https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

Integration of physics-informed deep learning with a distributed hydrological model (PIDL-DHM): A hybrid model for runoff simulation

Authors: Junping Wang[1,2], Baolin Xue[1,3,*], Guoqiang Wang[1]

Affiliations： [1]Innovation Research Center of Satellite Application (IRCSA), Faculty of Geographical Science, Beijing Normal University, Beijing, China

[2]National Institute of Natural Hazards, Ministry of Emergency Management of the People's Republic of China, Beijing, 100085, China

[3]Autonomous Region Collaborative Innovation Center for Integrated Management of Water Resources and Water Environment in the Inner Mongolia Reaches of the Yellow River, Hohhot 010018, China

Corresponding author: Baolin Xue, Innovation Research Center of Satellite Application (IRCSA), Beijing Normal University, Xinjiekouwai Street 19, Haidian, Beijing 100875, China

E-mail: xuebl@bnu.edu.cn (B. Xue)

Tel: +86-10-5880-4985

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

Abstract： Climate change and vegetation restoration have altered hydrological processes, but quantifying vegetation-runoff interactions remains challenging due to complex feedback and data scarcity. We developed a physics-informed deep learning hybrid model that couples process-based hydrology with deep learning to improve runoff prediction in vegetation-altered basins. The model explicitly represents transpiration, interception, snow dynamics, dual runoff generation, water-balance constraints, and distributed spatial heterogeneity. Evaluations in three Chinese basins spanning hydroclimatic and vegetation gradients show that the hybrid models outperform pure deep learning, with the strongest improvements in semi-arid and vegetation-transition basins. Correlation coefficients exceeded 0.8 at monthly/annual scales and 0.75 at daily scales in the most responsive basins. Compared with GR4J, the added complexity is most beneficial in heterogeneous semi-arid and ungauged sub-basins. These results demonstrate that physical constraints improve extrapolation reliability and support hybrid models for sustainable basin management under climate change.

Keywords： Hybrid models; Deep learning; Vegetation–runoff interactions; Climate change; Runoff prediction; Vegetation dynamics

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

Highlights

1, This study proposed a physics-informed deep learning model that incorporates the dual-regulation mechanisms of canopy and root systems as well as dynamic runoff processes.

2, It demonstrated better transferability to ungauged basins compared to traditional hydrological models.

3, Additionally, the model exhibited superior robustness under drought or hot climate conditions relative to other modeling approaches.

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

1. Introduction

Runoff prediction models are essential tools for flood forecasting, drought assessment, water allocation, and ecological restoration planning (Wang and Zeng, 2024). However, the reliability of runoff prediction is increasingly challenged by non-stationary environmental change (Konapala et al., 2020). Climate warming alters precipitation regimes, snow processes, and soil moisture dynamics, while vegetation restoration, afforestation, agricultural expansion, and land-cover change modify canopy interception, transpiration, and infiltration (Li et al., 2018; Wang et al., 2019). These changes jointly reshape the partitioning of precipitation into evapotranspiration, soil water storage, and surface runoff (Shen et al., 2018; Brient, 2020; Sit et al., 2020). Consequently, hydrological models developed under historically stationary may become unreliable when applied to basins undergoing rapid ecological or climate change (Li et al., 2022).

Traditional runoff prediction models predominantly rely on process-based hydrological models, such as the Soil and Water Assessment Tool (SWAT) (G. Arnold et al., 2012; Hamman et al., 2018) and the Variable Infiltration Capacity (VIC) model (Liu et al., 2025; Li et al., 2018), which offer physically-based representations of water and energy cycles (Bi et al., 2023; Zhou et al., 2021; Du et al., 2022; Li et al., 2024). A key strength of these models is their spatial discretization, which enables the representation of spatial heterogeneity (Shang et al., 2023; Xue et al., 2024). However, a major limitation of these models remains their vegetation-related parameters, rooting structures, phenological development, and stomatal regulation are commonly simplified or prescribed (Chen et al., 2025; Gaur and Drewry, 2024; Liu et al., 2020). Recent studies have shown that neglecting vegetation dynamics can lead to substantial uncertainty in evapotranspiration and runoff simulations. For example,

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

incorporating remotely sensed vegetation phenology into SWAT improved the representation of vegetation growth and ecohydrological processes, indicating that static phenological assumptions can bias water-balance simulations (Chen et al., 2023). Similarly, studies on land-surface and Earth-system models have shown that inaccurate sensitivity of evapotranspiration to vegetation change can propagate into misleading hydrological and climate feedbacks (Wang et al., 2024; Huang et al., 2016; Liu et al., 2025). These findings suggest that these traditional models may fail to adequately represent such complex interactions, leading to critical gaps in runoff prediction accuracy, especially during growing seasons, droughts, and in regions with limited data (Lawrence et al., 2019; Li et al., 2015; Yao et al., 2025; Barger et al., 2011; Xue et al., 2022).

These deficiencies are not merely theoretical. In a study of 156 Austrian catchments, Duethmann et al. (2020) found that a conceptual semi-distributed HBV-type hydrological model showed a mean gap of 95 mm yr^{-1} per 35 years between simulated and observed discharge trends. Accounting for vegetation dynamics using a satellite-derived vegetation index in the calculation of reference evaporation reduced this gap by 36 mm yr^{-1} per 35 years, highlighting the role of missing vegetation processes in long-term hydrological simulations. The challenge is particularly pronounced in basins where vegetation-water interactions vary strongly across space. In semi-arid or transitional basins, shallow-rooted grasslands and sparsely vegetated areas may respond rapidly to precipitation, whereas forested regions can sustain transpiration and baseflow through deeper soil-water access (Luo et al., 2020; Gao et al., 2022; Zhang et al., 2022). These contrasting mechanisms mean that a single lumped parameter set may reproduce outlet discharge but still fail to represent internal hydrological processes (Du et al., 2022; Camarero et al., 2024). Thus, evaluating models only by outlet

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

runoff can mask unrealistic spatial runoff generation patterns (Wei et al., 2021; Wood et al., 1992; Wang et al., 2007).

Deep learning models have recently achieved strong performance in rainfall-runoff prediction because of their ability to learn nonlinear relationships from large datasets (Ma et al., 2021; Van Katwyk et al., 2023). Models such as recurrent neural networks and long short-term memory networks can capture temporal dependence and complex hydroclimatic responses (Rumelhart et al., 1986). However, purely data-driven models are often difficult to interpret and may extrapolate poorly under conditions that differ from the training period, such as extreme droughts or ungauged subbasins (Khoshkalam et al., 2023; S. and Q., 2010). Their predictions may also violate water-balance consistency or produce hydrologically unreasonable internal fluxes (Sit et al., 2020). These limitations reduce their reliability for water-resource decision-making under climate and land-cover change.

Physics-informed deep learning provides a promising pathway to address these limitations by embedding hydrological process knowledge into differentiable neural network architectures (Tian et al., 2025; He et al., 2020; Shen et al., 2021). In such frameworks, physical equations, state variables, and process-based constraints can be integrated into neural networks, allowing model parameters to be optimized by gradient-based learning while maintaining physical interpretability (Lees et al., 2022; Li et al., 2024; Wang et al., 2018). Previous physics-informed hydrological models have demonstrated improved generalization by incorporating snow processes, soil-water dynamics, routing schemes, or water-balance constraints (Ji and Song et al., 2025; Shen & Lawson, 2021; Höge et al., 2022; Koppa et al., 2022). Following researches integrated more complex physical processes focusing on soil moisture dynamics (Konapala et al., 2020; Shang et al., 2023). However, most existing

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

physics-informed rainfall-runoff models still focus primarily on runoff generation and storage dynamics, while vegetation physiological regulation, canopy interception, dynamic stomatal resistance, and root-zone water uptake remain insufficiently represented (Zhong et al., 2023). This limits their ability to simulate runoff under vegetation-altered or climate-stressed conditions (Elghawi et al., 2023; Zhang et al., 2024; Wang et al., 2024; Xie et al., 2021). This limitation shows the need to further refine PIDL (Physics-informed Deep Learning) models to consider these critical processes (Dazzi, 2024; Vicente-Serrano et al., 2022; Niu et al., 2005).

This study focuses on the key questions: How can key ecological processes be dynamically embedded within a PIDL framework? How can reliable predictive performance be maintained in data-scarce regions? To address these challenges, this study develops a distributed physics-informed deep learning hydrological model, referred to as PIDL-DHM (Physics-informed Deep Learning with a Distributed Hydrological Model), that explicitly couples hydrological process constraints with dynamic vegetation-water interactions. The framework integrates three key components: (1) a physics-informed recurrent hydrological module that represents water-balance dynamics, snow processes, runoff generation, and baseflow; (2) a vegetation-process module that dynamically represents evapotranspiration, canopy interception, and vegetation-mediated infiltration using learnable time-varying parameters; and (3) a distributed routing and delay module that represents spatial heterogeneity in runoff generation and translation to the basin outlet.

We systematically evaluated the model across three climatically and ecologically contrasting Chinese basins from dry to wet climates to demonstrate its enhanced predictive capability and generalization in ungauged basins. We provide a new runoff prediction method

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

in response to climate change and offer valuable insights for water resource management and decision-making in these vulnerable regions.

2. Methodologies and Data Sources

2.1. Model Framework and major improvements

2.1.1. Foundational Architecture

The integration of physics and DL in this study is built upon the open-source contributions of Jiang (2020) and Zhong et al. (2023). Specifically, we adopt the core concept of encoding physical dynamics directly within computational graph of RNN (Figure 1). To implement this concept, we utilize a differentiable framework in PyTorch to rewrite the physical model (Paszke et al., 2019; Patil and Stieglitz et al., 2014). The fundamental innovation involves replacing the standard mathematical operations of RNN and LSTM cells —the weights and bias matrices that transform inputs and hidden states (Figure 1a)—with numerical solutions of physical governing equations and their parameters (Figure 1b). For instance, the cell state update is governed by equations simulating soil moisture movement (e.g., the soil water continuity equation) and runoff generation, while physical parameters (e.g., saturated hydraulic conductivity) become learnable parameters within the network. During the training process, the gradients of the loss function are backpropagated through these physical equations through automatic differentiation in PyTorch, enabling the joint optimization of both physical parameters and remaining network weights. Consequently, the hidden state variable h_t is exclusively composed of the physical state variables s_0 (e.g. snow water equivalent and soil moisture content), forming a physically interpretable “memory” of the hydrological system carried between time steps. All the neural components are unified

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

within a single end-to-end architecture, facilitating joint optimization across physical and data-driven modules.

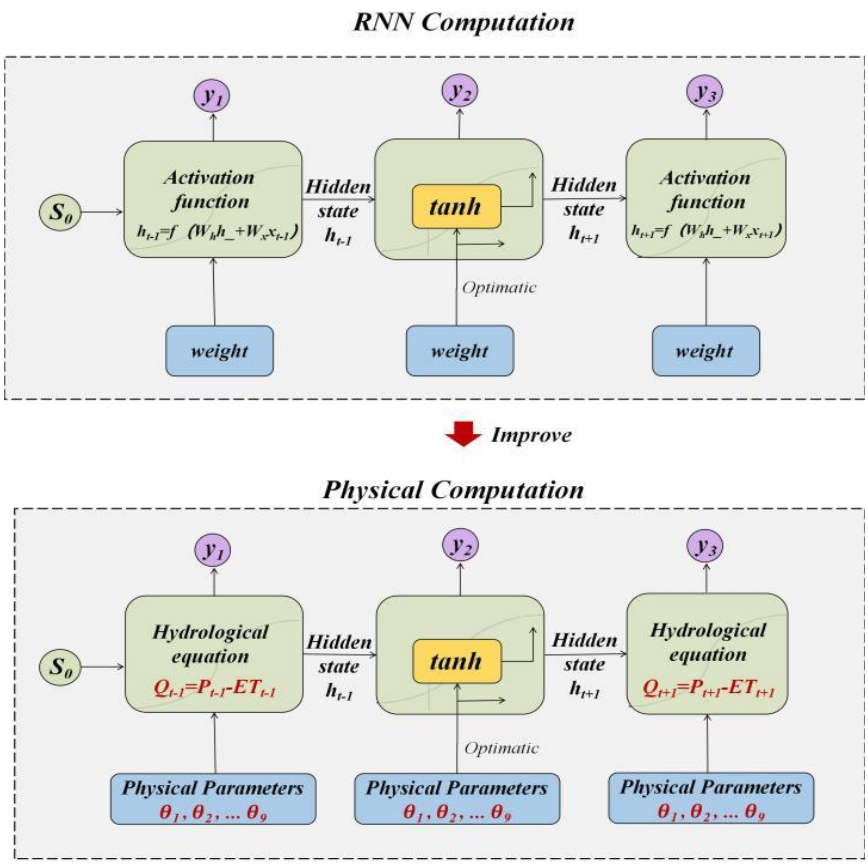


Figure 1. (a) Schematic diagram of the pure neural network unit of RNN and LSTM. (b) Schematic diagram of hybrid models, which add physical information to the neural network units of RNN and LSTM. The hidden state and weight in a recurrent neural network can be replaced by physical equations and parameters (e.g., the soil water content) and physical parameters

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

2.1.2. Major Improvements

Beyond above approaches, our novel framework introduces major improvements in representing runoff generation and vegetation processes. We rewrite the runoff generation and vegetation processes model, which are encoded as differentiable components within the PyTorch framework. Our key innovations within this framework include: (1) Dual-mechanism runoff generation. Dry basins exhibit both infiltration excess runoff during high-intensity storms and saturation-excess runoff when local soils reach field capacity after prolonged wetting. Unlike Jiang's (2020) single-mechanism approach, our PRNN incorporates a dynamic switching function that activates either infiltration-excess or saturation-excess runoff mechanisms based on real-time climate inputs, allowing temporal coexistence of both processes. (2) Plant–water interaction module. By incorporating a dedicated LSTM network for vegetation processes, we explicitly represent memory effects in plant-water interactions, including precipitation interception, stomatal transpiration, and root water uptake.

2.1.3 Physical Mechanisms

(1) Precipitation and Snowmelt module

In our model, precipitation and snow are divided based on the temperature thresholds T_{min} and T_{max}, and snow melt is modeled on the basis of the degree-day factor:

$$P_{snow}=\begin{cases}0 & T\geq T_{min}\\ P & T<T_{min}\end{cases} \tag{1}$$

$$P_{rain}=\begin{cases}P & T>T_{min}\\ 0 & T\leq T_{min}\end{cases} \tag{2}$$

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

$$M=\begin{cases} min\left|S_0,DDF\cdot\ (T-T_{max})\right| & T>T_{max} \\ 0 & T\leq T_{max} \end{cases} \tag{3}$$

where P and T are the precipitation and air temperature, respectively; DDF is the degree-day factor; P_{snow} and P_{rain} are the snowfall and precipitation, respectively; $T_{\min}$ is a training parameter representing the threshold temperature of the snowfall; $T_{\max}$ is a training parameter representing the snowmelt threshold temperature; and S_0 is the soil snow bucket content.

(2) Evapotranspiration

Evapotranspiration (ET) is calculated using the FAO56 Penman–Monteith (PM) Equation (5), which is a widely accepted method for estimating ET under various environmental conditions (Penman and Keen, 1997). Dynamic stomatal resistance is also used as a parameter, allowing for the incorporation of plant-specific responses to environmental stress (e.g., drought conditions).

$$ET=\begin{cases} ET_p\times(\frac{S_1}{S_{max}}) & 0\leq S_1\leq S_{max} \\ PET & S_1>S_{max} \end{cases} \tag{4}$$

$$e_s=0.611\times exp\left(\frac{17.3T}{T+237.3}\right) \tag{5}$$

$$ET_p=\frac{\Delta(R_n-G)+[\rho\cdot cp\cdot(e_s-e_a)]/r_a}{\Delta+\gamma(1+r_s/r_a)} \tag{6}$$

where S_1 and ET are water soil moisture content and actual evaporation, respectively, and where S_{max} is the storage capacity of the basin water bucket. R_n represents the net radiation, $MJ/m^2/day$. G is the soil heat flux density, which is set to 0 and is also measured in $MJ/m^2/day$. Δ is the slope of the saturation vapor pressure curve, expressed in $kPa/°C$. e_s

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

represents the saturation vapor pressure; e_a represents the actual water vapor pressure; ρ represents the air density; cp represents the specific heat capacity; and γ represents the hygrometer constant. $\mathrm{r_s}$ and r_a are the stomatal resistance and aerodynamic resistance, respectively.

(3) Runoff and baseflow

By integrating the core ideas of the soil water storage curve (SWAT) and VIC (superinfiltration runoff) (Dunne and Black, 1970), the baseflow and surface runoff calculation formulas are as follows:

$$q_{base}=\begin{cases} q_{max}\times(\frac{S_1}{S_{max}})^{-\beta} & 0\leq S_1\leq S_{max} \\ q_{max} & S_1>S_{max} \end{cases} \tag{7}$$

$$q_{sat}=$$

$$\begin{cases} P_{rain}-(i_{max}-i_0) & P_{rain}>0,\ P_a>i_{max}-i_0 \\ P_{rain}-\frac{i_{max}}{(1+b)}\times\ (1-\frac{i_0}{i_{max}})^{(1+b)}\ \ (1-\frac{i_0+P_{rain}}{i_{max}})^{(1+b)})) & P_{rain}>0,\ P_a\leq i_{max}-i_0 \end{cases} \tag{8}$$

$$q_{unsat}=\begin{cases} P_{rain}-q_{sat}-f_{max}[1-(1-\frac{P_{rain}-q_{sat}}{f_{max}})]^2 & P_{rain}>i_{max}-i_0 \\ S_1-S_{max} & P_{rain}\leq i_{max}-i_0 \end{cases} \tag{9}$$

$$q_{surf}=q_{unsat}+q_{sat} \tag{10}$$

$$i_0=i_{max}\times(1-(1-A)^{\frac{1}{b}}) \tag{11}$$

$$i_{max}=(1+b)\times S_{max} \tag{12}$$

$$A=\ 1-(1-\frac{S_1}{S_{max}})^{\frac{b}{1+b}} \tag{13}$$

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

where q_{base} and q_{surf} are the baseflow and surface runoff, respectively. q_{unsat} and q_{sat} represent saturated runoff and unsaturated runoff, respectively. q_{max}, S_{max} and f are training parameters representing the maximum baseflow when the catchment bucket is fully saturated, the maximum water storage, and a decay factor for the decline rate of runoff. f_{max} is the maximum infiltration velocity. b is the shape parameter of the storage capacity curve details. i_{max} is the maximum infiltration capacity.

(4) Vegetation module

In this study, precipitation and temperature were used as additional inputs to the vegetation LSTM module to implicitly represent vegetation growth processes (Zhong et al., 2023; Gao et al., 2021). The dynamic vegetation LSTM module includes evapotranspiration, infiltration, and canopy interception. In the vegetation canopy interception and water storage module, vegetation canopy interception is represented using the Holton vegetation formula (Holtan, 1961). The interception loss can be expressed as:

$$I_n=S_v+C_{sint}P_{rain} \tag{14}$$

where I_n represents the interception loss; S_v represents the water storage capacity of the vegetation in the canopy shelter area, S_v=0; and P_{rain} represents the precipitation in the vegetation cover area.

The evapotranspiration (ET) is calculated using the FAO56 Penman–Monteith (PM), but r_s used dynamic parameters for stomatal resistance. Additionally, the impact of vegetation on runoff is expressed using the Holtan formula (Holtan, 1961):

$$f(t)=GI\times A_a\times S_1{}^n+f_c \tag{15}$$

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

where $f(t)$ and f_c are the infiltration rate and stable infiltration rate, respectively, which are incorporated into the model to represent the physical properties of the soil and vegetation interaction. GI is the plant growth index, which is a dynamic parameter that represents the plant's growth stage. n is an empirical coefficient, and the experience value is 1.4. A_a is the ratio of the infiltration capacity and the soil surface water shortage, which represents the porosity connected with the surface and the vegetation root density index affecting infiltration. S_1 represents the water shortage capacity of the ground surface layer.

(5) Distributed delay kernel model equations

With the impact of spatial information on runoff modeling, the basin delay kernel model proposed by Sun et al. (2024) is like our fully distributed model but does not explicitly consider the runoff routing. Here we use a distributed hybrid-P-D configuration consider runoff generation move routing, so that land travel times are represented explicitly by a delay kernel. The delay kernel is a time-distribution operator applied after runoff generation, it reshapes timing while preserving mass at the daily step. Translation and dispersion from grid to the basin outlet are modeled by a learnable delay-kernel routing operator. The routed outlet discharge is computed as a discrete convolution:

$$Q = q_{surf} + q_{base} \tag{16}$$

$$Q_{basin,\,t} = \sum_{\tau=0}^{k-1} K(\tau)\, Q(t-\tau) \tag{17}$$

where Q_{basin} represents the runoff at the basin outlet. Q_i and α_i are the runoff and weighted calculations, respectively, which are based on the inverse distance of the distance from the grid point to the exit. N represents the total number of subbasins within the basin.

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

(6) Water balance constraint PRNNLayer

Mass conservation is enforced by construction through an explicit state-space update in the PRNN Layer, the mass conservation enforced by construction. At each time step, we enforced a basin-scale water balance residual using the model's internally simulated fluxes and storage:

$$P+M=I+ET+Q_{sub}-Q_{surf}-\Delta S \tag{18}$$

Which is identically 0.

2.2. Model Implementation

2.2.1. Architectural Overview

The proposed PIDL framework couples three functionally distinct neural network blocks (Table 1): (1) Physics-Informed Recurrent Neural Networks module (PRNN module), which encodes basin-scale water-balance equations and runoff generation (Figure 2b); (2) Plant Long Short-Term Memory network module (Plant LSTM module), which learns nonlinear vegetation water-use processes (Figure 2d); and (3) Distribution delay kernel module (Distribution module), which is divided into a fully distributed grid (Figure 2c). The three blocks exchange fluxes at each simulation time step, and the PRNN module provides soil moisture surface runoff and basic runoff to the Plant LSTM module, which returns dynamic evapotranspiration, interception loss, and root uptake, and the distribution module generates runoff through each grid.

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

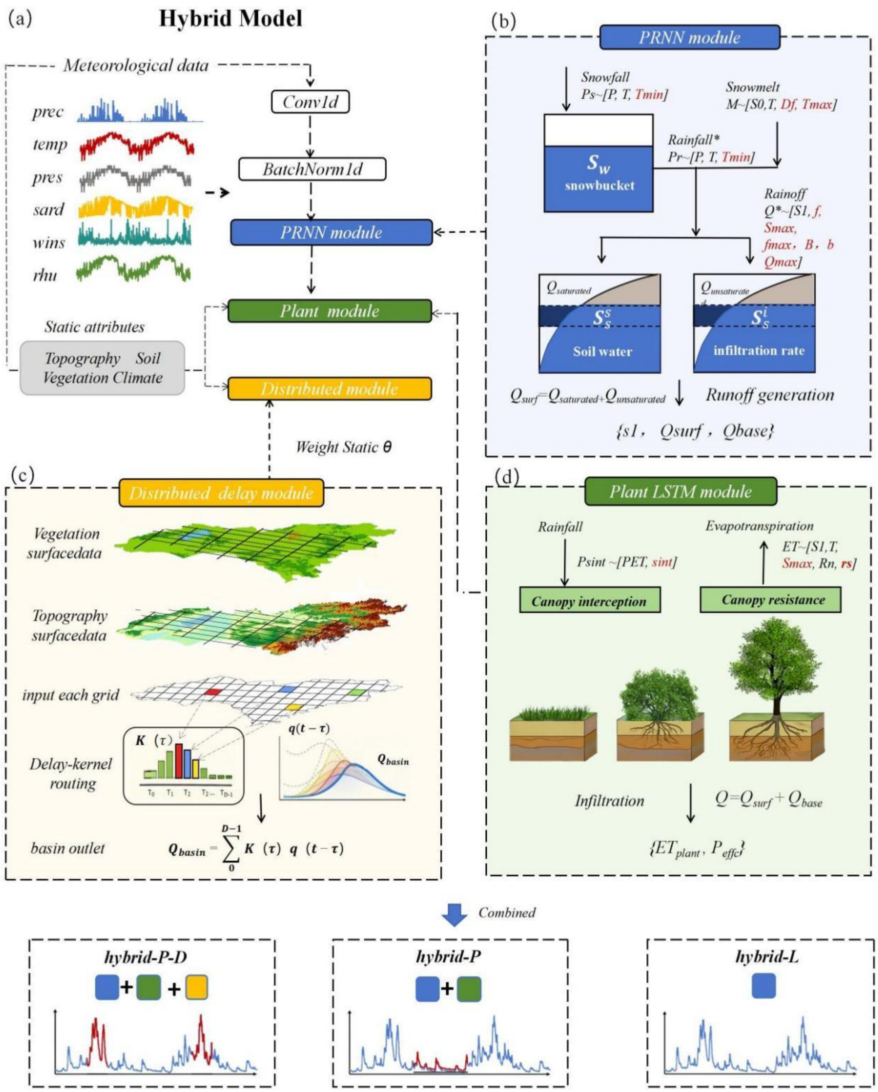


Figure 2. Structure of the physics-informed deep learning distributed hydrological modeling framework (PIDL-DHM). (a) Overview of the hybrid model architecture. (b) PRNN module diagram (highlighted in blue): The physics-informed recurrent neural network module combines DL with physical hydrological processes, including surface runoff and infiltration. For detailed formulations of these processes, refer to the main text (adapted from

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

S. Jiang, 2020). (c) Distributed delay module diagram (highlighted in yellow): This module handles spatially distributed hydrological processes, and each grid includes a PRNN module and a plant LSTM module. (d) Plant LSTM module diagram (highlighted in green): A long short-term memory (LSTM) network that synergistically combines physical mechanisms with vegetation processes. Meteorological inputs, including precipitation (prec), temperature (temp), atmospheric pressure (pres), solar radiation (sard), wind speed (wins), and relative humidity (rhu), drive the model. The figure depicts three model variations: hybrid-L: a hybrid lumped hydrological process model integrating the PRNN module; hybrid-P: a hybrid model that incorporates the Plant LSTM module; and hybrid-P-D: a fully distributed hybrid model incorporating the PRNN, Plant LSTM, and distributed modules. BatchNorm1d: batch normalization technique for stable training. Conv1D: extract spatial features from time series data. FCNN: fully connected neural network for feature integration.

Table 1. Architecture parameters of the hybrid physics-informed neural network model

Item	Definitions
Network	Fully connected neural network
Hidden layer	1 LSTM layer (64 neurons); 2 CNN layers (25 filters)
Activation funtion	Tanh /ReLU/ Heaviside
Optimizer	Adam optimizer
Initial learning rate	0.01
training time	About 300s per 1 epochs

2.2.2. Dynamic Parameterization

To account for dynamic vegetation phenology and its impacts on hydrological processes, we parameterize key vegetation-related coefficients as time-variant variables (Table 2). Vegetation growth dynamically alters canopy interception, transpiration, and root water uptake (Agnihotri et al., 2023), consequently modifying surface runoff generation and soil

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

water transport velocities. Specifically, the canopy interception coefficient (C_i) is treated as a dynamic parameter to capture these vegetation-mediated processes. To further extend this dynamic framework, we also designate the maximum water storage capacity (S_{max}) and the infiltration rate coefficient (f) governing the soil bucket model runoffs as time-dependent parameters. For evapotranspiration (ET) processes, stomatal resistance (r_s) is similarly implemented as a dynamic measure to mitigate the effects of vegetation phenology on ET estimates. All remaining parameters maintain static representations. When implemented as trainable variables, these parameters enable data-adaptive adjustments during model calibration, ensuring accurate representation of observed vegetation characteristics. Mechanistically, r_s modulates transpiration rates, whereas C_i controls canopy water retention capacity. This dynamic parameterization significantly enhances model adaptability across diverse climatic regimes and vegetation types. All physical mechanisms and their formulas used in the model are in supplemental material 1.

2.2.3. Training Configuration

To achieve these innovations, The overall computational workflow began with processing input data through the PRNN module (Figure 2b). Inputs include meteorological forcing and static catchment attributes (Figure 3). To enhance training stability and accelerate convergence, we applied batch normalization (BatchNorm1d) to the input features. For static parameter learning, we deployed a 1D convolutional encoder (Conv1D) to extract spatial features from time series data, which were subsequently integrated with other static features using a fully connected neural network (FCNN) (Fukushima and Miyake, 1982; Botterill and McMillan, 2023). The model was trained using the Adam optimizer with an initial learning rate of 0.01. To prevent overfitting, early stopping was implemented with a patience of 80

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

epochs based on the Nash-Sutcliffe Efficiency (NSE) metric. The detailed architecture parameters are summarized in Table 1.

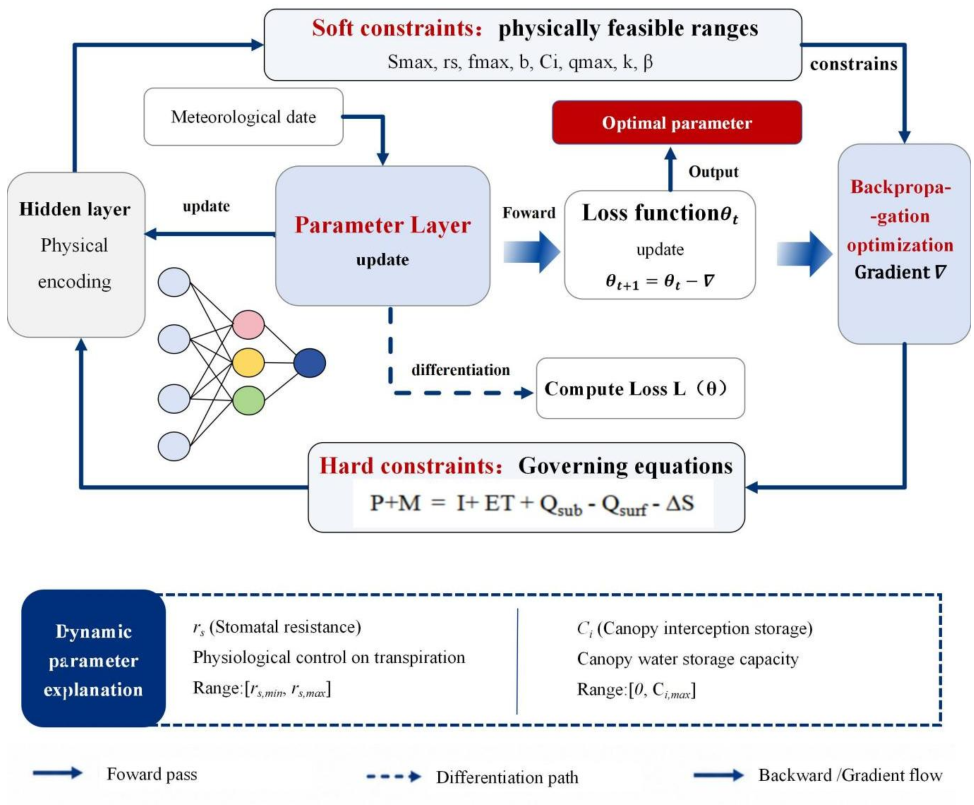


Figure 3. Updating of trainable hydrological parameters in the PIDL-DHM framework. Meteorological forcings and hydrological states are propagated through the differentiable physical model, while selected process parameters are dynamically updated within predefined physically feasible ranges. The governing hydrological equations impose hard physical constraints on model states and fluxes, whereas parameter bounds provide soft constraints on the optimization space. The loss gradient with respect to the trainable parameters is computed through automatic differentiation and backpropagated through the differentiable physical equations to iteratively update the parameters until the convergence criterion is satisfied.

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

344 **Table 2. All the dynamic physical parameters and setting value ranges**

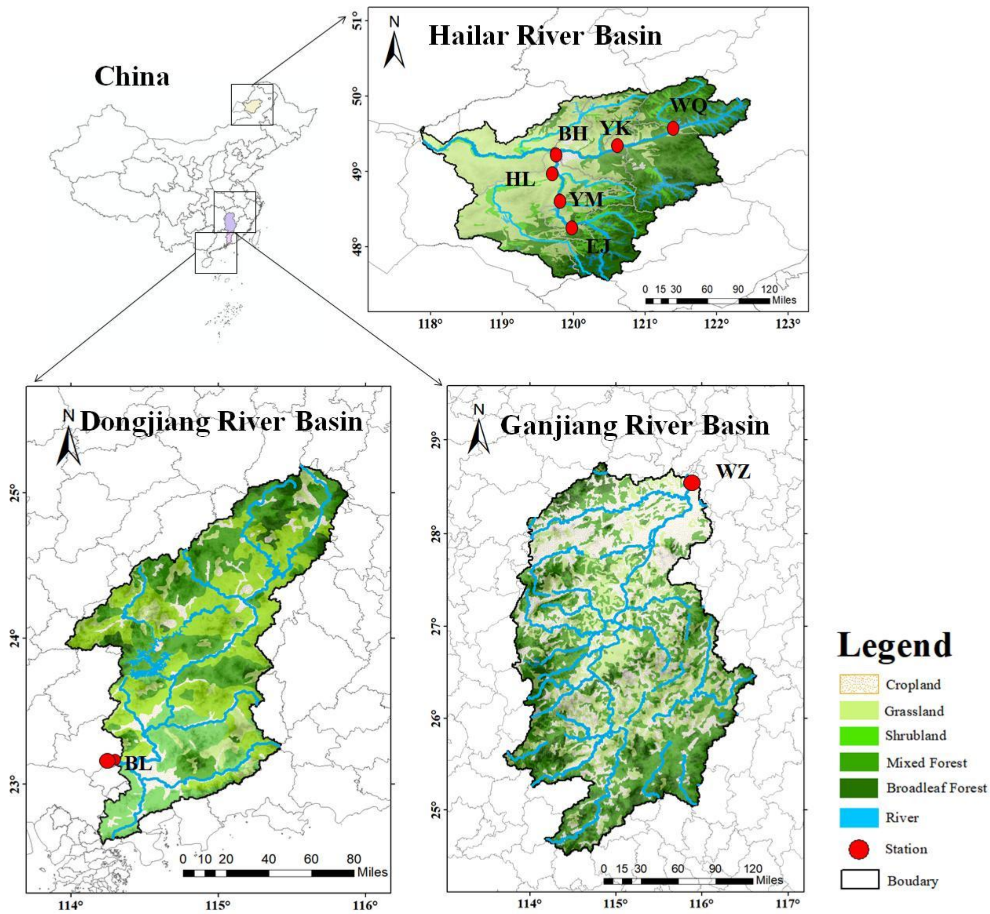
Parameter	Range	Definitions
T_{min}	[-3, 0]	Critical freezing temperature
ddf	[0,5]	Degree-day factor
T_{max}	[0, 3]	Snowmelt onset temperature
S_{max}	[0, 1,500]	Soil water storage capacity
r_s	[0, 500]	Stomatal resistance
f_{max}	[0, 50]	Infiltration capacity
b	[0, 5]	Soil water retention curve shape parameter
C_i	[0, 0.1]	Canopy interception coefficient
q_{max}	[0, 50]	Maximum baseflow rate
β	[0, 0.5]	Baseflow shape coefficient

345

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

346 **2.3. Experimental Design and Evaluation**

347 **2.3.1. Study Areas and Data Sources**



348

349 **Figure 4. Location map of the Hailar River Basin (HRB), Dongjiang River Basin (DRB)**
350 **and Ganjiang River (GRB).** The distributions of the vegetation types and hydrological
351 stations in the study area are shown, with different colors indicating various vegetation
352 categories, including forests, shrublands, grasslands, and croplands. Two basins are in

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

southern China and one in northern China, representing wet and dry climates, respectively. The runoff observation sites are BH, HL, YM, EJ, YK, WQ, BL, and WZ.

Table 3. Basin information for the Hailar River Basin (HRB), Dongjiang River Basin (DRB) and Ganjiang River Basin (GRB).

Station	HRB	DRB	GRB
Area	6.7×10^4	2.2×10^4	7.62×10^4
Vegetation	Grassland, forest	Forest	Forest, farmland
Temperature(℃)	-0.3	22.6	18.4
Precipitation(mm/year)	354.1	1,815.3	1,664.4
Station	BH	BL	WZ
Training year	2005-2014	2005-2009	2005-2014
Training data	1979-2004	1979-2004	1979-2004

The study was conducted in three contrasting river basins: the Hailar River Basin (HRB) in a semi-arid northern region, and the Dongjiang River Basin (DRB) and Ganjiang River Basin (GRB) in humid southern China (Figure 4, Table 3). The hydrological stations are the WZ station, the BL station and the BH station, respectively (Table 3). The vegetation types and areas across the three regions are diverse and exhibit distinct characteristics. Among the three basins, the largest basin, GRB, is 7.62×10^4 km^2 and is dominated by forest and farmland. The smallest basin, the Dongjiang River Basin, is 2.2×10^4 km^2 and is predominantly naturally forested. The Hailar River Basin is 6.7×10^4 km^2 and contains the

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

most varied vegetation types, including forests, shrubs, grasslands, and desertified grasslands. Moreover, the three basins also have different climate types and runoff generation mechanisms. The GRB and DRB are in humid areas, and the main runoff generation mechanism is saturation excess runoff, whereas the HRB is located in semiarid areas, and the main runoff generation mechanism is infiltration excess runoff.

We used daily meteorological data from the China Meteorological Forcing Dataset (CMFD) and daily runoff observations from January 1, 1979, to December 31, 2014. The CMFD was widely utilized in hydrological research throughout China. This dataset provides six variables at daily temporal resolution and 0.1°× 0.1° spatial resolution: near-surface air temperature, surface pressure, specific humidity, wind speed, downward shortwave radiation, and precipitation. The CMFD has been rigorously validated against station observations and demonstrates strength in representing precipitation phase partitioning and radiation budgets over complex terrain, making it suitable for driving process-based hydrological models in our study regions with diverse climatic conditions. To ensure the reliability of the model inputs and outputs, a systematic quality control procedure was applied to the raw hydrological data. Streamflow values deemed hydrologically implausible (e.g., anomalously high outlier in winter) were removed as outliers. These values are typically indicative of sensor errors or data logging mistakes. In the Dongjiang River Basin, due to abnormal precipitation recorded in the CMFD dataset, we evaluated model performance using daily runoff observation data from January 1, 2005, to December 31, 2009. We trained the model using 25 years of daily runoff observation data from January 1, 1979, to December 31, 2004, and evaluated its performance on an additional ten years of observation data spanning January 1, 2005, to December 31, 2014.

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

To validate the simulated evapotranspiration, we further utilized flux tower from the Hulunbuir Grassland Ecosystem National Field Scientific Observation and Research Station, located in the Xiertala Ranch of Hulunbuir City, Inner Mongolia (next to BH station). This flux dataset has been widely employed in ecohydrological studies of temperate grasslands in northern China (Yan et al., 2021).

2.3.2. Model Configurations

Another important issue is whether the added complexity of hybrid models is justified. Increasing model flexibility may improve outlet-flow performance, but it does not necessarily imply better process realism. A meaningful assessment should therefore compare the proposed model not only with pure deep learning models, but also with a traditional hydrological benchmark such as GR4J. To quantify the contribution of each innovation on runoff simulation capability, we designed five different configuration schemes (Figure 2) (1) Deep Learning model (DL model): pure deep learning RNN without physical information. (2) Hybrid-L model: The base PRNN module with lumped water-balance equations (following Jiang et al., 2020). configured with basic processes such as precipitation snowmelt and superinfiltration full runoff modules. (3) Hybrid-P model: a vegetation module is added on the basis of the hybrid-L model, including the vegetation transpiration process, vegetation infiltration process, etc. (4) Hybrid-P-D model: a fully distributed module is added on the basis of the hybrid-P model, which is divided into $0.1^\circ \times 0.1^\circ$ grids, and the 6 parameters of vegetation LAI, canopy height, maximum LAI, slope, surface elevation, minimum LAI, and the LAI change rate for each grid are input. (5) GR4J model: a traditional lumped conceptual rainfall-runoff model used as a low-complexity benchmark. The model was

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

calibrated using the same meteorological forcing and runoff observations as the hybrid models, with NSE as the objective function.

This progression enables tests that attribute performance gains to vegetation coupling and spatial distribution. The four models (the DL model, hybrid-L model, hybrid-P model, and hybrid-P-D model) share the same training parameters and loss function information and are trained 150 times. To verify the migration ability of the model in data-deficient basins, further research is necessary to determine whether models trained using basin exit points can effectively simulate hydrological processes at any untrained location within the same basin. We chose other runoff stations within the Hailar River Basin to validate the transferability of the hybrid model.

2.3.3. Evaluation Metrics

Model performance was evaluated using the Nash-Sutcliffe Efficiency (NSE), Kling-Gupta Efficiency (KGE), and Percent Bias (PBIAS). The statistical significance of performance differences between models was assessed using Meng's Z-test for correlated correlations (Meng et al., 1992; Nash and Sutcliffe, 1970, Legates and McCabe, 1999, Yilmaz et al., 2008).

NSE can be defined in the following general form:

$$NSE_{\alpha}=1-\frac{\sum_{i=1}^{T}\left|Q_{obs,i}-Q_{sim,i}\right|^{\alpha}}{\sum_{i=1}^{T}\left|Q_{obs,i}-\bar{Q}_{obs}\right|^{\alpha}} \tag{17}$$

where $Q_{obs,i}$ and $Q_{sim,i}$ are the simulated and observed runoff values, respectively, and where Q_{obs} is the average observed value. NSE measures the overall fit-of-goodness of the simulated and observed runoff data, with a NSE of 0.55 being the threshold for good performance (Knoben et al., 2019, Newman et al., 2015).

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

The Pbias is defined as follows:

$$Pbias=100\times\left|\frac{\sum_1^n(Q_{sim}-Q_{obs})}{\sum_1^n Q_{obs}}\right| \tag{18}$$

where Q_{sim} and Q_{obs} are the observed and simulated runoffs sorted in descending order, respectively.

Kling Gupta efficiency (KGE) is an indicator that comprehensively evaluates the difference between simulated model results and observed data (Gupta et al., 2009). KGE combines measures of correlation, bias, and variability. The calculation formula is as follows:

$$KGE=1-\sqrt{(r-1)^2+(\alpha-1)^2+(\beta-1)^2} \tag{19}$$

$$r=\frac{\sum_{i=1}^n(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum_{i=1}^n(X_i-\overline{X})^2\sum_{i=1}^n(Y_i-\overline{Y})^2}} \tag{20}$$

where the r Pearson correlation coefficient is used to measure the linear relationship between simulated values and observed values. $\alpha=\overline{X}/\overline{Y}$ represents the ratio of the average of the simulated and observed values, i.e., the ratio of the average of the simulated and observed values. $\beta=(\sigma_x/\overline{X})/(\sigma_y/\overline{Y})$ represents the ratio of the standard deviation σ_x, σ_y between the simulated and observed values, i.e., the ratio of the variability between the simulated and observed values. The closer the value is to 1, the better the model's performance in simulation is (Kling and Gupta, 2009).

The statistical significance of performance differences between models is assessed using Meng's Z-test for correlated correlations (Meng et al., 1992). Due to the interdependent (correlated) performance metrics, conventional tests like the t-test would be invalid here. Meng's test properly accounts for the high correlation between model forecasts, providing a more balanced and meaningful assessment of significance in the context of large hydrological

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

datasets. This method is particularly suitable for comparing dependent correlation coefficients (e.g., KGE) derived from the same dataset. The test statistic is computed as:

$$z=\frac{1}{2}In(\frac{1+r}{1-r})$$

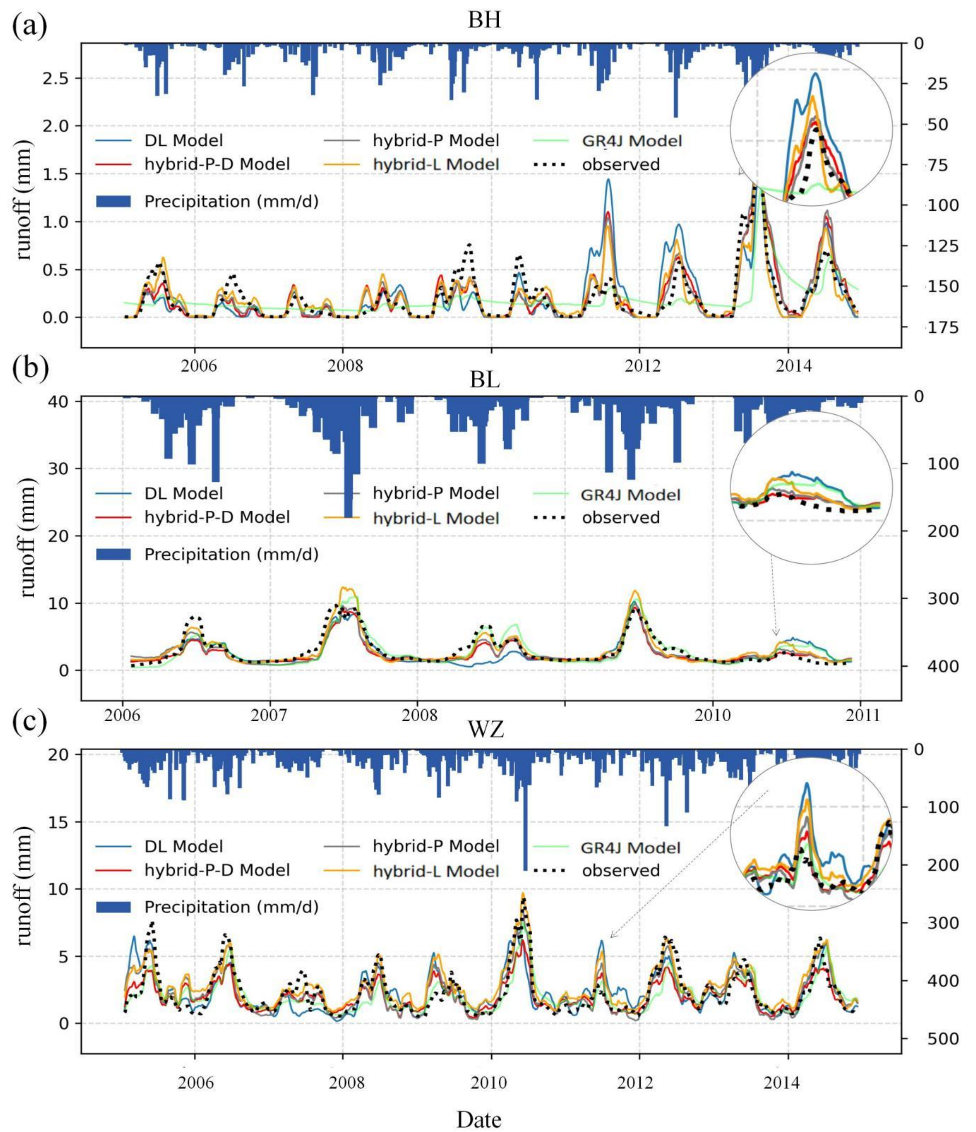
$$Z=(z_1-z_2)\times\sqrt{\frac{N-3}{2\times(1-r_{12})\times h}}$$

where: r is the Pearson correlation coefficient for a given model; z_1 and z_2 are Fisher's z-transformations of the correlation coefficients for Data A and Data B, respectively; N is the sample size; vr_{12} is the correlation between the two sets of data predictions; h is a correction factor accounting for the covariance structure

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

3. Results

3.1. Hybrid Model Evaluation Outlet Runoff in Trained Basins



https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

Figure 5. (a) (b) (c) Comparison of simulated and observed streamflows in the valid period. The black dots represent observations, the blue dots represent the DL model, the red dots represent the hybrid-P-D model, the gray dots represent the hybrid-P model, the yellow dots represent the hybrid-L model, the green dots represent the Physical model (GR4J) and the blue bars represent precipitation. Partial detail enlarged image on the right side.

Table 4. The results of 3 metrics (KGE, NSE and R^2) for five models (hybrid-L, hybrid-P, hybrid-P-D and DL models) in three study basins. Bold representation of the optimal model.

Basin	Station	Model	KGE	NSE	R^2
HRB	BH	hybrid-L	0.766	0.551	0.774
		hybrid-P	0.820	**0.655**	**0.823**
		hybrid-P-D	**0.821**	0.650	0.822
		GR4J	0.440	0.356	0.359
		DL	0.687	0.340	0.715
DRB	BL	hybrid-L	0.767	0.713	0.845
		hybrid-P	**0.785**	**0.736**	0.858
		hybrid-P-D	0.784	0.735	**0.858**
		GR4J	0.783	0.622	0.622
		DL	0.702	0.600	0.781
GRB	WZ	hybrid-L	0.617	0.491	0.706
		hybrid-P	0.616	0.528	0.727
		hybrid-P-D	**0.633**	**0.530**	**0.736**
		GR4J	0.548	0.508	0.521

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

	DL	0.607	0.379	0.644

Overall, compared with the DL method, the three hybrid models perform well in accurately capturing runoff with appropriate amplitudes and times in the three study basins (Figure 5). At the BH station, the KGE values of the hybrid-P-D, hybrid-P and hybrid-L models are 0.821, 0.820 and 0.766, respectively (Table 4). At the WZ station, the KGE values of the hybrid-P-D, hybrid-P and hybrid-L models are 0.633, 0.616 and 0.617, respectively. The GR4J model provides a useful conceptual hydrological benchmark, but its simulations generally show weaker capability. The comparison results show that the hybrid-P-D and hybrid-P models perform better. Hybrid-P-D has the most significant improvement in the WZ station simulations. At the BH and BL stations, the performance of the hybrid-P models is significantly improved, indicating that plant LSTM modules can further enhance model performance. This enhancement is manifested in the R^2 and KGE values, with a significant increase of 0.056 at the BH station.

Table 5. Statistical significance of performance differences among five models (hybrid-L, hybrid-P, hybrid-P-D, GR4J and DL models) based on Meng's Z-test.

Note: Significance level: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (**). Performance direction based on NSE metric.

Model	hybrid-L	hybrid-P	hybrid-P-D	DL	GR4J
hybrid-L	-	<0.001***	<0.001***	<0.001***	<0.001***
hybrid-P	<0.001***	-	0.044*	<0.001***	<0.001***

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

hybrid-P-D	<0.001***	0.044*	-		<0.001***
DL	<0.001***	<0.001***	<0.001***	-	<0.001***
GR4J	<0.001***	<0.001***	<0.001***	<0.001***	-

The pairwise statistical comparisons using Meng's Z-test revealed that all model performance differences were statistically significant ($p < 0.05$), confirming good overall skill relative to the pure DL benchmark (Table 5). Among them, The hybrid-P-D models demonstrated significantly superior performance, improving the NSE and R^2 values by 0.01-0.14. Its KGE spans 0.63-0.82, again surpassing the competing models. A comparison of the three models reveals that the plant LSTM module improves KGE most significantly at the BH station. The distribution module improved most significantly at the WZ station (Figure 5b). This difference can be attributed to the BH being situated in a semiarid vegetation-transition zone, where the plant LSTM module excels at capturing vegetation-water processes with high accuracy. The WZ station spans the largest area within this zone, and the distribution module performs better in large basins. A detailed picture reveals that hybrid-P-D models excel under high-runoff events, whereas all the models perform similarly during low-runoff periods. Finally, incorporating explicit spatial heterogeneity clearly benefits runoff simulations. When the distributed module is applied to the three study basins, the NSE ranges from 0.53 to 0.7—up to 0.14 higher than its lumped counterparts — demonstrating that spatially resolved vegetation and runoff mechanisms enhance the predictive accuracy without sacrificing robustness.

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

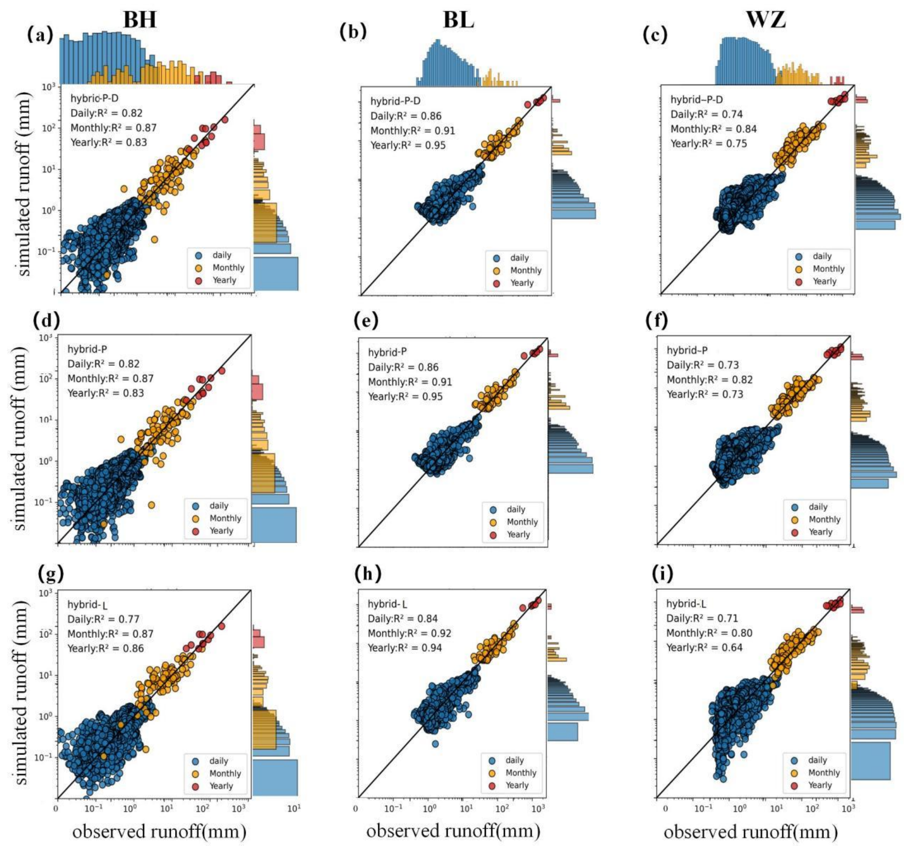


Figure 6. Scatter plots depicting the simulated and observed runoff at daily, monthly, and annual timescales for three models (hybrid-L, hybrid-P and hybrid-P-D) across three stations (BH, BL and WZ). The marginal histograms above and to the right of each scatter plot illustrate the frequency distributions of the observed (top) and simulated (right) runoff values at the three timescales. Each line represents the results for a specific model. Each column represents the results for a specific basin. The blue, yellow, and red markers denote daily, monthly, and yearly timescales, respectively. The solid 1:1 line indicates perfect agreement between the simulation and observation results.

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

3.2. Comparison of Model Seasonal Performances across Three Basins

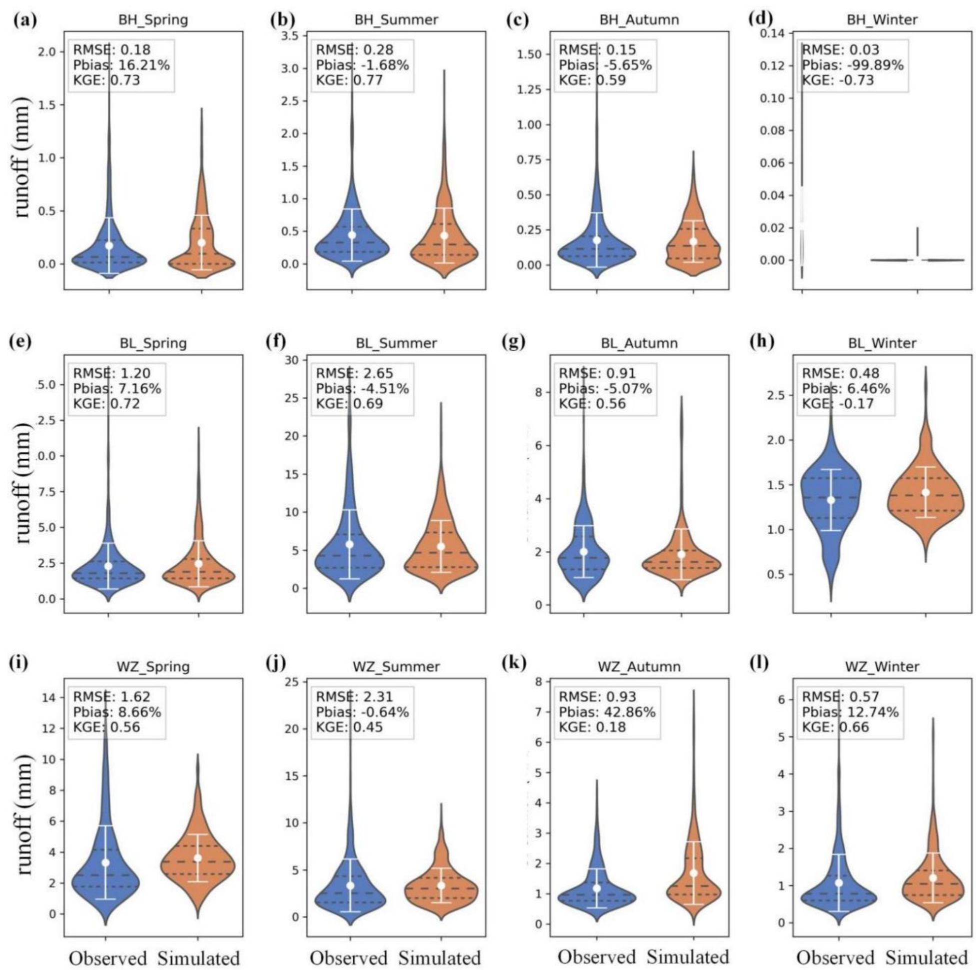


Figure 7. Violin plots comparing the distributions of observed (blue) and hybrid-P-D simulated (orange) runoff values for each season. From top to bottom: (a), (b), (c), and (d) are the spring, summer, autumn and winter seasons, respectively, at the BH station; (e), (f), (g), and (h) are the BL stations; and (i), (j), (k), and (l) are the WZ stations. Spring: March–May; Summer: June–August; Autumn: September–November; Winter: December–February. Each subplot displays statistical metrics, including the RMSE, NSE, and KGE, reflecting the hybrid-P-D model's performance in reproducing seasonal runoff.

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

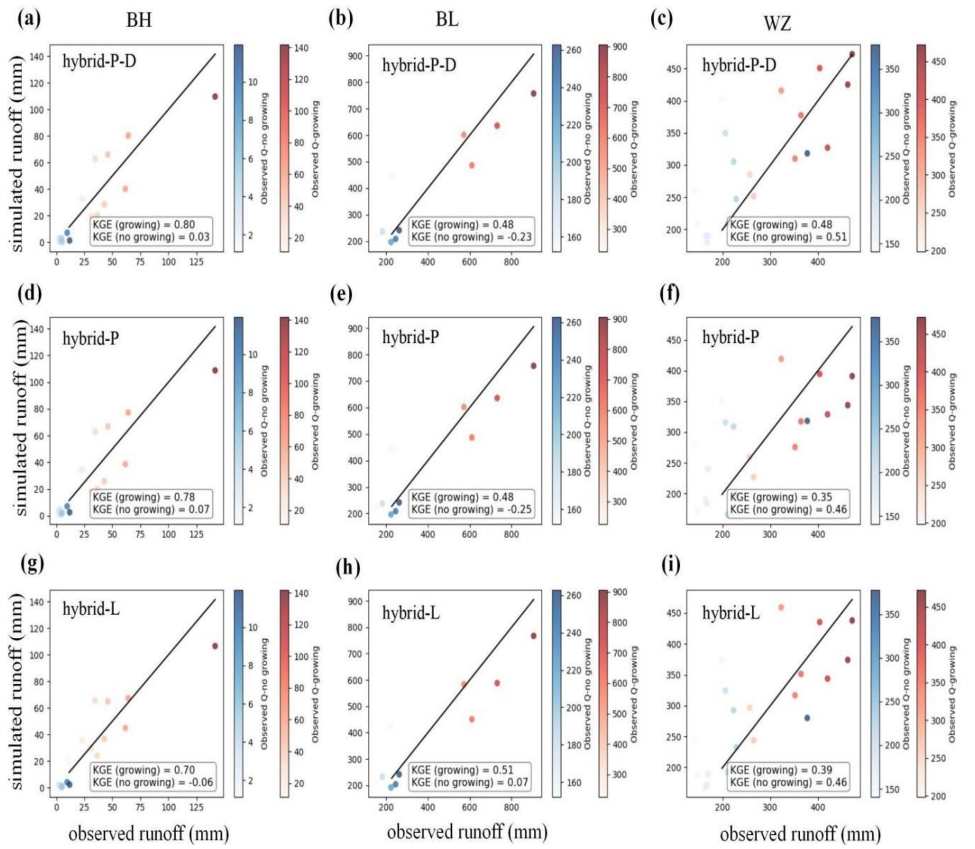


Figure 8. Scatter plot of the simulated and measured values for the growing season and nongrowing season. (a), (b), and (c) indicate the hybrid-P-D model; (d), (e), and (f) indicate the hybrid-P model; and (d), (e), and (f) indicate the hybrid-L model. Each line was evaluated at three stations (BH, BL and WZ). The growing season is from June to September, and the nongrowing season is from November to March of the following year. The blue color represents values during the nongrowing season, whereas the red color represents values during the growing season. Each point indicates the growing/nongrowing seasonally accumulated runoff, and the color intensity reflects the magnitude of the observed runoff.

For most stations, the "best" model of hybrid-P-D simulated distributions aligns with the observed distributions in terms of shape, mean, and variability during spring and summer. This is further supported by higher NSE and KGE values and lower RMSEs during these

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

seasons. However, the model performance declines during winter at certain sites, particularly at the BH station (Figure 7d). At this site, the simulated runoff distribution is notably wider and more skewed than the observed distribution, as shown by the broader and asymmetrical violin shape. This indicates that the model overestimates the frequency of low-runoff events during winter. These results suggest that the hybrid-P-D model tends to overestimate low runoff under cold and drought conditions, resulting in highly skewed distributions and reduced NSE and KGE values. Overall, while the model demonstrates better performance during high-runoff seasons, larger deviations are observed during low-runoff conditions, as reflected in both the violin plot shapes and the performance metrics.

The scatter plot of runoff during the growing season and nongrowing season in the basin with measured values illustrates the extent to which vegetation influences runoff. Figure 8 shows the simulation of the Q-growing/Q-no growing scatter plots using the three models. Although the KGE values for both the hybrid-P-D and hybrid-P models are lower in low streamflow regions (Figure 8b, 8e), this mainly reflects the difficulty of simulating base flow, where the true discharge variations are so small that they are masked by measurement error and natural fluctuations. In contrast, the performance of the hybrid-P-D model (Figure 8c) slightly surpasses that of the hybrid-P model (Figure 8f), with a KGE increase of 0.13 at the WZ station. The hybrid-P-D model simulation is notably better during the growing season, as reflected by substantially higher KGE values than those during the nongrowing season. These findings indicate that the hybrid-P-D model is more effective at capturing runoff dynamics during periods of high vegetation activity. Notably, the improvement at the BH station was the most pronounced, increasing from 0.70 to 0.80, whereas at the WZ station, it increased from 0.35 to 0.48. In contrast, the improvement at the BL station was negligible. These results suggest that the hybrid-P-D model is particularly effective in larger basins such as the WZ and in transitional vegetation zones such as the BH, where the interactions between vegetation and hydrology are more complex. Overall, the hybrid-P-D model outperforms the hybrid-P model in simulating Q-no growing.

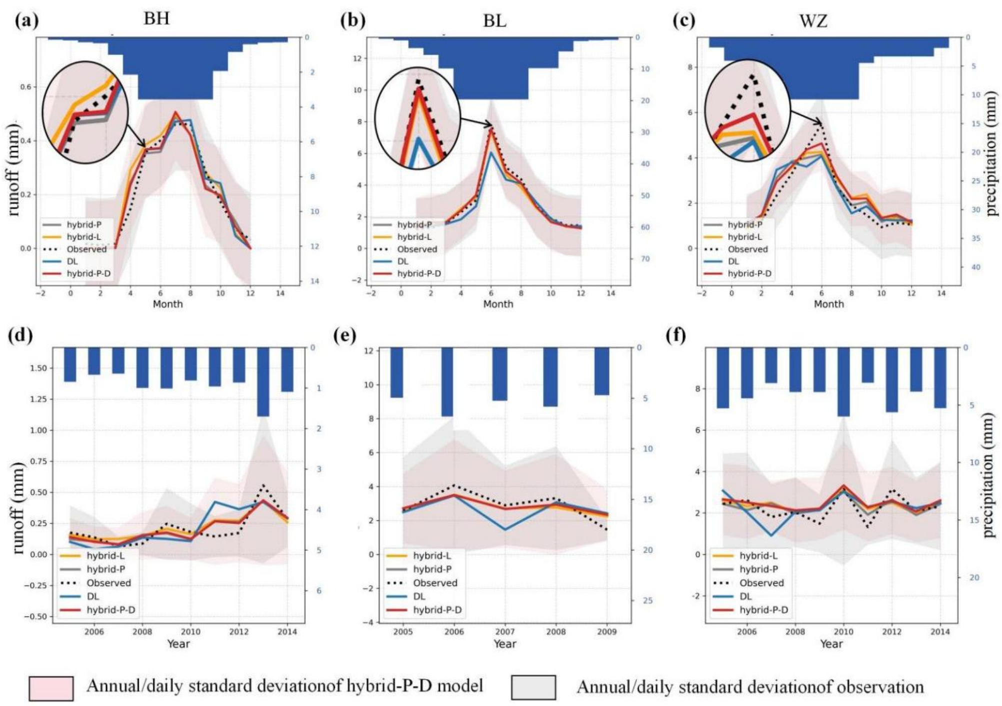


Figure 9. (a)(b)(c) Multiyear average monthly runoff simulations for four models (hybrid-L in yellow, hybrid-P in gray, hybrid-P-D in red, and DL in blue) compared with observations (black dashed line), illustrating each model's ability to capture monthly patterns. The red shaded area represents the variance in the hybrid-P-D simulation values, whereas the gray shaded area represents the variance in the hybrid-P simulation values. (d)(e)(f) Annual runoff simulations for the four models and observations, highlighting the interannual variability.

Because the impact of vegetation on runoff is often masked by meteorology, the simulation accuracy for monthly runoff was extracted and compared separately. Overall, Figure 9 indicates that the hybrid-P-D and hybrid-P models outperform the DL model in

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

capturing monthly scale hydrological dynamics. The hybrid model and deep learning model successfully simulate the bimodal phenomenon at the BH station (Figure 9a). However, when simulating flood peaks, it is evident that in hybrid models, DL models anticipate the timing of the spring flood peak in advance. Among the models, the hybrid-P-D model provides the best simulation of the spring peak occurrence time (Figure 9c). However, Figure 9d, 9e and 9f shows that the DL, hybrid-P-D, hybrid-P and hybrid-L models perform poorly when simulating drought years (2007) with low precipitation, especially at the WZ station. A significant overestimation was observed in 2007, 2009, and 2011, which were the years with the least precipitation. However, based on the distribution of variance in the shaded areas, the hybrid-P-D model has a smaller variance than the hybrid-P model does, indicating higher accuracy in simulating peak runoff events.

Figure 10 shows boxplots of the ratio of simulated to observed runoff (Q_{sim}/Q_{obs}) according to different precipitation and temperature ranges. Q_{sim}/Q_{obs} is compared from low to high to analyze the model's adaptability to climate change. The results indicate that the "best" model of hybrid-P-D demonstrates good adaptability to climate change at both the WZ and BL stations, with no significant trend observed in the ratio of Q_{sim}/Q_{obs} as temperature and precipitation increase. In contrast, at the BH station, the ratio clearly increases with increasing temperature and precipitation, suggesting that the hybrid-P-D model tends to overestimate runoff under these conditions. These findings highlight the model's differential response to varying climates, with the hybrid-P-D model exhibiting a stronger tendency to overestimate runoff in dry and cold basins.

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

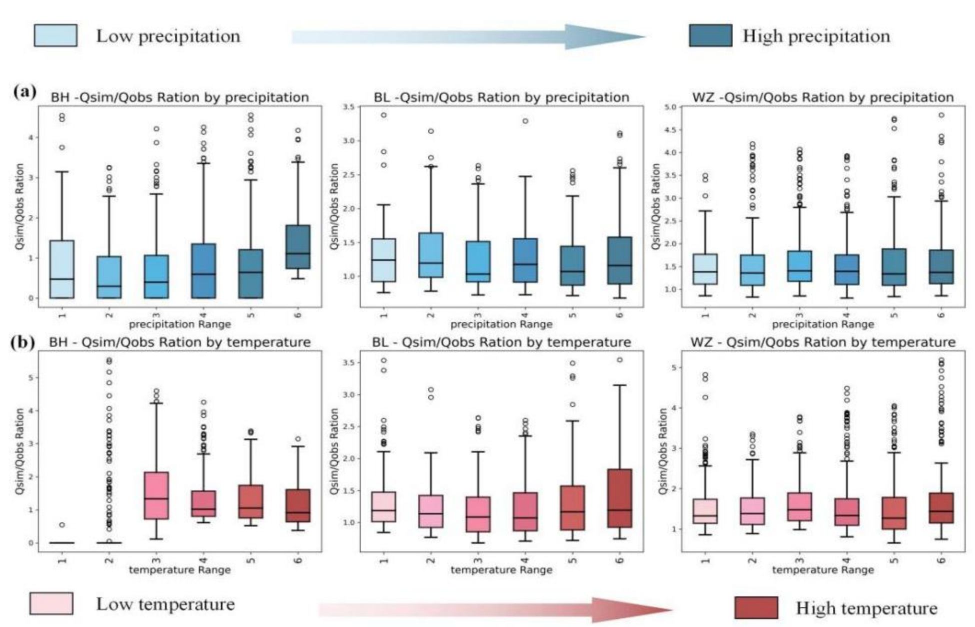


Figure 10. Boxplots of simulated and observed runoff under different conditions. (a) Boxplots of hybrid-P-D simulations of observed runoff (Q_{sim}/Q_{obs}) under different precipitation conditions (blue). It is classified into 6 equal parts to assess model performance across varying precipitation levels. The binned precipitation is classified as follows: for the BH site: <0.1, 0.1-0.2, 0.3-0.5, 0.6-1.3, 1.4-3.8, >3.9 mm/day; for the BL site: <0.1, 0.2-0.5, 0.6-2.0, 2.1-6.2, 6.3-16.2, >16.3 mm/day; for the WZ site: <0.1, 0.2-0.4, 0.5-1.4, 1.5-4.3, 4.4-12.4, >12.5 mm/day. **(b) Boxplots of simulated and observed runoff under different temperature conditions (red).** The precipitation and temperature gradually increase from light to dark colors. The binned temperature is classified as follows: for the BH site: < –18.0, –18.1- –4.7, –4.8-8.5, 8.6-15.7, 15.8-19.4, >19.5°C; for the BL site: <17.6, 17.7-22.8, 22.9-26.0, 26.1-27.4, 27.5-29.1, >29.2°C; for the WZ site: <8.1, 8.2-14.7, 14.8-20.7, 20.8-24.8, 24.9-28.7, >28.8°C.

3.3. Evaluating Model Transferability in Nested Subbasins

To evaluate the model's transferability, we applied the parameters learned at the basin outlet (BH) directly to five nested catchments—HL, EJ, YK, YM, and WQ—without further training (Figure 11). We used a hybrid-P-D configuration for simulation because it demands

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

minimal local data and has the best calibration performance at the BH. A pure DL model with identical hybrid-P-D model settings and training iterations was also employed for comparison.

The results show that hybrid-P-D achieved determination coefficients (R^2) between 0.56 and 0.76 across the five basins. Its PBIAS values were closer to zero than those of the DL models were, indicating substantially lower systematic error. The results demonstrated that the hybrid-P-D model, trained on data from basin outlet stations, performed impressively in simulating the runoff process for nested basins. As shown in Figure 11, the hybrid-P-D model also delivered a higher NSE than did the model based on the pure DL model, with the DL model suffering the most pronounced performance degradation when transferred to ungauged locations. Table 6 reveals that hybrid-P-D did not consistently outperform DL in terms of the correlation coefficient R^2. Instead, the key advantages of hybrid-P-D lie in the KGE and Pbias — it substantially reduces systematic error, with Pbias values much closer to zero than those of DL across all five transfer basins. However, the EJ station presented the lowest correlation coefficient of 0.56. This can be attributed to the influence of human activities on the hydrological processes in the basin above EJ station, such as the Honghuaerji Reservoir, which plays a significant role in the runoff process. These activities complicate hydrological simulation, leading to a decrease in model accuracy. To test spatial transferability, we applied the parameters calibrated at the basin outlet (BH) directly to five nested subcatchments—HL, EJ, YK, YM, and WQ — without any retraining. Under identical training conditions, the hybrid-P-D model consistently outperformed the pure DL benchmark: R^2 increased to 0.56-0.76, KGE increased by up to 0.85, R^2 increased 0.07, and absolute Pbias decreased by 11.65% at the midstream and upstream stations (EJ, YK, WQ). These gains demonstrate that hybrid-P-D can be generalized across space while requiring less calibration data.

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

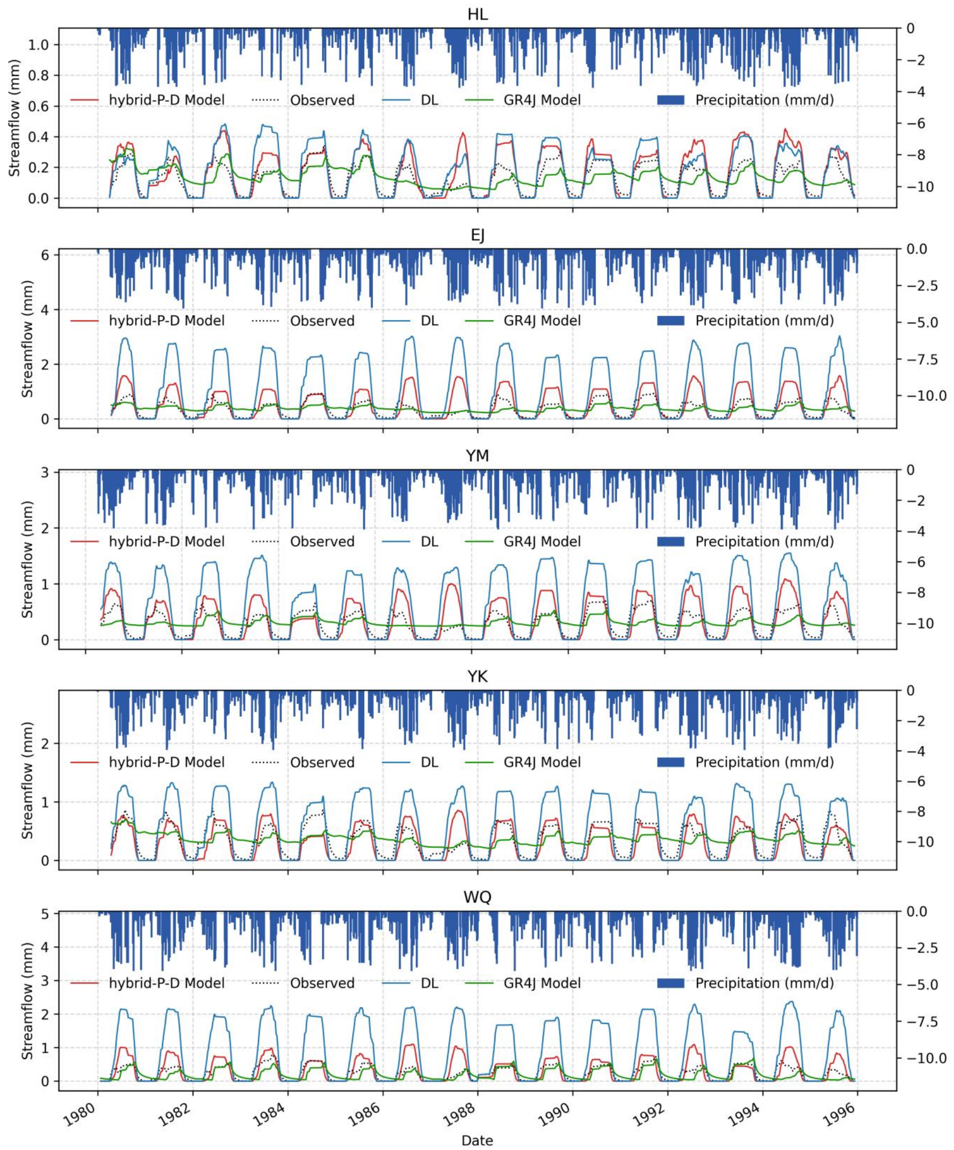


Figure 11. Comparison of simulated (hybrid-P-D, GR4J and DL models) and observed runoff during the valid period at the HL, EJ, YM, YK, and WQ stations in the HRB. The black dashed curve represents the observations, the blue line represents the DL model, the green line represents the GR4J model, the orange line represents the hybrid-P-D model, and the blue bars represent precipitation.

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

Table 6. Comparison of hybrid model performance in five transfer basins. Bold represents the optimal model.

Station	Model	R^2	Pbias (%)	KGE
HL	hybrid-P-D	0.77	**37.98**	**0.42**
	GR4J	**0.04**	7.94	0.17
	DL	**0.82**	43.42	0.33
EJ	hybrid-P-D	0.57	**37.35**	**0.35**
	GR4J	0.09	-5.46	0.05
	DL	**0.59**	172.30	-1.34
YM	hybrid-P-D	0.62	**15.86**	**0.58**
	GR4J	0.10	-11.48	0.04
	DL	**0.65**	93.79	-0.21
YK	hybrid-P-D	0.60	**-17.89**	**0.52**
	GR4J	0.04	-1.14	0.03
	DL	**0.66**	49.35	0.33
WQ	hybrid-P-D	0.61	**-7.35**	**0.49**
	GR4J	0.41	-33.95	0.39
	DL	**0.63**	127.59	-0.38

3.4 Spatial and temporal validation of evapotranspiration

To verify whether the improved runoff predictions are physically consistent, we compared model-simulated evapotranspiration (ET) against satellite-based ET products (PML-V2) for the HRB, both spatially and temporally. Figure 12 presents the spatial

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

distribution of mean annual ET from the Hybrid-P-D model, the PML-V2 reference, and their difference. The Hybrid-P-D model reproduces the observed spatial gradient of higher ET in the downstream forested areas and lower ET in the upstream grasslands, with a pixel-wise RMSE = 45.5 mm yr^{-1}. The figure reveals slight underestimation in forest-grassland transition zones. Despite these localized discrepancies, the overall bias is evenly distributed, with no significant spatial clustering. This spatial consistency confirms that the distributed physical constraints translate land-surface heterogeneity into realistic ET patterns.

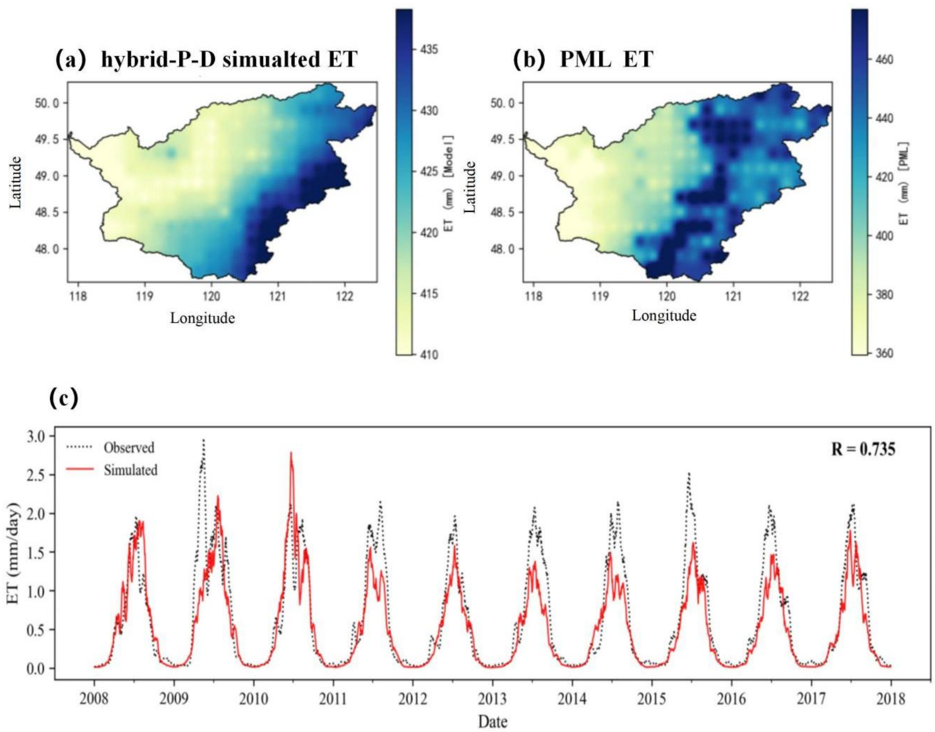


Figure 12. Spatial and temporal validation of evapotranspiration in the HRB. (a) Hybrid-P-D simulated annual ET; (b) satellite-based annual ET (PML-V2); (c) Temporal comparison of daily evapotranspiration between Hybrid-P-D simulation and observations at the Hulunbuir flux tower in the HRB

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

Figure 12(c) compares the temporal dynamics of daily ET between the Hybrid-P-D simulation and flux tower observations over the validation period. The model captures the seasonal cycle and interannual variability well, with a correlation coefficient of R = 0.735 between simulated and observed values. It reproduces the ET peak during the growing season and the low ET during winter dormancy, indicating that the dynamic vegetation module effectively tracks phenological water-use patterns.

4. Discussion

4.1. Comparison of Model Performance with Different Module Improvements

Accurate runoff simulations under climate change, particularly in data-scarce regions, are critical for effective water resource management. Our three hybrid models integrate dynamic vegetation–hydrological interaction mechanisms with distributed grid mechanisms by coupling deep learning and physical approaches. A validation across three Chinese basins representing dryness–wetness gradients revealed that the hybrid-P and hybrid-P-D models performed best. Notably, the hybrid-P-D model demonstrated superior performance in large basins with complex and heterogeneous underlying surface conditions. In contrast, the hybrid-P model, which does not rely on topographic or vegetation data, exhibited comparable accuracy to the hybrid-P-D model in medium- and small-sized basins, rendering it more suitable for regions with limited data availability (Figure 5). Compared with purely process-based or data-driven approaches, the hybrid-P-D model improved the KGE from 0.68 (DL model) to 0.82 (Table 4). By comparing the hybrid-P and hybrid-L models, we found that the plant LSTM module increased the KGE from 0.70 to 0.78 during the plant growing season. This indicates that the vegetation module improves the model's ability to extrapolate to

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

ungauged basins and capture seasonal hydrological dynamics (Figure 11) (Li et al., 2022; A et al., 2024; Bassani et al., 2024).

Although the fully distributed hybrid-P-D model achieved the best overall performance in the semi-arid HRB, its advantage over the much simpler GR4J benchmark was limited in the humid DRB, where both models produced comparable KGE values (GR4J:KGE=0.783 vs Hybrid-P-D:KGE=0.784)(Table 4). This result suggests that the benefit of model complexity is context dependent rather than universally required. In humid basins with relatively stable runoff-generation conditions, a parsimonious conceptual model may already capture the dominant hydrological response. By contrast, in the semi-arid HRB, particularly during extreme low-flow periods such as the 2007 drought at the WZ station, the hybrid-P-D model showed a more evident improvement over GR4J. For the ungauged sub-basins, including HL, EJ, YM, YK, and WQ, the hybrid-P-D model was able to generate transferable simulations using parameters learned from the basin outlet, whereas GR4J cannot be directly applied without local calibration.

The performance gain stems from the vegetation module, which dynamically parameterizes vegetation regulation within a distributed grid-based framework. Surface evapotranspiration is modeled as a linear function of soil water content (*SWC*) coupled with a dynamic stomatal resistance function (r_s), allowing accurate quantification of vegetation transpiration responses under water stress. This approach effectively mitigates the underestimation of runoff across multiple time scales (Figure 6) (Wu et al., 2024). The model incorporates the spatial heterogeneity of vegetation types (e.g., LAI and canopy height) within a 1×1 km^2 grid, improving flood peak simulations and peak timing (Figure 5). Such

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

improvements are critical for accurately simulating the impacts of vegetation on hydrological cycles across various land covers and climate regimes.

One of the key advantages of the hybrid-P-D model lies in its superior transferability compared with traditional DL models (Figure 11). In regions where the soil depth and climate closely resemble those in the training domain, the model performance has a correlation coefficient (R^2) greater than 0.76. This enhanced transferability of the hybrid-P-D model arises from its explicit integration of physical constraints and hydrological knowledge. Unlike standard DL models, which rely solely on training data and often struggle when applied to unseen environments, the hybrid-P-D model reduces this reliance by embedding physical equations, such as dynamic infiltration capacity. These embedded processes serve as internal priors that constrain model outputs within hydrologically reasonable ranges. As a result, the hybrid-P-D model can capture transferable process representations that remain valid across diverse climatic and ecohydrological regimes (Bindas et al., 2024; Feng et al., 2022). To further improve the robustness and generalizability of hybrid models across climatic gradients, we recommend extending the training durations and including larger catchments in future studies.

Independent ET evaluation provides evidence for the hydrological plausibility of the PIDL-DHM model. The comparison with PML-derived ET indicates that the model does not exhibit a systematic basin-wide bias, suggesting that the overall magnitude and large-scale spatial pattern of evapotranspiration are reasonably represented. However, relatively larger errors occur in the transitional zones, where climatic gradients, vegetation structure, and soil-moisture controls change more rapidly. These localized discrepancies imply that the current model can capture the first-order spatial pattern of ET, but still has limitations in

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

representing fine-scale ecohydrological transitions. Future improvements should therefore focus on strengthening the representation of spatially heterogeneous vegetation and soil-water processes, for example by incorporating remote-sensing-based dynamic vegetation structural parameters, canopy height, rooting depth, or soil-moisture constraints.

4.2. Model performance influenced by different climates

We validated the hybrid model across three diverse Chinese basins, which represent gradients of dryness, wetness, and vegetation types, to assess its adaptability to varying hydrological conditions. These basins, with distinct hydroclimatic characteristics, provide a comprehensive test of the model's robustness in different environmental contexts. The hybrid-P-D model exhibited strong adaptability to seasonal variations and dry–wet gradients, showing its ability to capture seasonal fluctuations in vegetation–water interactions. The simulations conducted for the spring, summer, and autumn seasons across these regions revealed that the simulated runoff distributions closely aligned with the observed values (Figure 8), reflecting good overall robustness and adaptability. These seasons correspond to peak root activity and transpiration, during which the dual-runoff generation mechanism module enhances simulation accuracy through soil moisture regulation (Figure 7). Additionally, the model was able to simulate winter runoff accurately in nonfrozen basins, such as the GRB and DRB, indicating its strong ability to handle seasonal shifts in runoff generation.

Additionally, our hybrid framework captures long-term trends remarkably well across most study basins. In dry years (such as 2007), low precipitation led to abnormal runoff fluctuations, and the DL models tended to underestimate runoff (Figure 9), whereas the hybrid-P-D model improved significantly. In the GRB and DRB basins, Q_{sim}/Q_{obs} maintains

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

stable performance under gradually increasing temperature and precipitation conditions, underscoring its ability to translate climatic signals into hydrological responses more faithfully than conventional data-driven approaches do (Figure 10). Nevertheless, limitations have emerged in semiarid grassland basins such as the HRB (Figure 8b). A significant positive trend was observed in the Q_{sim}/Q_{obs} with increasing temperature and precipitation, a pattern not observed in deep-rooted forest-dominated regions. This discrepancy suggests that the vegetation module may inadequately represent long-term vegetation–climate–runoff interactions in semiarid areas, potentially owing to the memory limitations of the RNN and LSTM architectures over extended time series (Du et al., 2024; Wei et al., 2024). Grasslands, characterized by shallow root systems and low water storage capacity, exhibit intensified evapotranspiration under drought, exacerbating water loss (Wang et al., 2024). In contrast, deep-rooted forests maintain stability during dry seasons by accessing deeper soil moisture, thereby delaying runoff decline and maintaining baseflow (Wang et al., 2023; Jiang et al., 2023). Explicit representations of dynamic rooting depth and moisture redistribution may help improve model performance for basins with contrasting vegetation, such as the HRB.

To clarify these complexities in runoff generation, we propose an improved runoff generation module that features dynamic switching between infiltration-excess (dry conditions) and saturation-excess (wet conditions) mechanisms (Jiang et al., 2023). This dual-mechanism approach allows the model to adaptively switch runoff generation mechanisms based on real-time environmental conditions (Liang et al., 1994). Specifically, parameters such as the infiltration curve shape factor (b) and maximum infiltration capacity (I_{max}) are dynamically adjusted according to changing climatic and soil conditions (Table 2). By dynamically modifying b and I_{max}, the model enables adaptive switching between infiltration-excess runoff during dry periods and saturation-excess runoff under wet

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

conditions. One key advantage of this approach is its ability to capture the "crust formation" of the soil surface, a phenomenon that often reduces infiltration capacity under climate stress. Traditional static runoff schemes frequently underestimate this process, which leads to inaccurate predictions, particularly under extreme climate conditions (Li et al., 2023; Patil et al., 2014). By integrating this dynamic process, the model provides a more accurate representation of the transitional dynamics between infiltration-excess and saturation-excess runoff mechanisms (Steenhuis et al., 2013).

Overall, integrating the dual runoff generalization mechanism into the hybrid model markedly enhances its ability to simulate key hydrological dynamics under climate change, particularly across moisture gradients, seasonal transitions, and vegetation-mediated responses. This improved simulation fidelity is not merely a statistical enhancement but holds significant practical value for increasing the reliability of seasonal water supply forecasts and refining climate change impact assessments in ecologically vulnerable regions. To further strengthen the model's utility, future work should focus on two key directions: (1) incorporating more detailed parameterizations of frozen soil and root depth dynamics to enhance physical realism, and (2) testing the generalizability of this framework by integrating it with other established hydrological models (e.g., SWAT, VIC) or applying it to diverse hydro-climatic regimes, such as arid basins or urbanizing catchments. These advancements will solidify the model's role as a robust tool for supporting adaptive water resource management and ecological protection strategies (Wang et al., 2024; Wang et al., 2023; Jiang et al., 2023).

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

5. Conclusions

This study develops a novel physics-informed deep learning framework that explicitly integrates vegetation physiological processes and a dual runoff generation mechanism within a fully distributed architecture. By coupling hydrological process representations with vegetation parameterization, the model captures vegetation-hydrology feedbacks and spatially heterogeneous runoff dynamics under changing climate conditions. The principal findings are as follows.

(1) According to multiple evaluation metrics and time scales, both the hybrid-P and hybrid-P-D models demonstrate relatively high accuracy (R^2 > 0.74 daily) and effectively capture the temporal changes in runoff during the growing season. The hybrid-P-D model performs better in large basins characterized by complex and heterogeneous surface conditions, whereas the hybrid-P model requires less terrain or vegetation data and achieves comparable accuracy in small- and medium-sized basins, making it more suitable for data-limited regions.

(2) The hybrid-P-D and hybrid-P model frameworks exhibit strong transferability and spatiotemporal generalizability by embedding unsaturated runoff processes. While the GR4J achieved comparable performance in humid basins, the hybrid models demonstrate advantages in more complex, data-scarce environments. The model more realistically captures the transitional dynamics between the two runoff mechanisms, requiring only downstream station data to simulate upstream and ungauged subbasin runoff satisfactorily. Augmenting training with additional upstream stations improves localized simulations, areas where depth and climate significantly deviate from the training domain.

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

(3) Spatial validation against PML-derived evapotranspiration indicates that the hybrid-P-D model does not exhibit a systematic basin-wide ET bias and can reasonably represent the overall magnitude and large-scale spatial pattern of evapotranspiration. However, This study also revealed the shortcomings of current models in transitional zones where climatic gradients, vegetation structure, and soil-moisture controls change rapidly. Future work should therefore incorporate more detailed parameterizations of dynamic rooting depth, canopy structure, frozen soil processes, and soil-moisture redistribution, as well as more diverse training datasets, to further improve model generalization under extreme climate events and seasonal edge conditions.

Future work should test the model on larger and more diverse datasets spanning broader climate and vegetation zones to rigorously evaluate its generalization ability, strengthening parameterization research on complex physical and physiological processes, such as freeze–thaw cycles, root depth dynamics, and soil microstructure evolution, in the interaction between vegetation and hydrology is necessary. Through these efforts, PIDL-DHM are expected to become important tools for water resource management and ecological protection in the context of climate change, providing solid scientific support for achieving sustainable basin management.

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

Credit Author Statement

Junping Wang: Data curation, Methodology, Writing - original draft. ; **Baolin Xue:** Conceptualization, Writing - review & editing, Funding acquisition; **Guoqiang Wang:** Project administration, Writing - review & editing, Supervision.

Source code availability: The source code is publicly available on Zenodo at record 18833916: https://zenodo.org/records/18833916. The archive includes the source code, model training scripts, and a pre-trained model checkpoint, along with scripts for data preprocessing and hydrological computations.

Data availability: The datasets used in this study are subject to confidentiality restrictions and cannot be publicly shared. However, example data (HLE_BH_data.csv and HLE_BH_surfacedata.csv) is available on Zenodo at record 18833916, : https://zenodo.org/records/18833916.

Acknowledgement

This work was supported by the National Science Fund for Distinguished Young Scholars (No. 52125901) and the National Natural Science Foundation of China (No. 52379002).

References

[1] A, Y.L., Wang, G.Q., Zhang, Q.Z., Wang, P.Z., Xue, B.L., Gao, Z.Y., Peng, Y.B., 2024. Water quality prediction based on an innovated physical and data driving hybrid model at basin scale. J Environ Inform. https://doi.org/10.3808/jei.202400510.

[2] Barger, N.N., Archer, S.R., Campbell, J.L., Huang, C., Morton, J.A., Knapp, A.K., 2011. Woody plant proliferation in north american drylands: a synthesis of impacts on ecosystem carbon balance. Journal of Geophysical Research. 116 (G4). https://doi.org/10.1029/2010JG001506.

[3] Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., Tian, Q., 2023. Accurate medium-range global weather forecasting with 3d neural networks. Nature (London). 619 (7970), 533-538. https://doi.org/10.1038/s41586-023-06185-3.

[4] Botterill, T.E., Mcmillan, H.K., 2023. Using machine learning to identify hydrologic signatures with an encoder-decoder framework. Water Resour Res. 59 (3). https://doi.org/10.1029/2022WR033091.

[5] Brient, F., 2020. Reducing uncertainties in climate projections with emergent constraints: concepts, examples and prospects. Adv Atmos Sci. 37 (1), 1-15. https://doi.org/10.1007/s00376-019-9140-8.

[6] Camarero, J.J., Pizarro, M., Gernandt, D.S., Gazol, A., 2024. Smaller conifers are more resilient to drought. Agric for Meteorol. 350, 109993. https://doi.org/10.1016/j.agrformet.2024.109993.

[7] Chen, A., Ricciuto, D., Mao, J., Wang, J., Lu, D., Meng, F., 2023. Improving e3sm land model photosynthesis parameterization via satellite sif, machine learning, and surrogate modeling. J Adv Model Earth Syst. 15 (4). https://doi.org/10.1029/2022MS003135.

[8] Chen, J., Qiu, B., Chakraborty, T.C., Miao, X., Cao, Y., Li, L., Zhao, S., Ni, Y., Tian, X., Qian, Y., Guo, W., 2025. Contrasting effects of urbanization on vegetation between the global south and global north. Nat Sustain. https://doi.org/10.1038/s41893-025-01520-0.

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

[9] Chen, S., Fu, Y.H., Wu, Z., Hao, F., Hao, Z., Guo, Y., Geng, X., Li, X., Zhang, X., Tang, J., Singh, V.P., Zhang, X., 2023. Informing the swat model with remote sensing detected vegetation phenology for improved modeling of ecohydrological processes. J Hydrol (Amst). 616, 128817. https://doi.org/10.1016/j.jhydrol.2022.128817.

[10] Dazzi, S., Physics-Informed Neural Networks for the Augmented System of Shallow Water Equations With Topography. Water Resources Research, 2024. 60(10): p. e2023WR036589.

[11] Duethmann, D., Blöschl, G., & Parajka, J. (2020). Why does a conceptual hydrological model fail to correctly predict discharge changes in response to climate change? Hydrology and Earth System Sciences, 24, 3493-3511. https://doi.org/10.5194/hess-24-3493-2020

[12] Du, Z., Yu, L., Yang, J., Xu, Y., Chen, B., Peng, S., Zhang, T., Fu, H., Harris, N., Gong, P., 2022. A global map of planting years of plantations. Sci Data. 9 (1). https://doi.org/10.1038/s41597-022-01260-2.

[13] Dunne, T., Black, R., 1970. Partial area contributions to storm runo in a small.

[14] Elghawi, R., Kraft, B., Reimers, C., Reichstein, M., Körner, M., Gentine, P., Winklerwinkler, A.J., 2023. Hybrid modeling of evapotranspiration: inferring stomatal and aerodynamic resistances using combined physics-based and machine learning. Environ Res Lett. 18 (3), 34039. https://doi.org/10.1088/1748-9326/acbbe0.

[15] Feng, D., Beck, H., Lawson, K., Shen, C., 2023. The suitability of differentiable, physics-informed machine learning hydrologic models for ungauged regions and climate change impact assessment. Hydrol Earth Syst Sci. 27 (12), 2357-2373. https://doi.org/10.5194/hess-27-2357-2023.

[16] Feng, D., Liu, J., Lawson, K., Shen, C., 2022a. Differentiable, learnable, regionalized process-based models with multiphysical outputs can approach state-of-the-art hydrologic prediction accuracy. Water Resour Res. 58 (10). https://doi.org/10.1029/2022WR032404.

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

[17] Feng, D., Liu, J., Lawson, K., Shen, C., 2022b. Differentiable, learnable, regionalized process-based models with multiphysical outputs can approach state-of-the-art hydrologic prediction accuracy. Water Resour Res. 58 (10). https://doi.org/10.1029/2022WR032404.

[18] Frame, J.M., Nair, T., Sunkara, V., Popien, P., Chakrabarti, S., Anderson, T., Leach, N.R., Doyle, C., Thomas, M., Tellman, B., 2024. Rapid inundation mapping using the us national water model, satellite observations, and a convolutional neural network. Geophys Res Lett. 51 (17). https://doi.org/10.1029/2024GL109424.

[19] Fukushima, K., Miyake, S., 1982. Neocognitron: a new algorithm for pattern recognition tolerant of deformations and shifts in position. Pattern Recognit. 15 (6), 455-469. https://doi.org/10.1016/0031-3203(82)90024-3.

[20] Gao, H., Han, C., Chen, R., Feng, Z., Wang, K., Fenicia, F., Savenije, H., 2022. Frozen soil hydrological modeling for a mountainous catchment northeast of the qinghai-tibet plateau. Hydrol Earth Syst Sci. 26 (15), 4187-4208. https://doi.org/10.5194/hess-26-4187-2022.

[21] Gaur, S., Drewry, D.T., 2024. Explainable machine learning for predicting stomatal conductance across multiple plant functional types. Agric for Meteorol. 350, 109955. https://doi.org/10.1016/j.agrformet.2024.109955.

[22] G. Arnold, J., N. Moriasi, D., W. Gassman, P., C. Abbaspour, K., J. White, M., Srinivasan, R., Santhi, C., D. Harmel, R., van Griensven, A., W. Van Liew, M., Kannan, N., K. Jha, M., 2012. Swat: model use, calibration, and validation. Trans Asabe. 55 (4), 1491-1508. https://doi.org/https://doi.org/10.13031/2013.42256.

[23] Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and nse performance criteria: implications for improving hydrological modelling. Journal of Hydrology (Amsterdam). 377 (1), 80-91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

[24] Hamman, J.J., Nijssen, B., Bohn, T.J., Gergel, D.R., Mao, Y., 2018. The variable infiltration capacity model version 5 (vic-5): infrastructure improvements for new applications

and reproducibility. Geosci Model Dev. 11 (8), 3481-3496. https://doi.org/10.5194/gmd-11-3481-2018.

[25] Höge, M., Scheidegger, A., Baity-Jesi, M., Albert, C., Fenicia, F., 2022. Improving hydrologic models for predictions and process understanding using neural odes. Hydrol Earth Syst Sci. 26 (19), 5085-5102. https://doi.org/10.5194/hess-26-5085-2022.

[26] Holtan, H.H., 1961. A concept for infiltration estimates in watershed engineering.

[27] Horton, R.E., 1933. The rôle of infiltration in the hydrologic cycle. Eos, Transactions American Geophysical Union. 14 (1), 446-460. https://doi.org/10.1029/TR014i001p00446.

[28] Huang, J., Yu, H., Guan, X., Wang, G., Guo, R., 2016. Accelerated dryland expansion under climate change. Nat Clim Chang. 6 (2), 166-171. https://doi.org/10.1038/nclimate2837.

[29] Ji, H. and Y. Song, et al. (2025). "Distinct hydrologic response patterns and trends worldwide revealed by physics-embedded learning." Nature Communications 16(1): 9169.

[30] Jiang, S., Zheng, Y., Solomatine, D., 2020. Improving AI system awareness of geoscience knowledge: symbiotic integration of physical approaches and deep learning. Geophys Res Lett. 47 (13). https://doi.org/10.1029/2020GL088229.

[31] Jiang, X., Wang, G., Wang, Y., Yao, J., Xue, B., A, Y., 2023. A hybrid framework for simulating actual evapotranspiration in data-deficient areas: a case study of the inner mongolia section of the yellow river basin. Remote Sens (Basel). 15 (9), 2234. https://doi.org/10.3390/rs15092234.

[32] Jingyi Bu, G.G.J.C., 2024. Dryland evapotranspiration from remote sensing solar-induced chlorophyll fluorescence: constraining an optimal stomatal model within a two-source energy balance model. Remote Sens Environ. (303). https://doi.org/10.1016/j.rse.2024.113999.

[33] Khoshkalam, Y., Rousseau, A.N., Rahmani, F., Shen, C., Abbasnezhadi, K., 2023. Applying transfer learning techniques to enhance the accuracy of runoff prediction produced by

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

long short-term memory networks with data integration. J Hydrol (Amst). 622, 129682. https://doi.org/10.1016/j.jhydrol.2023.129682.

[34] Konapala, G., Kao, S., Painter, S.L., Lu, D., Oak Ridge National Lab. Ornl, O.R.T.U., 2020. Machine learning assisted hybrid models can improve runoff simulation in diverse catchments across the conterminous us. Environ Res Lett. 15 (10), 104022. https://doi.org/10.1088/1748-9326/aba927.

[35] Koppa, A., Rains, D., Hulsman, P., Poyatos, R., Miralles, D.G., 2022. A deep learning-based hybrid model of global terrestrial evaporation. Nat Commun. 13 (1). https://doi.org/10.1038/s41467-022-29543-7.

[36] 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 (Amsterdam). 373 (3), 337-351. https://doi.org/10.1016/j.jhydrol.2009.04.031.

[37] Lawrence, D.M., Fisher, R.A., Koven, C.D., Oleson, K.W., Swenson, S.C., Bonan, G., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., Kluzek, E., Lawrence, P.J., Li, F., Li, H., Lombardozzi, D., Riley, W.J., Sacks, W.J., Shi, M., Vertenstein, M., Wieder, W.R., Xu, C., Ali, A.A., Badger, A.M., Bisht, G., van den Broeke, M., Brunke, M.A., Burns, S.P., Buzan, J., Clark, M., Craig, A., Dahlin, K., Drewniak, B., Fisher, J.B., Flanner, M., Fox, A.M., Gentine, P., Hoffman, F., Keppel Aleks, G., Knox, R., Kumar, S., Lenaerts, J., Leung, L.R., Lipscomb, W.H., Lu, Y., Pandey, A., Pelletier, J.D., Perket, J., Randerson, J.T., Ricciuto, D.M., Sanderson, B.M., Slater, A., Subin, Z.M., Tang, J., Thomas, R.Q., Val Martin, M., Zeng, X., 2019. The community land model version 5: description of new features, benchmarking, and impact of forcing uncertainty. J Adv Model Earth Syst. 11 (12), 4245-4287. https://doi.org/10.1029/2018MS001583.

[38] Lees, T., Reece, S., Kratzert, F., Klotz, D., Gauch, M., De Bruijn, J., Kumar Sahu, R., Greve, P., Slater, L., Dadson, S.J., 2022. Hydrological concept formation inside long short-term

memory (LSTM) networks. Hydrol Earth Syst Sci. 26 (12), 3079-3101. https://doi.org/10.5194/hess-26-3079-2022.

[39] Li, B., Sun, T., Tian, F., Ni, G., 2023. Enhancing process-based hydrological models with embedded neural networks: a hybrid approach. J Hydrol (Amst). 625, 130107. https://doi.org/10.1016/j.jhydrol.2023.130107.

[40] Li, B., Sun, T., Tian, F., Tudaji, M., Qin, L., Ni, G., 2024. Hybrid hydrological modeling for large alpine basins: a semi-distributed approach. Hydrol Earth Syst Sci. 28 (20), 4521-4538. https://doi.org/10.5194/hess-28-4521-2024.

[41] Li, H., Leung, L.R., Getirana, A., Huang, M., Wu, H., Xu, Y., Guo, J., Voisin, N., 2015. Evaluating global runoff simulations by a physically based routing model coupled with the community land model. J Hydrometeorol. 16 (2), 948-971. https://doi.org/10.1175/JHM-D-14-0079.1.

[42] Li, Q., Gao, X., Li, J., Yan, A., Chang, S., Song, X., Lo, K., 2024. Nonlinear time effects of vegetation response to climate change: evidence from qilian mountain national park in china. Sci Total Environ. 933, 173149. https://doi.org/10.1016/j.scitotenv.2024.173149.

[43] Li, W., Migliavacca, M., Forkel, M., Denissen, J.M.C., Reichstein, M., Yang, H., Duveiller, G., Weber, U., Orth, R., 2022. Widespread increasing vegetation sensitivity to soil moisture. Nat Commun. 13 (1). https://doi.org/10.1038/s41467-022-31667-9.

[44] Li, Y., Piao, S., Li, L., Chen, A., Wang, X., Ciais, P., Huang, L., Lian, X., Peng, S., Zeng, Z., Wang, K., Zhou, L., 2018. Divergent hydrological response to large-scale afforestation and vegetation greening in china. Sci Adv. 4 (5), eaar4182. https://doi.org/10.1126/sciadv.aar4182.

[45] Liu, M., Peñuelas, J., Trugman, A.T., Vargas G, G., Yang, L., Anderegg, W.R.L., 2025. Diverging responses of terrestrial ecosystems to water stress after disturbances. Nat Clim Chang. 15 (1), 73-79. https://doi.org/10.1038/s41558-024-02191-z.

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

[46] Liu, Y., Kumar, M., Katul, G.G., Feng, X., Konings, A.G., 2020. Plant hydraulics accentuates the effect of atmospheric moisture stress on transpiration. Nat Clim Chang. 10 (7), 691-695. https://doi.org/10.1038/s41558-020-0781-5.

[47] Luo, Q., Yang, K., Chen, Y., Zhou, X., 2020. Method development for estimating soil organic carbon content in an alpine region using soil moisture data. Science China Earth Sciences. 63 (4), 591-601. https://doi.org/10.1007/s11430-019-9554-8.

[48] Ma, K., Feng, D., Lawson, K., Tsai, W.P., Liang, C., Huang, X., Sharma, A., Shen, C., 2021. Transferring hydrologic data across continents - leveraging data-rich regions to improve hydrologic prediction in data-sparse regions. Water Resour Res. 57 (5). https://doi.org/10.1029/2020WR028600.

[49] Meng, X.-L., Rosenthal, R. and Rubin, D.B. (1992) Comparing Correlated Correlation Coefficients. Psychological Bulletin, 111, 172-175. https://doi.org/10.1037/0033-2909.111.1.172

[50] Medlyn, B.E., Barton, C., Broadmeadow, M., Ceulemans, R., De Angelis, P., Forstreuter, M., Freeman, M., Jackson, S.B., Kellomaki, S., Laitat, E., Rey, A., Roberntz, P., Sigurdsson, B.D., Strassemeyer, J., Wang, K., Curtis, P.S., Jarvis, P.G., 2001. Stomatal conductance of forest species after long-term exposure to elevated CO_2 concentration: a synthesis. New Phytol. 149 (2), 247-264. https://doi.org/10.1046/j.1469-8137.2001.00028.x.

[51] Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part discussion of principles. J Hydrol (Amst). 10 (3), 282-290. https://doi.org/10.1016/0022-1694(70)90255-6.

[52] Niu, G.Y., Yang, Z.L., Dickinson, R.E., Gulden, L.E., 2005. A simple topmodel-based runoff parameterization (simtop) for use in global climate models. Journal of Geophysical Research: Atmospheres. 110 (D21). https://doi.org/10.1029/2005JD006111.

[53] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). PyTorch: An imperative style, high-performance deep learning library. In Proceedings of the 33rd

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

International Conference on neural information processing systems. Curran Associates Inc. Article 721.

[54] Patil, S., & Stieglitz, M. (2014). Modelling daily streamflow at ungauged catchments: What information is necessary? Hydrological Processes, 28(3), 1159-1169. https://doi.org/10.1002/hyp.9660

[55] Penman, H.L., Keen, B.A., 1997. Natural evaporation from open water, bare soil and grass. Proceedings of the Royal Society of London. Series a. Mathematical and Physical Sciences. 193 (1032), 120-145. https://doi.org/10.1098/rspa.1948.0037.

[56] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533-536. https://doi.org/10.1038/323533a0

[57] S., J.P., Q., Y., 2010. A survey on transfer learning. Ieee Trans Knowl Data Eng. 22 (10), 1345-1359. https://doi.org/10.1109/TKDE.2009.191.

[58] Shang, K., Yao, Y., Di, Z., Jia, K., Zhang, X., Fisher, J.B., Chen, J., Guo, X., Yang, J., Yu, R., Xie, Z., Liu, L., Ning, J., Zhang, L., 2023. Coupling physical constraints with machine learning for satellite-derived evapotranspiration of the tibetan plateau. Remote Sens Environ. 289, 113519. https://doi.org/10.1016/j.rse.2023.113519.

[59] Shen, C., Laloy, E., Elshorbagy, A., Albert, A., Bales, J., Chang, F., Ganguly, S., Hsu, K., Kifer, D., Fang, Z., Fang, K., Li, D., Li, X., Tsai, W., 2018. Hess opinions: incubating deep-learning-powered hydrologic science advances as a community. Hydrol Earth Syst Sci. 22 (11), 5639-5656. https://doi.org/10.5194/hess-22-5639-2018.

[60] Singh, A., Kumar, S., Akula, S., Lawrence, D.M., Lombardozzi, D.L., 2020. Plant growth nullifies the effect of increased water-use efficiency on runoff under elevated CO_2 in the southeastern united states. Geophys Res Lett. 47 (4), n/a-n/a. https://doi.org/10.1029/2019GL086940.

[61] Sit, M., Demiray, B.Z., Xiang, Z., Ewing, G.J., Sermet, Y., Demir, I., 2020. A comprehensive review of deep learning applications in hydrology and water resources.

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

[62] Steenhuis, T.S., Hrnčíř, M., Poteau, D., Romero Luna, E.J., Tilahun, S.A., Caballero, L.A., Guzman, C.D., Stoof, C.R., šanda, M., Yitaferu, B., Císlerová, M., 2013. A saturated excess runoff pedotransfer function for vegetated watersheds. Vadose Zone J. 12 (4), vzj2013. https://doi.org/https://doi.org/10.2136/vzj2013.03.0060.

[63] Tian, Y., et al., Physics - Informed Neural Networks for Solving the Two - Dimensional Shallow Water Equations With Terrain Topography and Rainfall Source Terms. Water Resources Research, 2025. 61(9): p. e2025WR040052.

[64] Van Katwyk, P., Fox Kemper, B., Seroussi, H., Nowicki, S., Bergen, K.J., 2023. A variational lstm emulator of sea level contribution from the antarctic ice sheet. J Adv Model Earth Syst. 15 (12). https://doi.org/10.1029/2023MS003899.

[65] Vicente-Serrano, S.M., Miralles, D.G., Mcdowell, N., Brodribb, T., Domínguez-Castro, F., Leung, R., Koppa, A., 2022. The uncertain role of rising atmospheric CO_2 on global plant transpiration. Earth Sci Rev. 230, 104055. https://doi.org/10.1016/j.earscirev.2022.104055.

[66] Wang, J.: Integration of physics-informed deep learning with a distributed hydrological model (PIDL-DHM): A hybrid model for runoff simulation (v1), Zenodo [code and data], https://doi.org/10.5281/zenodo.18833916, 2026.

[67] Wang, C., Xu, J., Chen, Y., Bai, L., Chen, Z., 2018. A hybrid model to assess the impact of climate variability on runoff for an ungauged mountainous basin. Clim Dyn. 50 (7-8), 2829-2844. https://doi.org/10.1007/s00382-017-3775-x.

[68] Wang, D., Zeng, Z., 2024. Urgent need to improve modelled sensitivity of evaporation to vegetation change. Nature Water. https://doi.org/10.1038/s44221-024-00203-y.

[69] Wang, F., Zhang, J., Fonti, P., Sun, Q., Wang, Y., Zhang, F., Wang, Y., Yang, J., Gou, X., 2024. Effects of site aridity and species on stand transpiration in high-elevation dryland ecosystems. J Hydrol (Amst). 630, 130662. https://doi.org/10.1016/j.jhydrol.2024.130662.

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

[70] Wang, G., Wang, Y., Li, Y., Cheng, H., 2007. Influences of alpine ecosystem responses to climatic change on soil properties on the qinghai-tibet plateau, china. Catena (Amst). 70 (3), 506-514. https://doi.org/10.1016/j.catena.2007.01.001.

[71] Wang, P., Su, X., Zhou, Z., Wang, N., Liu, J., Zhu, B., 2023. Differential effects of soil texture and root traits on the spatial variability of soil infiltrability under natural revegetation in the loess plateau of china. Catena (Amst). 220, 106693. https://doi.org/10.1016/j.catena.2022.106693.

[72] Wang, P., Yao, J., Wang, G., Hao, F., Shrestha, S., Xue, B., Xie, G., Peng, Y., 2019. Exploring the application of artificial intelligence technology for identification of water pollution characteristics and tracing the source of water quality pollutants. Sci Total Environ. 693, 133440. https://doi.org/10.1016/j.scitotenv.2019.07.246.

[73] Wang, X., Xiao, J., Li, X., Cheng, G., Ma, M., Zhu, G., Altaf Arain, M., Andrew Black, T., Jassal, R.S., 2019. No trends in spring and autumn phenology during the global warming hiatus. Nat Commun. 10 (1). https://doi.org/10.1038/s41467-019-10235-8.

[74] Wi, S., Steinschneider, S., 2022. Assessing the physical realism of deep learning hydrologic model projections under climate change. Water Resour Res. 58 (9). https://doi.org/10.1029/2022WR032123.

[75] Xie, K., Liu, P., Zhang, J., Han, D., Wang, G., Shen, C., 2021. Physics-guided deep learning for rainfall-runoff modeling by considering extreme events and monotonic relationships. J Hydrol (Amst). 603, 127043. https://doi.org/10.1016/j.jhydrol.2021.127043.

[76] Xue, B., A, Y., Wang, G., Helman, D., Sun, G., Tao, S., Liu, T., Yan, D., Zhao, T., Zhang, H., Chen, L., Sun, W., Xiao, J., 2022. Divergent hydrological responses to forest expansion in dry and wet basins of china: implications for future afforestation planning. Water Resour Res. 58 (5). https://doi.org/10.1029/2021WR031856.

https://doi.org/10.5194/egusphere-2026-5146
Preprint. Discussion started: 15 September 2026

[77] Xue, J., Huete, A., Liu, Z., Wang, Y., Lu, X., 2024. Estimation of global ecosystem isohydricity from solar-induced chlorophyll fluorescence and meteorological datasets. Remote Sens Environ. 307, 114168. https://doi.org/10.1016/j.rse.2024.114168.

[78] Yan, R.R., Chen, B.R., Zhang, B.H., et al. Dataset of plant community composition dynamics in the Stipa baicalensis meadow steppe of Hulunbuir from 2009 to 2015. Science Data Bank, 2021. https://doi.org/10.11922/sciencedb.j00001.00243.

[79] Yao, L., Zhang, J., Cao, C., Zheng, F., 2025. Parameter estimation and uncertainty quantification of rainfall-runoff models using data assimilation methods based on deep learning and local ensemble updates. Environ Model Softw. 185, 106332. https://doi.org/10.1016/j.envsoft.2025.106332.

[80] Zhang, B., Ouyang, C., Cui, P., Xu, Q., Wang, D., Zhang, F., Li, Z., Fan, L., Lovati, M., Liu, Y., Zhang, Q., 2024. Deep learning for cross-region runoff and flood forecasting at a global scale. Innovation. 5 (3), 100617. https://doi.org/10.1016/j.xinn.2024.100617.

[81] Zhang, J., Zhang, Y., Sun, G., Song, C., Li, J., Hao, L., Liu, N., 2022. Climate variability masked greening effects on water yield in the yangtze river basin during 2001-2018. Water Resour Res. 58 (1). https://doi.org/10.1029/2021WR030382.

[82] Zhang, X., Zhang, Y., Tian, J., Ma, N., Wang, Y., 2022. CO_2 fertilization is spatially distinct from stomatal conductance reduction in controlling ecosystem water-use efficiency increase. Environ Res Lett. 17 (5), 54048. https://doi.org/10.1088/1748-9326/ac6c9c.

[83] Zhong, L., Lei, H., Gao, B., 2023. Developing a physics-informed deep learning model to simulate runoff response to climate change in alpine catchments. Water Resour Res. 59 (6). https://doi.org/10.1029/2022WR034118.

[84] Zhong, L., Lei, H., Yang, J., 2024. Development of a distributed physics-informed deep learning hydrological model for data-scarce regions. Water Resour Res. 60 (6). https://doi.org/10.1029/2023WR036333.