https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

From prediction improvement to uncertainty propagation structure: revisiting hybrid hydrological models

Bu Li[1], Ting Sun[2], Fuqiang Tian[1], Guangheng Ni[1]

[1]State Key Laboratory of Hydro-science and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing 100084, China
[2] Institute for Risk and Disaster Reduction, University College London, London WC1E 6BT, UK

Correspondence to: G. Ni (ghni@tsinghua.edu.cn)

Abstract. Hybrid hydrological models integrating embedded neural networks (ENNs) have demonstrated strong potential for improving streamflow prediction. However, how uncertainties induced by ENNs propagate through hydrological processes and internal system states remains poorly understood. This limits a mechanistic understanding of the stability and reliability of hybrid hydrological systems. This study therefore moves beyond performance evaluation and focuses on the internal uncertainty propagation structure of fully coupled hybrid hydrological systems. We investigate the internal uncertainty propagation mechanisms of hybrid hydrological models with different levels of ENN embedding across 31 cold-region basins. An ensemble-based framework driven by stochastic optimization was used to quantify uncertainty across hydrological fluxes and internal state variables. Multiple complementary metrics were employed to characterize uncertainty magnitude, ensemble consistency, coverage and spread. Results show that ENNs improve streamflow simulation performance but alter the internal uncertainty structure of hydrological systems. Uncertainty amplification is primarily localized within ENN-replaced hydrological processes, while propagation to physically based unreplaced processes remains limited. Instead, internal state variables with memory effects act as key “uncertainty reservoirs”, where uncertainty accumulates over time. This reveals a structured rather than uniform pattern of uncertainty propagation in hybrid hydrological systems. Furthermore, increased ensemble coverage is accompanied by wider uncertainty bands, indicating a trade-off between predictive reliability and sharpness induced by stochastic optimization. Overall, this study provides new insights into how ENNs reshape uncertainty propagation pathways in hydrological models. The results highlight the importance of jointly considering predictive performance, uncertainty structure and internal system stability in hydrological modeling.

1 Introduction

Accurate hydrological simulation plays an essential role in understanding hydrological dynamics and supporting applications such as flood forecasting and water resources management (Blöschl et al., 2019; Solanki et al., 2025; Thébault et al., 2026). Physical hydrological models have been widely developed and applied for decades (Cui et al., 2023; Li et al., 2021). By explicitly describing hydrological mechanisms, these models provide valuable insights into the functioning of hydrological

https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

systems (Singh et al., 2023; Wang et al., 2026). Nevertheless, their predictive capability can be constrained by uncertainties arising from model structures, parameterization schemes, and scale representation issues (Nearing et al., 2021; Renard et al., 2010). Deep learning (DL) approaches have shown remarkable capability in identifying complex nonlinear relationships between meteorological inputs and hydrological responses, leading to their increasing application in hydrological modelling (Kratzert et al., 2018; Li, Li, et al., 2023; Nearing et al., 2021; Shen & Lawson, 2021). The limited transparency of DL-based representations and their weak connection with established hydrological knowledge have raised concerns regarding their interpretability and broader application in hydrological science (Yihan Wang et al., 2025; Zhang et al., 2025). Hybrid hydrological models that combine process-based formulations with DL techniques have therefore emerged as an alternative approach, aiming to integrate the predictive capability of data-driven methods with the process understanding provided by hydrological models (Feng et al., 2022; Höge et al., 2022; Jiang et al., 2022).

Recently, hybrid models with different coupling frameworks have been increasingly developed in various studies. These strategies include using DL models as post-processors to correct simulation errors (Feigl et al., 2022; Frame et al., 2021; Quilty et al., 2022), incorporating physical knowledge into neural networks through physics-informed approaches (Cai et al., 2022; Hoedt et al., 2021; Pokharel et al., 2023; Yihan Wang et al., 2025; Wi & Steinschneider, 2024), employing neural networks for process-based model parameterization (Tsai et al., 2021), and embedding neural networks directly into differentiable process-based models (Feng et al., 2022; Li et al., 2024; Zhong et al., 2023). Among these strategies, embedding neural networks (ENNs) into differentiable process-based hydrological models enables a deeper integration between data-driven and physical representations. ENNs can parameterize, enhance, or replace specific hydrological modules while remaining coupled with underlying process dynamics. Such architectures enable bidirectional interactions between data-driven representations and physically based simulations, allowing hydrological states and processes to evolve within a unified modelling framework. (Shen et al., 2023; Tao et al., 2024).

Embedded DL within process-based models (referred to as "hybrid hydrological models" hereafter) is increasingly regarded as one of the most promising approaches in hydrological modeling (Li, Sun, et al., 2023; Wang et al., 2024; Zhong et al., 2024). The development of hybrid hydrological models has gradually evolved from lumped conceptual frameworks (Feng et al., 2022; Li, Sun, et al., 2023) to distributed and semi-distributed modeling systems (Li et al., 2024; Wang et al., 2024; Zhong et al., 2024). Beyond streamflow prediction, hybrid hydrological models are also applied in multiple hydrological processes simulation (Yi Wang et al., 2025). These advances demonstrate that hybrid hydrological models possess considerable potential for advancing hydrological prediction. Despite growing interest in such hybrid systems, most existing studies have primarily focused on improvements in simulation accuracy, while treating uncertainty as an external evaluation metric rather than an internal system property. As a result, the mechanisms by which uncertainties are generated, propagated and accumulated within fully coupled hydrological systems remain insufficiently understood. In particular, it is unclear how ENNs within process equations alters the internal structure of hydrological uncertainty, including whether uncertainty propagates uniformly across hydrological fluxes or becomes localized within specific process–state interactions.

https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

Understanding these mechanisms is essential for assessing the stability and physical consistency of hybrid hydrological models under stochastic parameterization and optimization.

Traditional hydrological uncertainty studies primarily quantified parameter uncertainty, forcing uncertainty and structural uncertainty (McMillan et al., 2018; Renard et al., 2010). Various approaches, such as Monte Carlo simulation, Bayesian inference, and the Generalized Likelihood Uncertainty Estimation (GLUE) framework, have been extensively developed and widely applied (Ragab et al., 2020; Wang et al., 2017; Zhao et al., 2015; Zheng & Keller, 2007). These methods are developed based on physical model structures and relatively explicit uncertainty sources, which facilitates uncertainty attribution and analysis in traditional hydrological models. The applicability of above approaches to hybrid hydrological models remains insufficiently understood. By integrating ENNs, hybrid models introduce increased nonlinearity, high-dimensional parameters and data-driven representations. Uncertainty propagation becomes more complex than in traditional physical models, posing significant challenges for uncertainty attribution and quantification. In addition, existing uncertainty studies largely emphasized streamflow simulations, while uncertainty associated with internal processes and state variables received less attention (Khatami et al., 2019). Such internal uncertainties yet remain poorly characterized. Therefore, a comprehensive uncertainty evaluation framework is needed to investigate uncertainty propagation across both hydrological states and predictive outputs in hybrid hydrological models.

This study aims to address the following questions in hybrid hydrological models:

(1) how do ENNs influence runoff predictive uncertainty?

(2) how does uncertainty propagate through hydrological processes and internal state variables?

To answer the above questions, this study develops a comprehensive ensemble-based framework to investigate uncertainty propagation in hybrid hydrological models. Three hydrological models with different levels of neural-network embedding were developed. Ensemble simulations were conducted to evaluate uncertainty propagation from ENN-replaced hydrological processes to unreplaced physically processes and internal state variables. Multiple uncertainty metrics were further adopted to quantify uncertainty behavior. The remainder of this paper is structured as follows: Sect. 2 describes our hybrid hydrological models, ensemble simulation framework, study area, and the data used. In Sect. 3, we evaluate uncertainty propagation in hybrid hydrological models using various hydrological metrics. The results and their implications are discussed in Sect. 4. Finally, we conclude in Sect. 5.

2 Methods

2.1 Hybrid hydrological models

This study utilizes the hybrid hydrological models NN_x developed in our previous work (Li, Sun, et al., 2023) for uncertainty analysis. The models adopts the conceptual hydrological model EXP-Hydro (hereafter referred to as EXP) as the physical backbone and integrates ENNs within a differentiable programming framework. In the hybrid configurations, ENNs are introduced to replace selected hydrological modules while retaining the original water balance formulation and key physical

https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

constraints of the process-based model. The differentiable framework enables joint optimization of process-based parameters and neural network representations during model training..

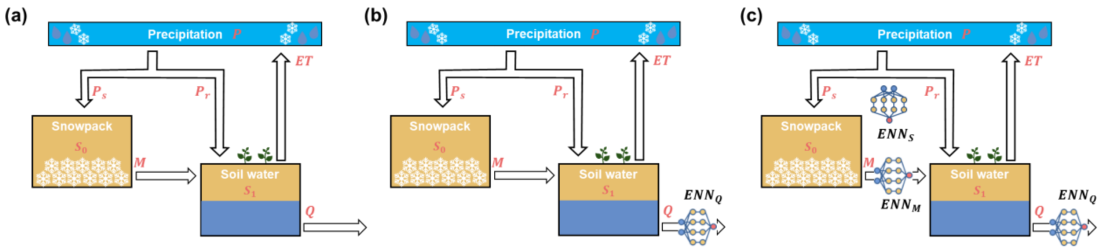

Figure 1. The Schematic diagram of the EXP (a), NN_Q (b) and NN_{QSM} (c) models. Three internal modules can be optionally replaced by embedded neural networks (shown as network insets labeled as ENN_x with x denotes the corresponding hydrological process).

The EXP framework represents hydrological dynamics through two internal storage variables, including snow storage S_0 and basin water storage S_1, and four primary hydrological processes: precipitation partitioning S, snowmelt M, evapotranspiration E, and runoff generation Q (Fig. 1a). Hybrid configurations are constructed by replacing one or multiple process modules of EXP with ENNs. Detailed model structures, formulations and implementations can be found in Appendix A and Li, Sun, et al. (2023). This study employs the original EXP model, the runoff-enhanced hybrid model (NN_Q, Fig. 1b) and the hybrid model replacing runoff and two snow-related processes simultaneously (NN_{QSM}, Fig. 1c), to represent the increasing levels of hybridization complexity.

2.2 Optimization-induced uncertainty

All trainable parameters in the three models were jointly optimized within the differentiable programming framework. The training procedure is sensitive to random parameter initialization and stochastic optimization, so ensemble simulations can produce different hydrological processes using identical forcing datasets. In this study, such variability among ensemble simulations is interpreted as optimization-induced uncertainty. The objective of this study is to investigate the optimization-induced uncertainty propagation throughout different components of the hybrid hydrological systems. Multiple ensemble simulations were firstly generated for each model structure using different random initialization seeds. Different components of ensemble simulations were further evaluated using multiple metrics (Sect. 2.5).

2.3 Study area and Data used

We selected 31 basins from the Catchment Attributes and Meteorology for Large-sample Studies (CAMELS) dataset in the contiguous United States as study sites (Fig. 2; Addor et al., 2017; Newman et al., 2015). Because this study focuses on uncertainty behaviour associated with snow-related processes, we first identified hydrologically suitable regions characterized by relatively cold conditions. Three hydrological zones (HUC 1, 4, and 17) with mean annual air temperatures below 8 °C and more than 20 available basins were selected, resulting in 124 candidate basins. Considering the substantial

https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

computational cost associated with ensemble simulations of hybrid hydrological models, 31 representative basins were further selected from these candidate basins. The 124 basins were first sorted according to their original ordering in the CAMELS dataset. Every fourth basin (i.e., 1st, 5th, 9th, …) was selected to ensure spatial representativeness across the study regions.

The model simulations were driven by three meteorological variables: precipitation, air temperature and day length. Observed runoff was used as the training target for all model configurations. Meteorological forcing data are obtained from the Daymet subset of the CAMELS dataset, which has demonstrated superior performance compared with other available forcing products (Feng et al., 2022; Kratzert et al., 2021; Newman et al., 2015).

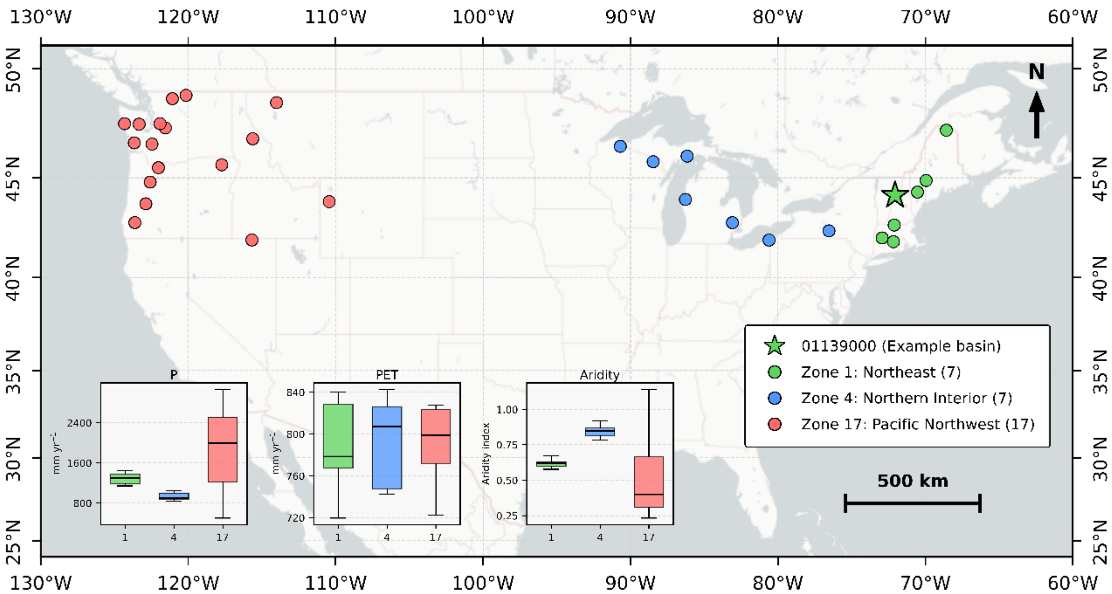

Figure 2. Spatial distributions of the 31 basins used in this study across three hydrological zones. Insets show the distributions of mean annual precipitation (P), potential evapotranspiration (PET), and aridity index (AI) for the selected basins in different hydroclimatic regions.

2.4 Experiment design

This study applied the EXP, NN_Q and NN_{QSM} models in the 31 selected basins. For each basin, the three model configurations were separately optimized and evaluated using the Nash–Sutcliffe efficiency (NSE; Nash & Sutcliffe, 1970) as the loss function. Model training was conducted from 1 October 1981 to 30 September 2000, while the evaluation period extended from 1 October 2000 to 30 September 2010. A one-year warm-up period preceding the training interval was included to initialize model states. For each model structure, 50 ensemble simulations were conducted using different random parameter initialization and stochastic optimization. The ensemble size was selected to balance computational efficiency and ensemble stability. Preliminary experiments indicated that increasing the number of ensemble members beyond this threshold resulted in only marginal changes in performance statistics (Appendix B). All ensemble members

https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

shared identical meteorological forcing data, model structures, and training configurations to ensure consistent comparison among different hybridization strategies. This study employed these ensemble simulations to evaluate how ENNs affects model uncertainty across three types of hydrological components:

(1) **direct uncertainty effects on ENN-replaced processes**: runoff generation process in NN_Q model relative to EXP model; precipitation partitioning and snowmelt process in NN_{QSM} model relative to NN_Q and EXP model. Besides, runoff generation process in NN_{QSM} model relative to EXP model was utilized to investigate whether additional ENN replacement of hydrological processes further enhances uncertainty in the ENN-replaced runoff generation process.

(2) **indirect uncertainty propagation to unreplaced processes**: precipitation partitioning and snowmelt process in NN_Q model relative to EXP model; evapotranspiration process in NN_Q and NN_{QSM} models relative to EXP model.

(3) **uncertainty accumulation within internal state variables**: snow storage (S_0) and basin water storage (S_1) in NN_Q and NN_{QSM} models relative to EXP model.

This design allows us to distinguish whether uncertainty amplification primarily occurs in hydrological processes directly replaced by ENNs, propagates indirectly to unreplaced physically based processes, or accumulates within internal state variables.

2.5 Evaluation metrics

The following metrics are used to evaluate the model performance and uncertainty in the later analysis.

2.5.1 Model performance evaluation metrics

The predictive capability of different models was assessed using NSE, modified NSE (mNSE; Legates & McCabe Jr, 1999), and the absolute value of peak flow bias (PFAB; Yilmaz et al., 2008). NSE and mNSE were calculated using the following generalized formulation:

$$\mathrm{NSE}_{\alpha} = 1 - \frac{\sum_{i=1}^{T} |Q_{obs,i} - Q_{sim,i}|^{\alpha}}{\sum_{i=1}^{T} |Q_{obs,i} - \overline{Q_{obs}}|^{\alpha}} \tag{1}$$

where $Q_{sim,i}$ and $Q_{obs,i}$ denote simulated and observed runoff at time step i , T represents the total number of simulation steps, and $\overline{Q_{obs}}$ is the averaged observed values. NSE and mNSE are defined by Equation (1) with α being 2 and 1, respectively.

Peak flow bias was quantified using PFAB:

$$\mathrm{PFAB} = 100 \times \left| \frac{\sum_{i=1}^{L} (Q_{sim,i} - Q_{obs,i})}{\sum_{i=1}^{L} Q_{obs,i}} \right| \tag{2}$$

where $Q_{sim,i}$ and $Q_{obs,i}$ are the simulated and observed runoff sorted in descending order, respectively. L denotes the number of samples included in the top 1% of observed flows.

https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

2.5.2 Model uncertainty evaluation metrics

Model uncertainty was evaluated using four complementary metrics, including the interquartile range ($\mathrm{IQR_x}$), mean pairwise correlation ($\mathrm{MPC_x}$), P-factor and R-factor. These four metrics can quantify uncertainty magnitude, ensemble consistency, uncertainty coverage and uncertainty width, respectively.

The interquartile range ($\mathrm{IQR_x}$) was calculated as:

$$\mathrm{IQR_x} = IQ_{75}(\mathrm{x}) - IQ_{25} \quad (3)$$

where $IQ_{75}(\mathrm{x})$ and $IQ_{25}(\mathrm{x})$ denote the 75th and 25th percentiles of ensemble simulations for variable or metric x, respectively. Larger $\mathrm{IQR_x}$ values indicate stronger uncertainty spread among ensemble members. In this study, x represents different hydrological variables and evaluation metrics.

The mean pairwise correlation $\mathrm{MPC_x}$ was defined as:

$$\mathrm{MPC_x} = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} Corr(x_i, x_j) \quad (4)$$

where N is the number of ensemble members, x_i and x_j represent the i-th and j-th ensemble simulations of variable x, respectively. $Corr$ denotes the Pearson correlation coefficient. Higher $\mathrm{MPC_x}$ values indicate stronger consistency and stability among ensemble simulations.

The P$-factor$ represents the percentage of observed runoff values bracketed by the 95% prediction uncertainty (95PPU) band (Abbaspour et al., 2007):

$$\mathrm{P}-factor = \frac{1}{N} \sum_{i=1}^{N} I\left(Q_{obs,i} \in \left[Q_{2.5,i}, Q_{97.5,i}\right]\right) \quad (5)$$

where $Q_{2.5,i}$ and $Q_{97.5,i}$ denote the 2.5th and 97.5th percentiles of ensemble simulations at time step i, respectively, and I is the indicator function.

The R$-factor$ represents the uncertainty width and was calculated as:

$$\mathrm{R}-factor = \frac{1}{N} \sum_{i=1}^{N} \frac{Q_{97.5,i} - Q_{2.5,i}}{\sigma_{obs}} \quad (6)$$

where σ_{obs} is the standard deviation of observed runoff. Larger R$-factor$ values indicate wider uncertainty bands. The P$-factor$ and R$-factor$ can provide complementary evaluation of uncertainty reliability and uncertainty spread in ensemble simulations.

Among the above metrics, three model performance metrics as well as the P$-factor$ and R$-factor$ require observed data and were therefore only applied to runoff-related evaluation. $\mathrm{IQR_x}$ and $\mathrm{MPC_x}$ do not require observational references and were used to evaluate uncertainty in snow-related processes, evapotranspiration processes and internal state variables. For the calculation of $\mathrm{IQR_x}$ and $\mathrm{MPC_x}$, this study utilizes S, M, ET, S0 and S1 to denote precipitation partitioning, snowmelt, evapotranspiration, snow storage and basin water storage variables, respectively. For example, $\mathrm{IQR_{S0}}$ represents the interquartile range of snow storage simulations. Furthermore, to eliminate the influence of differences in the magnitudes of snowfall, snowmelt, evapotranspiration, and internal state variables among basins, all variables were first normalized prior to

https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

IQR_x and MPC_x calculations. Specifically, snowfall, snowmelt, and evapotranspiration processes were normalized by observed precipitation, while internal state variables were normalized by their maximum values within each basin.

3 Results

3.1 Performance improvements of hybrid hydrological models

Fig. 3 presents the median results of 50 ensemble runoff simulations for the EXP, NN_Q and NN_{QSM} models across the 31 study basins. Consistent with our previous study in (Li, Sun, et al., 2023), both NN_Q and NN_{QSM} exhibited higher NSE and mNSE values than the EXP model and NN_{QSM} model achieved the highest median NSE and mNSE values (Table 1). It is indicated that embedding ENNs into hydrological processes enhanced the capability of capturing overall runoff dynamics. Replacing snow-related processes with ENNs can further improved predictive performance. Besides, PFAB values slightly increased in two hybrid models relative to the EXP model, indicating that hybridization simultaneously introduce larger biases in peak-flow representation. These results suggest that enhanced performance in hybrid hydrological models does not necessarily imply improved hydrological consistency or reduced uncertainty, motivating further investigation into uncertainty propagation within hybrid hydrological models.

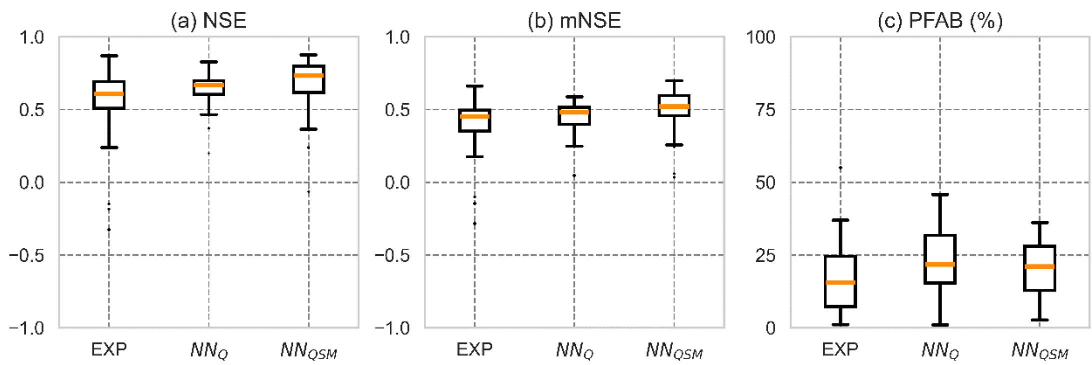

Figure 3. Comparison of streamflow simulation performance among the EXP, NN_Q, and NN_{QSM} models across the 31 study basins. For each basin and model structure, NSE, mNSE, and PFAB values represent the median results of 50 ensemble simulations.

Table 1. Median streamflow simulation performance and uncertainty evaluation metrics of the EXP, NN_Q and NN_{QSM} models across the 31 study basins based on 50 ensemble simulations.

	EXP	NN_Q	NN_{QSM}
NSE	0.608	0.668	0.733
mNSE	0.451	0.480	0.520
PFAB	15.5%	21.7%	24.6%
IQR_{NSE}	0.039	0.067	0.073

https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

$\mathbf{IQR_{mNSE}}$	0.024	0.100	0.105
$\mathbf{IQR_{PFAB}}$	5.4%	14.6%	12.8%
$\mathbf{P\text{–}}\boldsymbol{factor}$	0.209	0.777	0.837
$\mathbf{R\text{–}}\boldsymbol{factor}$	0.264	0.891	0.885
$\mathbf{IQR_{S}}$	0.020	0.023	0.134
$\mathbf{IQR_{M}}$	0.016	0.020	0.143
$\mathbf{IQR_{ET}}$	0.018	0.022	0.013
$\mathbf{IQR_{S0}}$	0.017	0.015	0.057
$\mathbf{IQR_{S1}}$	0.037	0.129	0.121
$\mathbf{MPC_{S}}$	0.955	0.943	0.915
$\mathbf{MPC_{M}}$	0.846	0.862	0.902
$\mathbf{MPC_{ET}}$	0.999	0.994	0.997
$\mathbf{MPC_{S0}}$	0.982	0.979	0.940
$\mathbf{MPC_{S1}}$	0.990	0.974	0.960

3.2 Uncertainty effects on ENN-replaced processes

This section investigates uncertainty effects on ENN-replaced hydrological processes based on simulated runoff, snowfall and snowmelt results from EXP, NN_Q, and NN_{QSM} models. First we evaluate the uncertainty of the ENN-replaced runoff generation processes through compraison between EXP and NN_Q models. Fig. 4 and Table 1 showed that all three IQR-related metrics $\mathrm{IQR_{NSE}}$, $\mathrm{IQR_{mNSE}}$ and $\mathrm{IQR_{PFAB}}$ increased from 0.039, 0.024 and 5.4% in EXP to 0.067, 0.100 and 14.6% in NN_Q model. The amplified uncertainty exhibited distinct characteristics under different runoff conditions. Compared with overall runoff uncertainty represented by $\mathrm{IQR_{NSE}}$, uncertainty amplification was stronger under low-flow and peak-flow conditions, as reflected by larger increases in $\mathrm{IQR_{mNSE}}$ and $\mathrm{IQR_{PFAB}}$. In particular, the more than fourfold increase in $\mathrm{IQR_{mNSE}}$ suggests that low-flow simulations were especially sensitive to ENN embedding within runoff generation processes. Fig. 4 further demonstrated that the NN_Q model exhibited wider IQR distributions than the EXP model across all three metrics, indicating stronger ensemble variability after the introduction of ENN.

To further investigate uncertainty coverage and uncertainty width under different runoff conditions, the $\mathrm{P}\text{–}factor$/$\mathrm{R}\text{–}factor$ analyses (Fig. 5) and an ensemble hydrograph comparison for an example basin (Fig. 6) were conducted. The NSE, $\mathrm{P}\text{–}factor$, and $\mathrm{R}\text{–}factor$ values in this example basin were all close to the median results across the 31 study basins (Fig. 2). It is believed that this basin can provide a representative characterization of ensemble simulation uncertainty.

https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

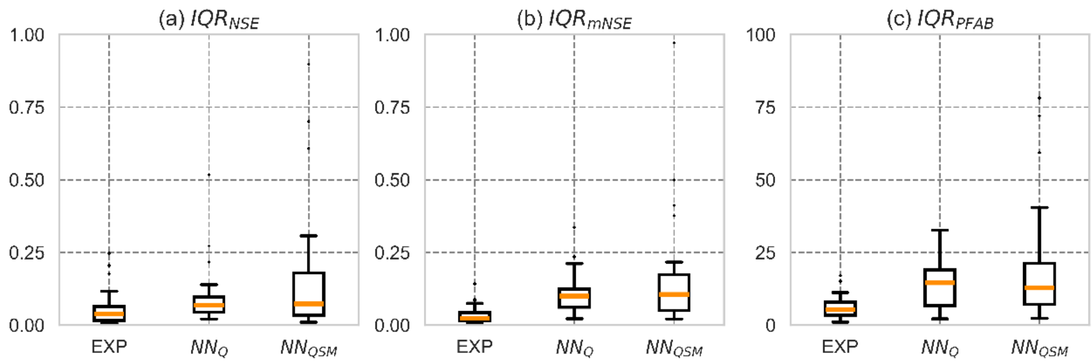

Figure 4. Comparison of runoff uncertainty metrics–$\mathbf{IQR_{NSE}}$, $\mathbf{IQR_{mNSE}}$ and $\mathbf{IQR_{PFAB}}$–among EXP, $\boldsymbol{NN_Q}$ and $\boldsymbol{NN_{QSM}}$ models across the 31 study basins based on 50 ensemble simulations.

As shown in Fig. 5 and Table 1, the NN_Q model exhibited higher P$-factor$ values than the EXP model, with median values increasing from 0.209 to 0.777. Most basins in the NN_Q model achieved P$-factor$ values greater than 0.7, whereas almost no basins in the EXP model exceeded this threshold. The results also revealed that distinct uncertainty coverage characteristics under different runoff conditions. Under both high-flow and low-flow conditions, the NN_Q model consistently exhibited higher P$-factor$ values than the EXP model. However, compared with overall flow conditions, P$-factor$ values generally decreased under high-flow conditions but increased under low-flow conditions for both models. Most basins in the NN_Q model achieved P$-factor$ values approaching 1 under low-flow conditions. These results indicate that the NN_Q model improved the coverage of observed runoff within ensemble prediction intervals, with uncertainty coverage under low-flow conditions being better than that under high-flow conditions.

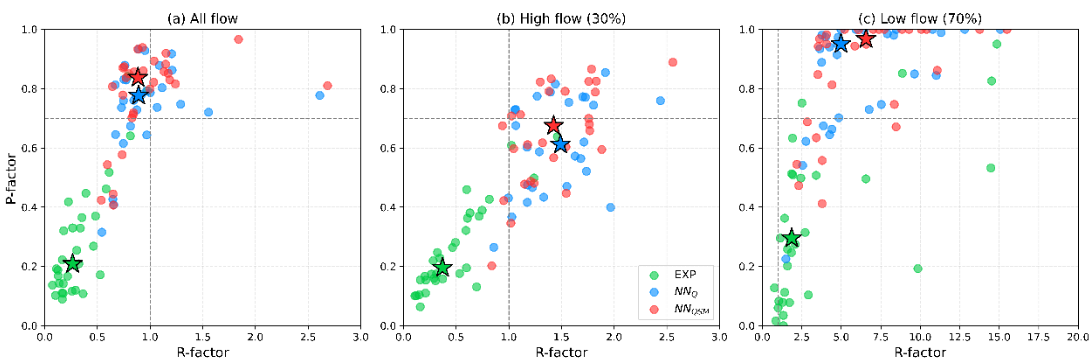

Figure 5. Comparison of P$-factor$ and R$-factor$ relationships among EXP, $\boldsymbol{NN_Q}$ and $\boldsymbol{NN_{QSM}}$ models under overall flows, high flows (top 30%), and low flows (bottom 70%) conditions.

In addition, the NN_Q model consistently exhibited larger R$-factor$ values than the EXP model under overall, high-flow and low-flow conditions. Under overall runoff conditions, the median R$-factor$ values of the EXP and NN_Q models were 0.264 and 0.891, respectively, both remaining below 1. Compared with overall runoff conditions, R$-factor$ values increased

https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

under high-flow and low-flow conditions. The increase is more pronounced under low-flow conditions, with a considerably larger proportion of basins exhibiting R$-factor$ values greater than 1. Overall, the NN_Q model generated wider uncertainty bands than the EXP model, particularly under low-flow conditions where the largest R-factor values were observed.

Combined with the above analysis, the increase in P$-factor$ closely tracked the increase in R$-factor$, indicating a clear trade-off between uncertainty coverage and uncertainty band width. The NN_Q model achieved improved uncertainty coverage at the expense of considerably wider uncertainty bands rather than improved calibration. This tradeoff was especially pronounced under low-flow conditions, where the NN_Q model simultaneously exhibited the highest uncertainty coverage and the widest uncertainty bands. The ensemble hydrographs of the example basin further support these findings (Fig. 6). The NN_Q model displayed larger ensemble spread and wider 95PPU ranges than the EXP model, particularly during low-flow periods.

Following the runoff uncertainty analysis, we further investigated the direct effects of ENN-replaced snow-related processes on uncertainty through comparisons between the NN_Q and NN_{QSM} models (Fig. 7 and Table 1). In the NN_Q model, precipitation partitioning and snowmelt processes remained physically based, whereas these processes were directly replaced by ENNs in the NN_{QSM} model. Uncertainty differences between these two models can reflect the direct uncertainty effects introduced by ENNs embedding within snow-related processes.

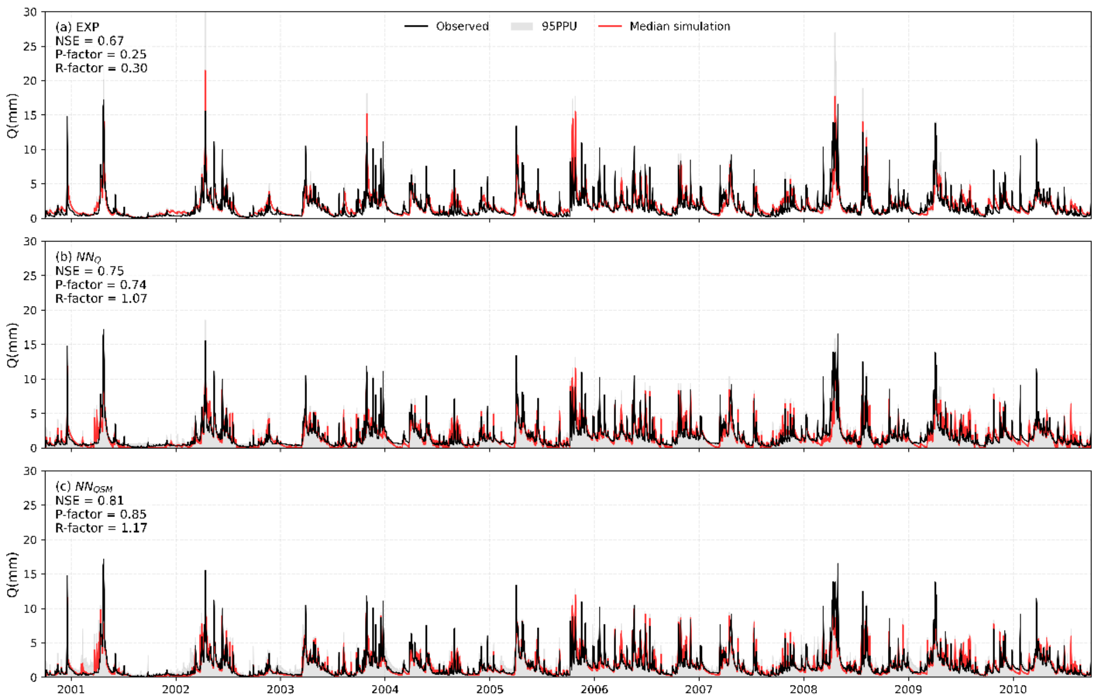

Figure 6. Ensemble median streamflow simulations and 95% prediction uncertainty (95PPU) bands of the EXP, NN_Q and NN_{QSM} models for a representative study basin (Fig. 2) during the evaluation period.

https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

As shown in Fig. 7 and Table 1, the NN_{QSM} model exhibited larger $\mathrm{IQR_S}$ and $\mathrm{IQR_M}$ values than the NN_Q model, with median values increasing from 0.023 and 0.020 in the NN_Q model to 0.134 and 0.143 in the NN_{QSM} model, respectively. Relative to the comparatively concentrated uncertainty distributions in the NN_Q model, both precipitation partitioning and snowmelt processes in the NN_{QSM} model exhibited enlarged IQR distributions and stronger inter-basin variability. These results indicate pronounced uncertainty amplification after direct ENNs embedding within snow-related hydrological processes.

The amplified uncertainty characteristics were further reflected by the MPC-based ensemble consistency analysis. For precipitation partitioning processes, the NN_{QSM} model exhibited lower $\mathrm{MPC_S}$ values than the NN_Q model with median values decreasing from 0.943 to 0.915. For snowmelt processes, although the median $\mathrm{MPC_M}$ value in the NN_{QSM} model was slightly higher than that in the NN_Q model (0.902 versus 0.862), the NN_{QSM} model showed lower mean values together with a much wider distribution toward lower MPC values. These results demonstrated the weakened ensemble consistency of two snow-related processes in the NN_{QSM} model. Therefore, utilizing ENNs to replace hydrological processess can amplify the uncertainty within these processes.

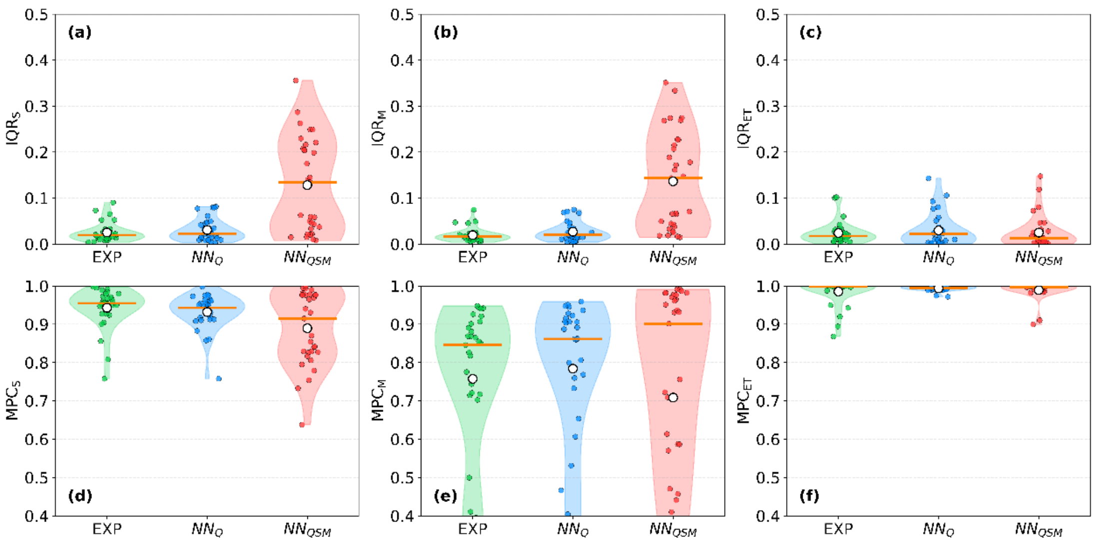

Figure 7. Comparison of IQR and MPC for normalized snowfall, snowmelt and ET variables among the EXP, NN_Q and NN_{QSM} models across the 31 study basins. Orange horizontal lines denote median values, while white circles represent mean values.

Furthermore, runoff uncertainty in the NN_{QSM} model relative to the EXP model was compared with that in the NN_Q model relative to the EXP model. This analysis aimed to investigate whether utilizing ENNs to replace other hydrological processes can further enhance uncertainty in the ENN-replaced runoff generation process. Results showed that the runoff uncertainty metrics of the NN_{QSM} and NN_Q models remained generally similar, while most uncertainty metrics in the NN_{QSM} model were

https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

slightly higher and showed wider IQR distributions. These findings indicate that increasing hybridization complexity through simultaneous ENN replacement of multiple hydrological processes slightly enhanced runoff uncertainty propagation.

3.3 Uncertainty propagation to unreplaced processes

This section investigates whether uncertainty introduced by ENNs propagates to physically based hydrological processes that were not directly replaced by ENNs. Specifically, comparisons between EXP and NN_Q models were used to evaluate uncertainty propagation from the ENN-replaced runoff process to unreplaced snow-related processes. EXP, NN_Q and NN_{QSM} models were further compared to assess uncertainty propagation to the unreplaced ET process under different levels of hybridization complexity.

Results showed that uncertainty propagation is highly constrained by physical process structure and does not exhibit system-wide amplification compared with the direct uncertainty amplification observed in ENN-replaced processes (Fig. 7). For snow-related processes, the uncertainty metrics of the NN_Q and EXP models remained highly similar. Relative to the EXP model, the NN_Q model exhibited slightly wider uncertainty spreads, with $\mathrm{IQR_S}$ and $\mathrm{IQR_M}$ increasing marginally from 0.020 and 0.016 to 0.023 and 0.020, respectively. These increases were negligible compared with the pronounced uncertainty amplification caused by direct ENN replacement of snow-related processes in the NN_{QSM} model. Besides, the NN_Q model exhibited slightly weaker consistency in precipitation partitioning processes ($\mathrm{MPC_S}$: 0.943 versus 0.955) but slightly stronger consistency in snowmelt processes ($\mathrm{MPC_M}$: 0.862 versus 0.846) relative to the EXP model. Uncertainty propagation to the ET process was even weaker. The uncertainty metrics of the EXP, NN_Q and NN_{QSM} models remained highly consistent and similar uncertainty patterns with only minimal differences among models. $\mathrm{IQR_{ET}}$ values remained consistently low across all three models and $\mathrm{MPC_{ET}}$ values were consistently close to 1. These results suggest that uncertainty amplification induced by ENNs does not propagate to unreplaced physically based processes.

3.4 Uncertainty accumulation within internal state variables

This section compare uncertainty characteristics of internal state variables among three models to evaluate whether uncertainty amplification induced by ENNs accumulates within memory states (Fig. 8). For snow storage (S_0), both EXP and NN_Q models showed similarly low uncertainty levels, with similarly low $\mathrm{IQR_{S0}}$ (0.017 and 0.015, respectively) and high $\mathrm{MPC_{S0}}$ (0.982 and 0.979, respectively) values. NN_{QSM} models achieved higher $\mathrm{IQR_{S0}}$ (0.057) and lower $\mathrm{MPC_{S0}}$ (0.940) values, suggesting the uncertainty of S_0 is larger. Basin water storage (S_1) showed a different uncertainty pattern that uncertainty in the EXP model is small while that in the NN_Q and NN_{QSM} models is large. $\mathrm{IQR_{S1}}$ increased from 0.037 in the EXP model to 0.129 and 0.121 in the NN_Q and NN_{QSM} models, respectively, while $\mathrm{MPC_{S1}}$ decreased progressively from 0.990 to 0.974 and 0.960. These results demonstrated that uncertainty accumulation within internal state variables was closely associated with their corresponding ENN-replaced hydrological processes. In theNN_{QSM} model, ENNs replace snow-

https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

related processes, resulting in larger uncertainty in snow storage. In the EXP and NN_Q models, ENNs replace the runoff process, leading to larger uncertainty in basin water storage.

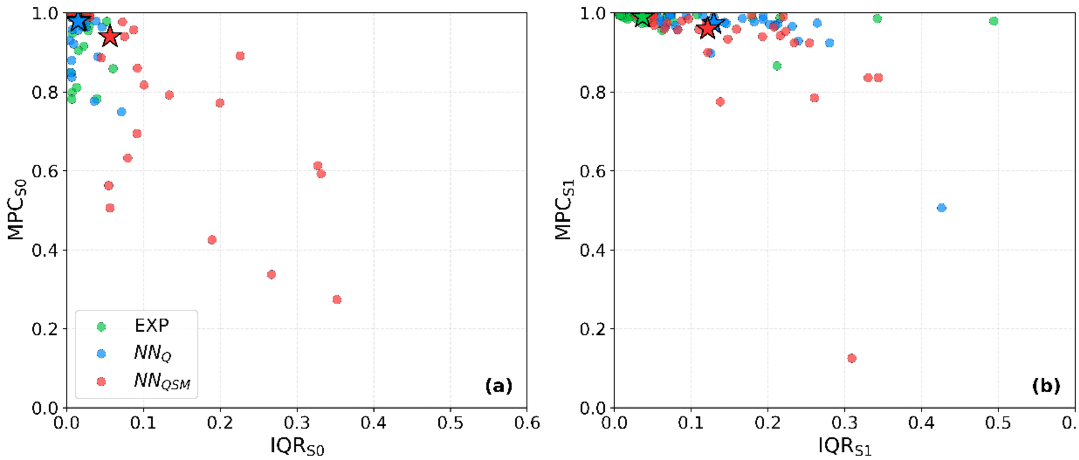

Figure 8. Relationships between normalized IQR and MPC for internal state variables (S_0 and S_1) among the EXP, NN_Q and NN_{QSM} models across the 31 study basins. Circles represent individual basins, while stars denote the mean values of all basins.

In conjunction with the results presented in Sect. 3.1.3, the above findings indicate that uncertainty arising from ENNs is preferentially accumulated in internal state variables possessing memory effects, as opposed to being propagated to other unreplaced hydrological processes. This observation further suggests that internal state variables function as significant reservoirs for the storage and accumulation of uncertainty within hybrid hydrological systems, thereby constituting a critical pathway via which the embedding of ENNs exerts influence on model stability.

4 Discussions

4.1 Rationale behind uncertainty propagation in hybrid hydrological models

This study demonstrated that uncertainty amplified by ENNs was primarily concentrated within directly replaced hydrological processes and their corresponding internal state variables. Uncertainty propagation to unreplaced physically based hydrological processes remained relatively limited. Uncertainty propagation in hybrid hydrological systems is not uniformly distributed across different hydrological components, but instead remains largely localized within directly associated process–state systems.

The amplified uncertainty likely originates from the differences between physically based formulations and ENN-based process representations in constraining hydrological behaviors. Physically based hydrological processes are constrained by predefined equations and physically interpretable relationships. For example, traditional snowmelt formulations commonly impose monotonic temperature–melt relationships above a threshold temperature. Under these formulations, parameter

https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

perturbations generally produce limited divergence in hydrological responses among ensemble simulations. In contrast, ENN-replaced processes generally contain more trainable parameters and weaker explicit process constraints. They thus enlarge the feasible hydrological response space within hybrid hydrological systems. Different parameter initializations and stochastic optimization trajectories may converge toward substantially different hydrological response behaviors. This mechanism is consistent with the pronounced uncertainty amplification observed in ENN-replaced runoff generation and snow-related processes in this study. It also helps explain why the increased uncertainty coverage of hybrid models was accompanied by wider uncertainty bands. The larger feasible hydrological response space introduced by ENNs increased ensemble spread among simulations. As a result, observations were more likely to be enclosed by prediction intervals, but the sharpness of uncertainty estimates was reduced.

The uncertainty amplified by ENNs was more likely to accumulate within internal state variables than to propagate directly to unreplaced hydrological fluxes. The physical modules in hybrid models still preserve predefined hydrological relationships and process constraints, thereby dampening excessive uncertainty propagation originating from ENN-replaced components. This buffering and stabilizing effect was particularly evident in the ET process, where uncertainty metrics remained highly consistent across all three model structures. In contrast, internal state variables possess memory and integration effects that allow uncertainty to persist and gradually accumulate over time. As a result, internal hydrological storages acted as important uncertainty reservoirs within hybrid hydrological systems.

4. 2 Implications and limitations for hybrid hydrological models

This study has several important implications for hybrid hydrological modeling. First, although ENN substantially improved runoff simulation accuracy, it simultaneously amplified uncertainty within directly replaced hydrological processes and internal state variables. This suggests that conventional performance metrics alone may be insufficient for evaluating hybrid hydrological models. Practical applications of hybrid hydrological models should jointly consider predictive accuracy, uncertainty characteristics, and internal system stability. Using the ensemble median as the final simulation output may help reduce the influence of stochastic optimization–induced uncertainty and improve simulation robustness.

Second, the results highlight the importance of explicitly evaluating uncertainty propagation and accumulation within internal hydrological states. Existing hybrid hydrological studies primarily focus on runoff performance, while uncertainty behaviors within internal process representations and state variables are rarely investigated. Uncertainty amplification caused by ENNs may remain hidden if model evaluations rely solely on runoff performance. Incorporating uncertainty diagnostics for both hydrological processes and internal states can provide a more comprehensive assessment of model reliability and physical consistency.

Third, hybrid hydrological modeling strategies can benefit from retaining physically constrained formulations for hydrological components that strongly regulate overall system stability. Future hybrid hydrological models may benefit from introducing stronger physical constraints into ENN-based process representations, such as conservation relationships, monotonicity constraints, and physics-guided regularization strategies. In addition, incorporating multisource observations

https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

and datasetsfor joint training and physical constraint integration may further improve model physical consistency and reduce excessive uncertainty amplification while preserving predictive flexibility.

Several limitations in this study should also be acknowledged. First, this study mainly focused on uncertainty induced by parameter perturbations and stochastic optimization, while additional uncertainty sources, such as forcing uncertainty, structural uncertainty, and observational uncertainty, were not explicitly considered. Second, only one ENN-based hybridization strategy was investigated in this study. Different neural-network architectures, hybridization structures, and optimization schemes may produce different uncertainty propagation characteristics. Future studies should further investigate how multiple uncertainty sources interact and propagate within hybrid hydrological systems under different hybridization strategies.

5 Conclusions

This study systematically investigated uncertainty propagation in hybrid hydrological models induced by ENNs. Three hybrid model structures with different levels of ENN complexity were used and evaluated in 31 cold-region basins. The main conclusions are summarized as follows.

(1) ENNs improved simulation performance but simultaneously amplified uncertainty within directly replaced hydrological processes. Compared with physically based formulations, ENN-replaced runoff generation and snow-related processes exhibited larger uncertainty spreads, reduced ensemble consistency, and wider ensemble prediction intervals. Increasing hybridization complexity through simultaneous ENN replacement of multiple hydrological processes slightly enhanced uncertainty amplification.

(2) Uncertainty induced by ENNs was more likely to accumulate within internal hydrological state variables than to propagate to unreplaced physically based hydrological processes. Internal states variables with memory effects acted as important uncertainty reservoirs within hybrid hydrological systems, whereas unreplaced physically based processes remained comparatively stable across different hybridization levels.

(3) The amplified uncertainty induced by ENN embedding likely originated from the larger parameter space, stronger nonlinear flexibility and weaker explicit process constraints of ENN-based process representations compared with physically based formulations.

Overall, this study demonstrates that uncertainty propagation in hybrid hydrological systems is highly structured rather than uniformly distributed across hydrological components. The findings highlight the importance of jointly evaluating predictive performance, uncertainty characteristics, and internal system stability in hybrid hydrological modeling. Future developments of hybrid hydrological models may benefit from incorporating stronger physical constraints and uncertainty-aware modeling strategies to improve predictive robustness while preserving physical consistency.

https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

Appendix A: Equations of ENN-replaced hydrological processes in the NN_Q and NN_{QSM} models

The EXP-Hydro model was used as the physically based benchmark framework in this study. To highlight the hybridization configurations considered in the NN_Q and NN_{QSM} models, this appendix summarizes the formulations of hydrological components replaced by ENNs, including precipitation partitioning, snowmelt, and runoff generation .

(1) Precipitation partition

$$P_s = P \times NN_S\,(P,\,T) \tag{A1}$$

where $P_s T$ and P denote snowfall, precipitation, and mean air temperature, respectively. $NN_S\,(P,\,T)$ represents the neural network used for precipitation partitioning with a Sigmoid activation function.

(2) Snowmelt

$$M = S_0 \times NN_M\,(T) \tag{A2}$$

where M is the snowmelt and $NN_M\,(T)$ denotes the neural network used for snowmelt estimation with a Sigmoid activation function.

(3) Runoff

$$Q = NN_Q\,(M + P_r, S_1) \tag{A3}$$

where Q is the runoff, S_1 is the basin water storage, P_r is the rainfall, and $NN_Q\,(M + P_r, S_1)$ is the neural network with a ReLU activation function.

Appendix B: Sensitivity of uncertainty statistics to ensemble size

To evaluate the sensitivity of model performance and uncertainty statistics to ensemble size, a preliminary sensitivity analysis was conducted using an example basin (Fig. 2) with 150 ensemble simulations. Performance metrics (NSE, mNSE, and PFAB) and uncertainty metrics ($\text{P}-factor$ and $\text{R}-factor$) were recalculated using different numbers of ensemble members. Results showed that both performance and uncertainty statistics became generally stable when the ensemble size reached 50, and further increasing the ensemble size produced only marginal changes (Fig. B1). Therefore, 50 ensemble members were selected in this study as a reasonable compromise between computational efficiency and ensemble stability.

https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

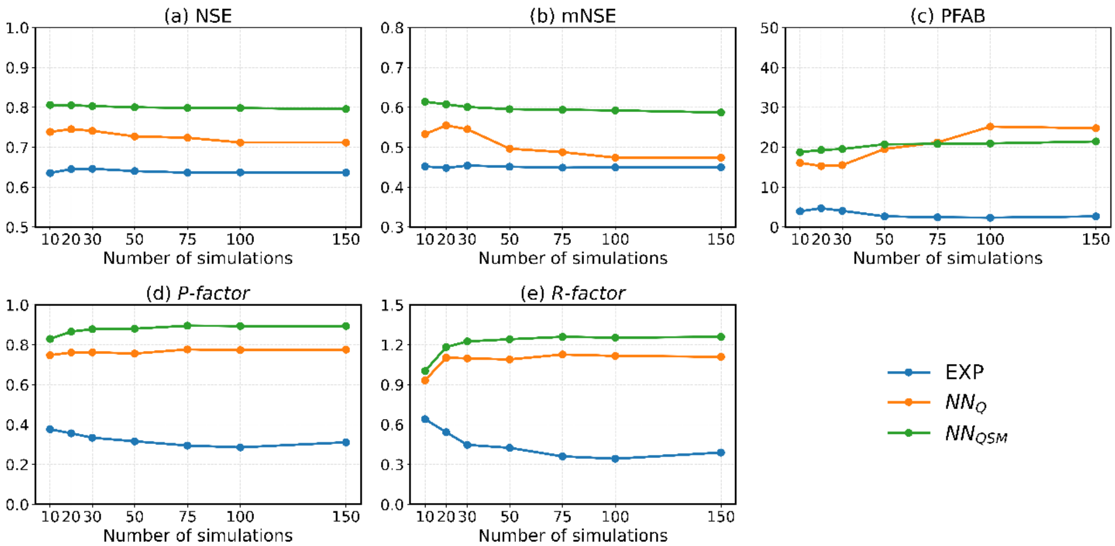

Figure B1. Sensitivity of performance metrics (NSE, mNSE, and PFAB) and uncertainty metrics (P–$factor$ and R–$factor$) to ensemble size for an example basin (Fig. 2).

Code and data availability

CAMELS dataset used in this study is openly available(Addor et al., 2017; Newman et al., 2015). The model source code will be released after the completion of the first round of peer review.

Author contributions

BL conceived the idea, collected the data, designed the experiments, and carried them out. All authors conducted the analysis. GN contributed to the supervision. BL drafted the manuscript, and all authors reviewed and edited the manuscript

Competing interests

At least one of the (co-)authors is a member of the editorial board of Hydrology and Earth System Sciences.

Financial support

This work was funded by the State Key Laboratory of Hydroscience and Engineering (sklhse-TD-2024-C01) and Science and Technology Department of Gansu Province (22ZD6WA043).

References

Abbaspour, K. C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J., and Srinivasan, R.: Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT, Journal of Hydrology, 333(2-4), 413–430, 2007.

Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrology and Earth System Sciences, 21(10), 5293–5313, https://doi.org/10.5194/hess-21-5293-2017, 2017.

Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J. W., McDonnell, J. J., Savenije, H. H. G., Sivapalan, M., Stumpp, C., Toth, E., Volpi, E., Carr, G., Lupton, C., Salinas, J., Széles, B., Viglione, A., Aksoy, H., Allen, S. T., Amin, A., Andréassian, V., Arheimer, B., Aryal, S. K., Baker, V., Bardsley, E., Barendrecht, M. H., Bartosova, A., Batelaan, O., Berghuijs, W. R., Beven, K., Blume, T., Bogaard, T., Borges de Amorim, P., Böttcher, M. E., Boulet, G., Breinl, K., Brilly, M., Brocca, L., Buytaert, W., Castellarin, A., Castelletti, A., Chen, X., Chen, Y., Chen, Y., Chifflard, P., Claps, P., Clark, M. P., Collins, A. L., Croke, B., Dathe, A., David, P. C., de Barros, F. P. J., de Rooij, G., Di Baldassarre, G., Driscoll, J. M., Duethmann, D., Dwivedi, R., Eris, E., Farmer, W. H., Feiccabrino, J., Ferguson, G., Ferrari, E., Ferraris, S., Fersch, B., Finger, D., Foglia, L., Fowler, K., Gartsman, B., Gascoin, S., Gaume, E., Gelfan, A., Geris, J., Gharari, S., Gleeson, T., Glendell, M., Gonzalez Bevacqua, A., González-Dugo, M. P., Grimaldi, S., Gupta, A. B., Guse, B., Han, D., Hannah, D., Harpold, A., Haun, S., Heal, K., Helfricht, K., Herrnegger, M., Hipsey, M., Hlaváčiková, H., Hohmann, C., Holko, L., Hopkinson, C., Hrachowitz, M., Illangasekare, T. H., Inam, A., Innocente, C., Istanbulluoglu, E., Jarihani, B., Kalantari, Z., Kalvans, A., Khanal, S., Khatami, S., Kiesel, J., Kirkby, M., Knoben, W., Kochanek, K., Kohnová, S., Kolechkina, A., Krause, S., Kreamer, D., Kreibich, H., Kunstmann, H., Lange, H., Liberato, M. L. R., Lindquist, E., Link, T., Liu, J., Loucks, D. P., Luce, C., Mahé, G., Makarieva, O., Malard, J., Mashtayeva, S., Maskey, S., Mas-Pla, J., Mavrova-Guirguinova, M., Mazzoleni, M., Mernild, S., Misstear, B. D., Montanari, A., Müller-Thomy, H., Nabizadeh, A., Nardi, F., Neale, C., Nesterova, N., Nurtaev, B., Odongo, V. O., Panda, S., Pande, S., Pang, Z., Papacharalampous, G., Perrin, C., Pfister, L., Pimentel, R., Polo, M. J., Post, D., Prieto Sierra, C., Ramos, M.-H., Renner, M., Reynolds, J. E., Ridolfi, E., Rigon, R., Riva, M., Robertson, D. E., Rosso, R., Roy, T., Sá, J. H. M., Salvadori, G., Sandells, M., Schaefli, B., Schumann, A., Scolobig, A., Seibert, J., Servat, E., Shafiei, M., Sharma, A., Sidibe, M., Sidle, R. C., Skaugen, T., Smith, H., Spiessl, S. M., Stein, L., Steinsland, I., Strasser, U., Su, B., Szolgay, J., Tarboton, D., Tauro, F., Thirel, G., Tian, F., Tong, R., Tussupova, K., Tyralis, H., Uijlenhoet, R., van Beek, R., van der Ent, R. J., van der Ploeg, M., Van Loon, A. F., van Meerveld, I., van Nooijen, R., van Oel, P. R., Vidal, J.-P., von Freyberg, J., Vorogushyn, S., Wachniew, P., Wade, A. J., Ward, P., Westerberg, I. K., White, C., Wood, E. F., Woods, R., Xu, Z., Yilmaz, K. K., and Zhang, Y.: Twenty-three unsolved problems in hydrology (UPH) – a community perspective, Hydrological Sciences Journal, 64(10), 1141–1158, https://doi.org/10.1080/02626667.2019.1620507, 2019.

https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

Cai, H., Liu, S., Shi, H., Zhou, Z., Jiang, S., and Babovic, V.: Toward improved lumped groundwater level predictions at catchment scale: Mutual integration of water balance mechanism and deep learning method, Journal of Hydrology, 613, 128495, 2022.

Cui, T., Li, Y., Yang, L., Nan, Y., Li, K., Tudaji, M., Hu, H., Long, D., Shahid, M., Mubeen, A., He, Z., Yong, B., Lu, H., Li, C., Ni, G., Hu, C., and Tian, F.: Non-monotonic changes in Asian Water Towers' streamflow at increasing warming levels, Nat Commun, 14(1), 1176, https://doi.org/10.1038/s41467-023-36804-6, 2023.

Feigl, M., Roesky, B., Herrnegger, M., Schulz, K., and Hayashi, M.: Learning from mistakes-Assessing the performance and uncertainty in process-based models, Hydrol Process, 36(2), e14515, https://doi.org/10.1002/hyp.14515, 2022.

Feng, D., Liu, J., Lawson, K., and Shen, C.: Differentiable, Learnable, Regionalized Process-Based Models With Multiphysical Outputs can Approach State-Of-The-Art Hydrologic Prediction Accuracy, Water Resources Research, 58(10), https://doi.org/10.1029/2022wr032404, 2022.

Frame, J. M., Kratzert, F., Raney, A., Rahman, M., Salas, F. R., and Nearing, G. S.: Post-Processing the National Water Model with Long Short-Term Memory Networks for Streamflow Predictions and Model Diagnostics, JAWRA Journal of the American Water Resources Association, 57(6), 885–905, https://doi.org/10.1111/1752-1688.12964, 2021.

Hoedt, P.-J., Kratzert, F., Klotz, D., Halmich, C., Holzleitner, M., Nearing, G. S., Hochreiter, S., and Klambauer, G. (2021). Mc-lstm: Mass-conserving lstm. International conference on machine learning,

Höge, M., Scheidegger, A., Baity-Jesi, M., Albert, C., and Fenicia, F.: Improving hydrologic models for predictions and process understanding using neural ODEs, Hydrology and Earth System Sciences, 26(19), 5085–5102, https://doi.org/10.5194/hess-26-5085-2022, 2022.

Jiang, S., Zheng, Y., Wang, C., and Babovic, V.: Uncovering Flooding Mechanisms Across the Contiguous United States Through Interpretive Deep Learning on Representative Catchments, Water Resources Research, 58(1), e2021WR030185, https://doi.org/10.1029/2021wr030185, 2022.

Khatami, S., Peel, M. C., Peterson, T. J., and Western, A. W.: Equifinality and flux mapping: A new approach to model evaluation and process representation under uncertainty, Water Resources Research, 55(11), 8922–8941, 2019.

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

Kratzert, F., Klotz, D., Hochreiter, S., and Nearing, G. S.: A note on leveraging synergy in multiple meteorological data sets with deep learning for rainfall–runoff modeling, Hydrology and Earth System Sciences, 25(5), 2685–2703, https://doi.org/10.5194/hess-25-2685-2021, 2021.

Legates, D. R., and McCabe Jr, G. J.: Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation, Water Resources Research, 35(1), 233–241, 1999.

https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

Li, B., Li, R., Sun, T., Gong, A., Tian, F., Khan, M. Y. A., and Ni, G.: Improving LSTM hydrological modeling with spatiotemporal deep learning and multi-task learning: a case study of three mountainous areas on the Tibetan Plateau, Journal of Hydrology, 129401, 2023.

Li, B., Sun, T., Tian, F., and Ni, G.: Enhancing process-based hydrological models with embedded neural networks: A hybrid approach, Journal of Hydrology, 130107, 2023.

Li, B., Sun, T., Tian, F., Tudaji, M., Qin, L., and Ni, G.: Hybrid hydrological modeling for large alpine basins: a semi-distributed approach, Hydrology and Earth System Sciences, 28(20), 4521–4538, 2024.

Li, B., Zhou, X., Ni, G., Cao, X., Tian, F., and Sun, T.: A multi-factor integrated method of calculation unit delineation for hydrological modeling in large mountainous basins, Journal of Hydrology, 597, 126180, https://doi.org/10.1016/j.jhydrol.2021.126180, 2021.

McMillan, H. K., Westerberg, I. K., and Krueger, T.: Hydrological data uncertainty and its implications, Wiley Interdisciplinary Reviews: Water, 5(6), e1319, 2018.

Nash, J. E., and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology, 10(3), 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., and Gupta, H. V.: What Role Does Hydrological Science Play in the Age of Machine Learning? , Water Resources Research, 57(3), e2020WR028091, https://doi.org/10.1029/2020wr028091, 2021.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, Hydrology and Earth System Sciences, 19(1), 209–223, https://doi.org/10.5194/hess-19-209-2015, 2015.

Pokharel, S., Roy, T., and Admiraal, D.: Effects of mass balance, energy balance, and storage-discharge constraints on LSTM for streamflow prediction, Environmental Modelling & Software, 166, 105730, 2023.

Quilty, J. M., Sikorska-Senoner, A. E., and Hah, D.: A stochastic conceptual-data-driven approach for improved hydrological simulations, Environmental Modelling & Software, 149, https://doi.org/10.1016/j.envsoft.2022.105326, 2022.

Ragab, R., Kaelin, A., Afzal, M., and Panagea, I.: Application of Generalized Likelihood Uncertainty Estimation (GLUE) at different temporal scales to reduce the uncertainty level in modelled river flows, Hydrological Sciences Journal, 65(11), 1856–1871, 2020.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resources Research, 46(5), 2010.

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., Baity-Jesi, M., Fenicia, F., Kifer, D., and Li, L.: Differentiable modelling to unify machine learning and physical models for geosciences, Nature Reviews Earth & Environment, 4(8), 552–567, 2023.

Shen, C., and Lawson, K.: Applications of deep learning in hydrology, Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science, and Geosciences, 283–297, 2021.

Singh, S., Mishra, K., Chavan, R., and Tiwari, H. (2023). Advancements and challenges in hydrological modeling: a comprehensive review. International Conference on Hydraulics, Water Resources and Coastal Engineering,

Solanki, H., Vegad, U., Kushwaha, A., and Mishra, V.: Improving streamflow prediction using multiple hydrological models and machine learning methods, Water Resources Research, 61(1), e2024WR038192, 2025.

Tao, H., Abba, S. I., Al-Areeq, A. M., Tangang, F., Samantaray, S., Sahoo, A., Siqueira, H. V., Maroufpoor, S., Demir, V., and Bokde, N. D.: Hybridized artificial intelligence models with nature-inspired algorithms for river flow modeling: A comprehensive review, assessment, and possible future research directions, Engineering Applications of Artificial Intelligence, 129, 107559, 2024.

Thébault, C., Knoben, W., Addor, N., Newman, A., and Clark, M.: Varying the combination of hydrological models in time and space: Toward a more accurate representation of streamflow in large-sample hydrology, Water Resources Research, 62(5), e2025WR042272, 2026.

Tsai, W. P., Feng, D., Pan, M., Beck, H., Lawson, K., Yang, Y., Liu, J., and Shen, C.: From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling, Nat Commun, 12(1), 5988, https://doi.org/10.1038/s41467-021-26107-z, 2021.

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

Wang, H., Wang, C., Wang, Y., Gao, X., and Yu, C.: Bayesian forecasting and uncertainty quantifying of stream flows using Metropolis–Hastings Markov Chain Monte Carlo algorithm, Journal of Hydrology, 549, 476–483, 2017.

Wang, K., Onodera, S.-i., Saito, M., and Nang, Y. W.: Advancing Physical Realism in Hydrological Modelling: Selection and Integration—A Review and Synthesis, Water resources management, 40(5), 181, 2026.

Wang, Y., Wang, W., Ma, Z., Zhao, M., Li, W., and Ye, F.: Integrating multiple physical processes with deep learning for the prediction of coupled water-vapor-heat water flux in unsaturated zone, Journal of Hydrology, 661, 133654, 2025.

Wang, Y., Zhang, L., Erichson, N. B., and Yang, T.: Investigating the streamflow simulation capability of a new mass-conserving long short-term memory (MC-LSTM) model across the contiguous United States, Journal of Hydrology, 658, 133161, 2025.

Wi, S., and Steinschneider, S.: On the need for physical constraints in deep learning rainfall–runoff projections under climate change: a sensitivity analysis to warming and shifts in potential evapotranspiration, Hydrology and Earth System Sciences, 28(3), 479–503, 2024.

Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, Water Resources Research, 44(9), W09417, 2008.

https://doi.org/10.5194/egusphere-2026-4628
Preprint. Discussion started: 25 August 2026

Zhang, T., Zhang, R., Li, J., and Feng, P.: Deep learning of flood forecasting by considering interpretability and physical constraints, Hydrology and Earth System Sciences, 29(21), 5955–5974, 2025.

Zhao, T., Wang, Q., Bennett, J. C., Robertson, D. E., Shao, Q., and Zhao, J.: Quantifying predictive uncertainty of streamflow forecasts based on a Bayesian joint probability model, Journal of Hydrology, 528, 329–340, 2015.

Zheng, Y., and Keller, A. A.: Uncertainty assessment in watershed-scale water quality modeling and management: 1. Framework and application of generalized likelihood uncertainty estimation (GLUE) approach, Water Resources Research, 43(8), 2007.

Zhong, L., Lei, H., and Gao, B.: Developing a Physics-Informed Deep Learning Model to Simulate Runoff Response to Climate Change in Alpine Catchments, Water Resources Research, 59(6), https://doi.org/10.1029/2022wr034118, 2023.

Zhong, L., Lei, H., and Yang, J.: Development of a distributed physics-informed deep learning hydrological model for data-scarce regions, Water Resources Research, 60(6), e2023WR036333, 2024.