy-limited hydrological regime with minimal human regulation. The Niger river basin, with an area of

https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

about 2,000,000 km^2, is a semi-arid to sub-humid system with strong rainfall seasonality, delayed runoff, and substantial variability in flow timing and volume. It is particularly relevant for assessing model performance under seasonal water scarcity and lagged hydrological responses. The Orange river basin, with an area of approximately 973,000km^2, is a predominantly semi-arid system with low and highly variable rainfall, frequent dry spells, and episodic high-intensity rainfall events that generate sharp streamflow peaks. The basin is heavily regulated, and its runoff dynamics are influenced by both climatic extremes and water management infrastructure, making it a useful test case for model performance under arid and drought-prone conditions. Missouri river basin, with an area of about 1,400,000 km^2, is a snow-dominated basin originating in the Rocky Mountains and flowing through semi-arid to temperate plains.

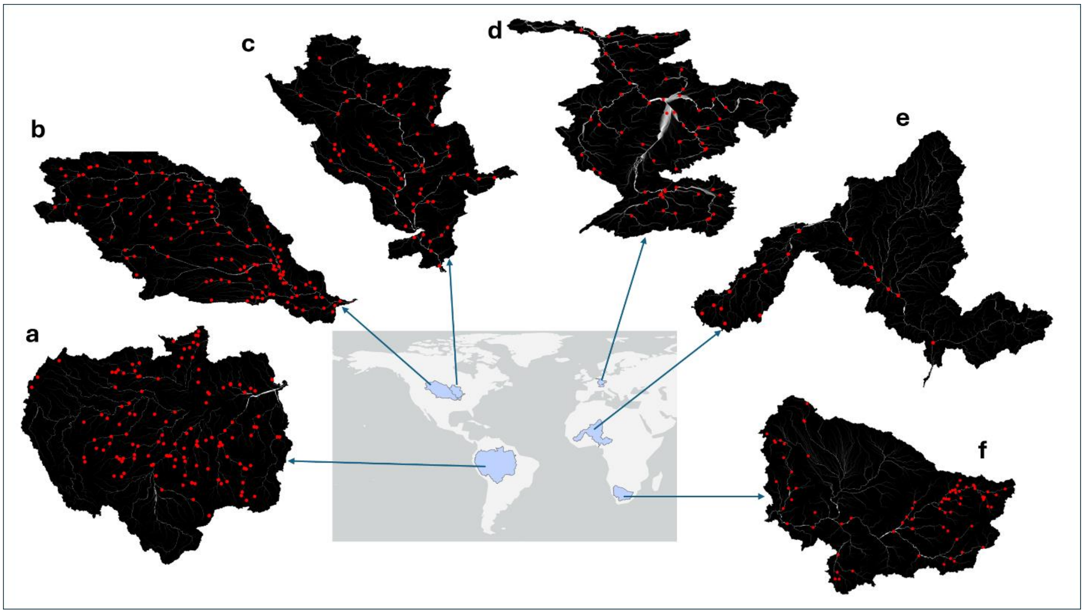

Figure 4. Maps showing the location of the river basins and the hydrological stations used in training and evaluating model performance: **a)** is Amazon river basin; **b)** is Missouri basin; **c)** is Upper Mississippi basin; **d)** is Rhine basin; **e)** is Niger basin and **f)** is Orange basin.

Upper Mississippi Basin, with an area of about 500,000 km^2, is a mixed snow and rain-fed basin with temperate climate, extensive agricultural land use, and moderate hydrological regulation. It offers a contrasting test case for evaluating model performance in intensively managed, mid-latitude catchments. Rhine Basin, with an area of approximately 185,000km^2, is a temperate, well-instrumented basin with both rain and snowmelt contributions, heavily influenced by urbanization and water management

https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

infrastructure. It represents a densely monitored system with well-known anthropogenic impacts. Together, these basins cover a spectrum of climate zones (humid tropical to semi-arid and temperate), flow regimes (perennial, seasonal, snowmelt-dominated, and regulated), and land cover types (forest, savanna, cropland, and urbanized areas).

4.2. Model configuration

For each basin, the relevant Bakaano-Hydro modules were used to automatically download and preprocess input data. Meteorological data was obtained from the CHELSA-W5E5 database for all basins. The analysis period was set from 1981 to 2016, aligning with the availability of the CHELSA-W5E5 data, which extends until 2016. Observed streamflow data were obtained from the GRDC (GRDC, 2025). The spatial resolution for all analysis was 1km^2. Hydrological stations were included in the analysis only if they had at least three years of observed records within the study period and a catchment area greater than 1000 km^2. Based on this criteria, 19 stations were selected for the Niger, 64 for the Upper Mississippi, 68 for the Rhine, and 131 for the Amazon, 53 for the Orange basin and 115 stations for the Missouri basin. Observed streamflow records for each station were split into training (1989–2016) and independent evaluation (1982–1988) periods. Due to variations in record lengths among stations, the number of stations available for training and evaluation differed from the total number initially selected. The three-input branch architecture was used for the Niger basin because it provided the best performance. For the rest of the basins, the two-input branch configuration provided the best performance.

4.3 Performance and diagnostic evaluation

Model performance across hydroclimatic gradients was evaluated using a suite of diagnostic metrics that characterize different aspects of streamflow simulation. The raincloud plots (Fig.5) reveal distinct patterns of model behavior across the six selected basins, reflecting how climatic and hydrological conditions influence predictive skill. Overall performance was assessed using the Nash–Sutcliffe Efficiency (NSE) and the Kling–Gupta Efficiency (KGE). These metrics evaluate the ability of the model to capture the magnitude and timing of streamflow. In humid and temperate basins such as the Amazon, Rhine, and Upper Mississippi, both NSE and KGE values were high and tightly distributed across stations, indicating strong and stable model performance (Fig.5). The Niger basin, though semi-arid to sub-humid, also showed consistently positive NSE and KGE values with limited spread, suggesting robust performance under seasonally dry conditions. In contrast, broader variability and lower median values were observed in the

https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

Missouri and Orange basins, where snowmelt processes, episodic rainfall, and regulation introduce greater hydrological complexity and less predictable flow regimes.

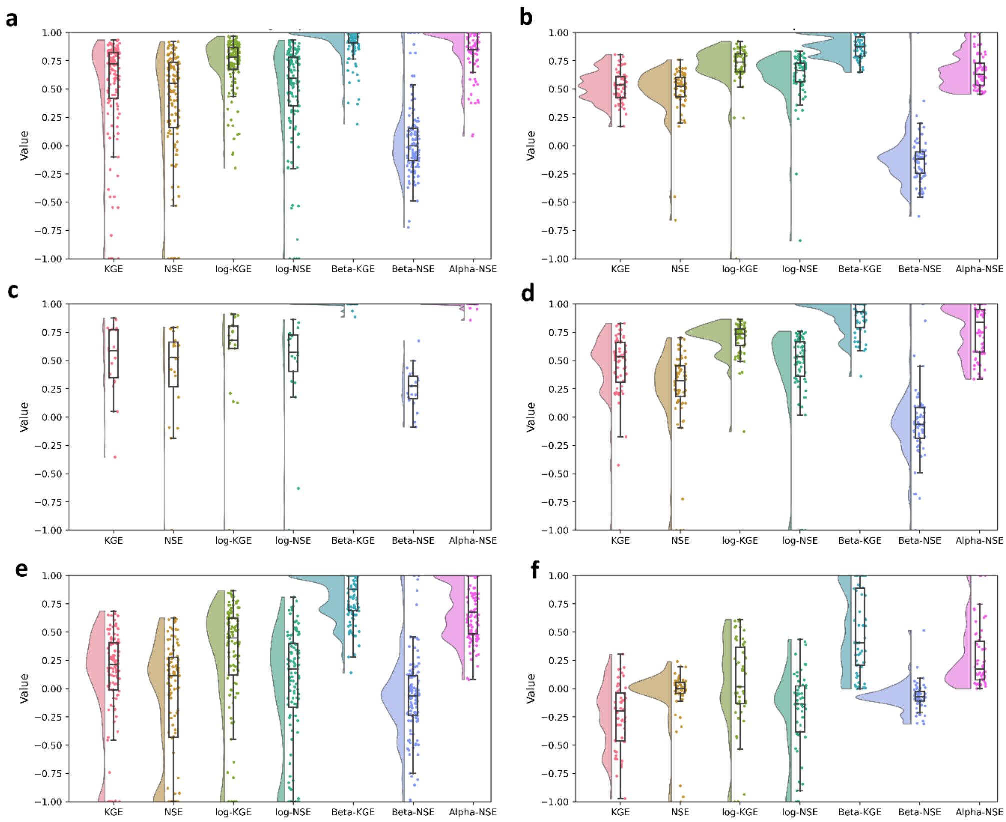

Figure 5. Performance evaluation of Bakaano-Hydro across multiple river basins; a is Amazon river basin, b is Rhine basin, c is Niger basin, d is Upper Mississippi basin, e is Missouri basin and f is Orange basin. The evaluation shows performance across multiple metrics KGE is NSE is . Each panel presents half violin plot, boxplots and density plots to visualize the distribution of the seven metrics. The boxplot shows distribution quartiles and whiskers show the full range; the points show the distribution for the gauging stations and the half-violin plots show the distribution density. The evaluation period was from 1982 to 1989 for each station and was not included during training for each basin.

To assess Bakaano-Hydro performance under low-flow conditions, log-transformed versions of NSE and KGE were used. Once again, the Amazon, Rhine, and Upper Mississippi basins showed strong results, with high log-KGE and log-NSE values and narrow distributions. Niger also performed adequately in low-flow periods, though with more variation in log-NSE, indicating occasional underestimation of recession flows.

https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

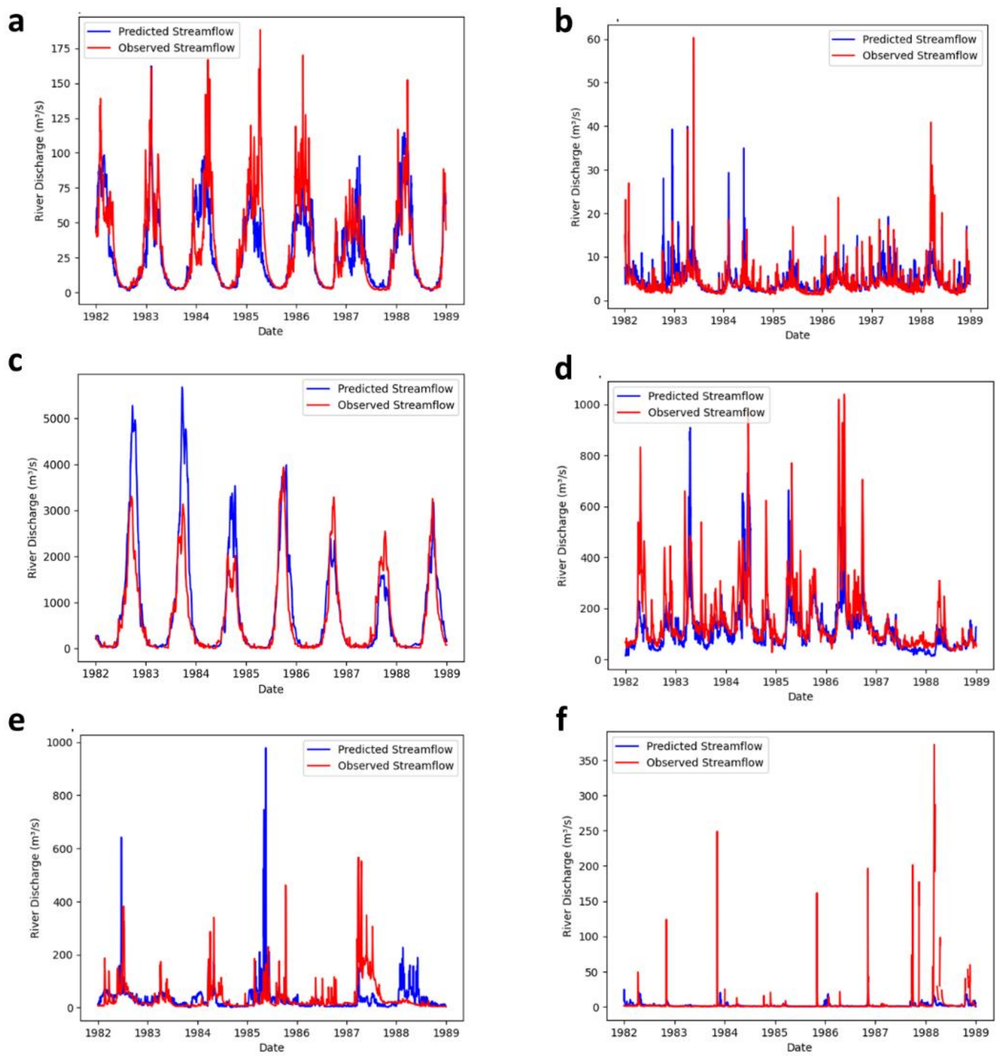

Figure 6. Hydrographs of stations with the median KGE during the evaluation in each river basin: a) is Amazon river basin; b) is Rhine basin; c) is Niger basin; d) is Upper Mississippi basin; e) is Missouri basin and f) is Orange basin. The evaluation period was from 1982 to 1989 for each station and was not included during training for each basin.

Performance in the Orange and Missouri basins was notably weaker, consistent with the challenge of capturing low, variable flows in dryland or snow-affected systems (Fig.6). Model bias was analyzed using the Beta-KGE and Beta-NSE components. Beta-KGE values were consistently close to 1 across all basins,

https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

indicating that the model accurately reproduces mean discharge with low absolute bias. However, Beta-NSE values were generally lower and more variable, particularly in the Missouri and Orange basins. This discrepancy reflects the fact that while the absolute bias may be small, it is large relative to the natural variability of observed flows in these regions. The Alpha-NSE metric was used to assess how well the model captures streamflow variability. Alpha values close to 1 indicate a good match between the standard deviation of observed and simulated flows. Performance was strongest in the Amazon, Rhine, and Upper Mississippi, where Alpha-NSE values were high and consistent across stations. The Niger basin also showed strong performance, reflecting the model's ability to reproduce seasonal variability. Lower and more variable Alpha-NSE values in the Missouri and Orange basins suggest that the model tends to underestimate variability in streamflow. VegET uses a saturation-excess runoff generation mechanism, where runoff occurs only after soil saturation. The simulated runoff is subsequently routed and fed into the neural network component to predict streamflow. This configuration is particularly effective in humid, energy-limited basins with frequent rainfall and sustained soil moisture, where saturation-excess dynamics align with observed runoff generation processes (Beven, 2012; Kidron, 2021; Dunne and Black, 1970; Chow et al., 1988). However, in semi-arid or subtropical basins such as the Orange and parts of the Missouri, where rainfall is infrequent and intense, and soils are often dry or hydrophobic, runoff is more likely generated through infiltration-excess (Hortonian) mechanisms (Chow et al., 1988; Kidron, 2021; Beven, 2012). Because VegET does not initiate runoff until the soil is saturated, it may fail to simulate rapid runoff events triggered by high-intensity storms, leading to underestimation of peaks and timing errors. During extended dry periods, soil moisture remains well below saturation thresholds, resulting in negligible runoff even when surface runoff may occur in reality. These limitations hinder the neural network's ability to reproduce sharp hydrograph responses in water-limited systems.

Overall, the diagnostic evaluation highlights a key structural constraint: the model performs best in humid or temperate basins where saturation-excess runoff processes dominate and flow regimes are gradual and persistent. In contrast, performance declines in dry or snow-affected basins with intermittent or threshold-driven hydrology. Despite these limitations, the model demonstrates strong potential for cross-regional application, particularly in environments where vegetation, soil moisture, and water availability interact predictably to drive runoff production as well as in data-scarce basins (Fig.7). This supports its broader utility in global-scale hydrological modeling, especially when complemented by improvements that incorporate infiltration-excess processes or hybrid runoff generation schemes.

https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

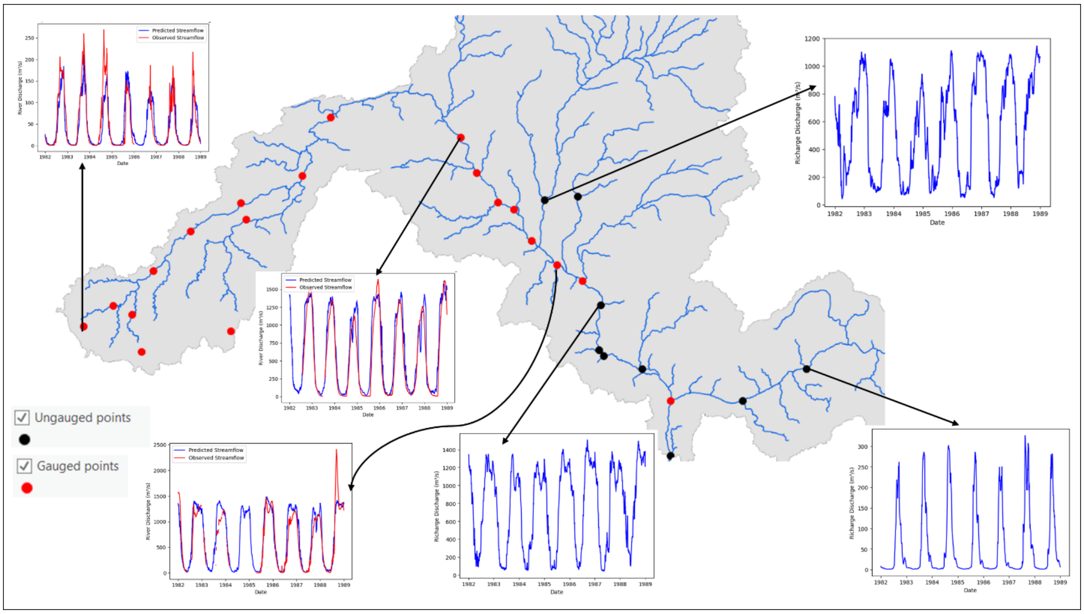

Figure 7. Demonstration of Bakaano-Hydro's fully distributed streamflow prediction capability across a river basin. The map shows gauged points (red circles) used for training and evaluation and ungauged prediction points (black circles). Insets illustrate predicted vs. observed streamflow hydrographs at selected gauged during the evaluation period and predicted streamflow hydrographs at ungauged locations, highlighting the model's ability to generalize across space and provide reliable streamflow estimates at any point within the basin.

5. Conclusions

This paper presents Bakaano-Hydro, a fully distributed hybrid modeling framework that integrates physically based runoff generation, topographic flow routing, and deep learning-based streamflow prediction. By coupling gridded process-based components with a temporal neural network architecture, the model effectively captures spatial heterogeneity, hydrological connectivity, and temporal dynamics within a cohesive structure. In contrast to traditional lumped or semi-distributed data-driven approaches, Bakaano-Hydro enables spatially explicit streamflow prediction at any location within a basin, informed by physically meaningful processes. The model was evaluated across six river basins of varying sizes, spanning four continents and representing diverse hydroclimatic regimes and land-use patterns. Results demonstrate strong predictive skill in humid, snow-fed, and seasonally dynamic systems, and consistent generalization across basins of varying size and complexity. Performance was highest in catchments dominated by saturation-excess runoff, while reduced accuracy in dryland basins highlights the need for future integration of infiltration-excess processes to enhance robustness in arid environments. By embedding hydrological realism within a data-driven framework, Bakaano-Hydro enhances interpretability and improves generalizability—key advantages for next-generation streamflow modeling. Importantly, Bakaano-Hydro is modular, fully automated, and open-source, allowing users to easily deploy it in new regions or for new scenarios. These features make the framework particularly well-suited for application in data-scarce regions, where conventional physically based models may be hindered by parameter uncertainty and limited calibration data. With minimal tuning, the model can be used to generate reliable streamflow predictions in poorly monitored catchments, offering a valuable tool for regions where water-related decision-making is often most urgent but least supported by data infrastructure. Beyond its use in research, Bakaano-Hydro offers strong potential for operational and policy-relevant applications, including real-time streamflow forecasting, climate impact assessments, drought and flood risk monitoring, and long-term adaptation planning. In future work, we aim to extend Bakaano-Hydro by incorporating infiltration-excess, integrating uncertainty quantification, and enhancing coupling with land surface or vegetation models. Together, these additions will further improve the model's robustness, transparency, and applicability to a wider range of hydrological settings and societal needs. The source code is publicly available at https://github.com/confidence-duku/bakaano-hydro to support reproducibility, collaborative development, and wider adoption by the hydrological modeling and Earth system science communities.

Code availability

The minted version of Bakaano-Hydro v1.1 is available on Zenodo at https://doi.org/10.5281/zenodo.15227201. Bakaano-Hydro is also available on GitHub at https://github.com/confidence-duku/bakaano-hydro. The *README.md* file on https://github.com/confidence-duku/bakaano-hydro provides documentation on installation and usage requirements. The *requirements.txt* also on GitHub https://github.com/confidence-duku/bakaano-hydro provides details on the required dependencies. Additionally, a Jupyter Notebook, *quick_start.ipynb*, is available and provides a stepwise guidance to the full Bakaano-Hydro pipeline from download to model training and simulation.

Data availability

Meteorological data for the six case studies are available from

https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

https://doi.org/10.48364/ISIMIP.836809.2 (Karger, 2021). Tree cover and herbaceous cover data are available from https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD44B#bands (Dimiceli et al., 2015). NDVI data are available from https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD13A2 (Didan, 2015). Soil data are available from https://files.isric.org/soilgrids/ (Poggio et al., 2021). Elevation data are available from https://data.hydrosheds.org/file/hydrosheds-v1-dem/hyd_glo_dem_30s.zip (Lehner et al., 2008). Observed streamflow data are available from https://portal.grdc.bafg.de/applications/public.html?publicuser=PublicUser#dataDownload/Home (Grdc, 2025).

Competing interests

The author declares no conflict of interest.

Reference

Anderson, S. and Radić, V.: Evaluation and interpretation of convolutional long short-term memory networks for regional hydrological modelling, Hydrol. Earth Syst. Sci., 26, 795-825, 10.5194/hess-26-795-2022, 2022.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R.: Large area hydrologic modeling and assessment part I: model development 1, JAWRA Journal of the American Water Resources Association, 34, 73-89, 1998.

Arsenault, R., Martel, J. L., Brunet, F., Brissette, F., and Mai, J.: Continuous streamflow prediction in ungauged basins: long short-term memory neural networks clearly outperform traditional hydrological models, Hydrol. Earth Syst. Sci., 27, 139-157, 10.5194/hess-27-139-2023, 2023.

Bai, S., Kolter, J. Z., and Koltun, V.: An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv, arXiv preprint arXiv:1803.01271, 10, 2018.

Beven, K. J.: Rainfall-runoff modelling: the primer, John Wiley & Sons2012.

Cassagnole, M., Ramos, M. H., Zalachori, I., Thirel, G., Garçon, R., Gailhard, J., and Ouillon, T.: Impact of the quality of hydrological forecasts on the management and revenue of hydroelectric reservoirs – a conceptual approach, Hydrol. Earth Syst. Sci., 25, 1033-1052, 10.5194/hess-25-1033-2021, 2021.

Chow, V. T., Maidment, D. R., and Mays, L. W.: Applied Hydrology, MacGraw-Hill, Inc., New York, 572, 1988.

De Roo, A. P. J., Wesseling, C. G., and Van Deursen, W. P. A.: Physically based river basin modelling within a GIS: the LISFLOOD model, Hydrological Processes, 14, 1981-1992, https://doi.org/10.1002/1099-1085(20000815/30)14:11/12<1981::AID-HYP49>3.0.CO;2-F, 2000.

Depetris, P. J.: The Importance of Monitoring River Water Discharge, Frontiers in Water, 3, 10.3389/frwa.2021.745912, 2021.

Didan, K.: MOD13A2 MODIS/terra vegetation indices 16-day L3 global 1km SIN grid V006, (No Title), 2015.

https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

Didan, K.: MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V061 [Data set]. NASA EOSDIS Land Processes Distributed Active Archive Center, Accessed 2023-11-09 from https://doi. org/10.5067/MODIS/MOD13Q1. 061, 2021.

DiMiceli, C., Carroll, M., Sohlberg, R., Kim, D.-H., Kelly, M., and Townshend, J.: MOD44B MODIS/Terra vegetation continuous fields yearly L3 global 250m SIN grid V006, NASA EOSDIS Land Processes DAAC, 10, 2015.

Dunne, T. and Black, R. D.: Partial area contributions to storm runoff in a small New England watershed, Water resources research, 6, 1296-1311, 1970.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., and Michaelsen, J.: The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes, Scientific Data, 2, 150066, 10.1038/sdata.2015.66, 2015.

Gauch, M., Kratzert, F., Klotz, D., Nearing, G., Lin, J., and Hochreiter, S.: Rainfall–runoff prediction at multiple timescales with a single Long Short-Term Memory network, Hydrol. Earth Syst. Sci., 25, 2045-2062, 10.5194/hess-25-2045-2021, 2021.

Golden, H. E., Christensen, J. R., McMillan, H. K., Kelleher, C. A., Lane, C. R., Husic, A., Li, L., Ward, A. S., Hammond, J., Seybold, E. C., Jaeger, K. L., Zimmer, M., Sando, R., Jones, C. N., Segura, C., Mahoney, D. T., Price, A. N., and Cheng, F.: Advancing the science of headwater streamflow for global water protection, Nature Water, 3, 16-26, 10.1038/s44221-024-00351-1, 2025.

GRDC: The Global Runoff Data Centre, 56068 Koblenz, Germany (https://grdc.bafg.de/), 2025.

Hamman, J. J., Nijssen, B., Bohn, T. J., Gergel, D. R., and Mao, Y.: The Variable Infiltration Capacity model version 5 (VIC-5): infrastructure improvements for new applications and reproducibility, Geosci. Model Dev., 11, 3481-3496, 10.5194/gmd-11-3481-2018, 2018.

Hargreaves, G. H. and Samani, Z. A.: Reference crop evapotranspiration from temperature, Applied engineering in agriculture, 1, 96-99, 1985.

Herrera, D., Ellis, A., Fisher, B., Golden, C. D., Johnson, K., Mulligan, M., Pfaff, A., Treuer, T., and Ricketts, T. H.: Upstream watershed condition predicts rural children's health across 35 developing countries, Nature Communications, 8, 811, 10.1038/s41467-017-00775-2, 2017.

Herrera, P. A., Marazuela, M. A., and Hofmann, T.: Parameter estimation and uncertainty analysis in hydrological modeling, WIREs Water, 9, e1569, https://doi.org/10.1002/wat2.1569, 2022.

Hunt, K. M. R., Matthews, G. R., Pappenberger, F., and Prudhomme, C.: Using a long short-term memory (LSTM) neural network to boost river streamflow forecasts over the western United States, Hydrol. Earth Syst. Sci., 26, 5449-5472, 10.5194/hess-26-5449-2022, 2022.

Ioffe, S. and Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift, International conference on machine learning, 448-456,

Karger, D. N., Lange, S., Hari, C., Reyer, C. P. O., Conrad, O., Zimmermann, N. E., and Frieler, K.: CHELSA-W5E5: daily 1 km meteorological forcing data for climate impact studies, Earth Syst. Sci. Data, 15, 2445-2464, 10.5194/essd-15-2445-2023, 2023.

Karger, D. N. L., S; Hari, C.; Reyer , P. O. C.; Zimmermann, E. N.: CHELSA-W5E5 v1.1: W5E5 v1.0 downscaled with CHELSA v2.0, ISIMIP Repository. [dataset], 2021.

Kidron, G. J.: Comparing overland flow processes between semiarid and humid regions: Does saturation overland flow take place in semiarid regions?, Journal of Hydrology, 593, 125624, https://doi.org/10.1016/j.jhydrol.2020.125624, 2021.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrol. Earth Syst. Sci., 22, 6005-6022, 10.5194/hess-22-6005-2018, 2018.

https://doi.org/10.5194/egusphere-2025-1633
Preprint. Discussion started: 5 May 2025

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., and Nearing, G.: Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets, Hydrol. Earth Syst. Sci., 23, 5089-5110, 10.5194/hess-23-5089-2019, 2019.

Lehner, B., Verdin, K., and Jarvis, A.: New Global Hydrography Derived From Spaceborne Elevation Data, Eos, Transactions American Geophysical Union, 89, 93-94, https://doi.org/10.1029/2008EO100001, 2008.

Luong, M.-T., Pham, H., and Manning, C. D.: Effective approaches to attention-based neural machine translation, arXiv preprint arXiv:1508.04025, 2015.

Moges, E., Demissie, Y., Larsen, L., and Yassin, F.: Review: Sources of Hydrological Model Uncertainties and Advances in Their Analysis, Water, 13, 28, 2021.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J. N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth Syst. Sci. Data, 13, 4349-4383, 10.5194/essd-13-4349-2021, 2021.

Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., Hassidim, A., Klotz, D., Kratzert, F., Metzger, A., Nevo, S., Pappenberger, F., Prudhomme, C., Shalev, G., Shenzis, S., Tekalign, T. Y., Weitzner, D., and Matias, Y.: Global prediction of extreme floods in ungauged watersheds, Nature, 627, 559-563, 10.1038/s41586-024-07145-1, 2024.

Palmer, M. and Ruhi, A.: Linkages between flow regime, biota, and ecosystem processes: Implications for river restoration, Science, 365, eaaw2087, doi:10.1126/science.aaw2087, 2019.

Perez, E., Strub, F., De Vries, H., Dumoulin, V., and Courville, A.: Film: Visual reasoning with a general conditioning layer, Proceedings of the AAAI conference on artificial intelligence,

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D.: SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty, SOIL, 7, 217-240, 10.5194/soil-7-217-2021, 2021.

Quinn, P., Beven, K., Chevallier, P., and Planchon, O.: The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models, Hydrological processes, 5, 59-79, 1991.

Schaphoff, S., von Bloh, W., Rammig, A., Thonicke, K., Biemans, H., Forkel, M., Gerten, D., Heinke, J., Jägermeyr, J., Knauer, J., Langerwisch, F., Lucht, W., Müller, C., Rolinski, S., and Waha, K.: LPJmL4 – a dynamic global vegetation model with managed land – Part 1: Model description, Geosci. Model Dev., 11, 1343-1375, 10.5194/gmd-11-1343-2018, 2018.

Senay, G. B.: Modeling Landscape Evapotranspiration by Integrating Land Surface Phenology and a Water Balance Algorithm, Algorithms, 1, 52-68, 2008.

Senay, G. B., Kagone, S., Parrish, G. E. L., Khand, K., Boiko, O., and Velpuri, N. M.: Improvements and Evaluation of the Agro-Hydrologic VegET Model for Large-Area Water Budget Analysis and Drought Monitoring, Hydrology, 10, 168, 2023.

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

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting, The journal of machine learning research, 15, 1929-1958, 2014.

Yu, Q., Tolson, B. A., Shen, H., Han, M., Mai, J., and Lin, J.: Enhancing long short-term memory (LSTM)-based streamflow prediction with a spatially distributed approach, Hydrol. Earth Syst. Sci., 28, 2107-2122, 10.5194/hess-28-2107-2024, 2024.