



Original Research

Physics-guided networks for probabilistic hydrodynamic forecasting in canal systems

Wangjiayi Liu^a, Guanghua Guan^{a,*}, Xiaonan Chen^b, Liangsheng Shi^a, Guangtao Fu^c, Dragan Savic^{c,d}^a State Key Laboratory of Water Resources Engineering and Management, Wuhan University, Wuhan, 430072, China^b Construction and Administration Bureau of the Middle-Route of the South-to-North Water Division Project of China, Beijing, 100038, China^c Centre for Water Systems, University of Exeter, Exeter, EX4 4QF, United Kingdom^d KWR Water Research Institute, Groningenhaven 7, Nieuwegein, 3430 BB, Netherlands

ARTICLE INFO

Article history:

Received 1 July 2025

Received in revised form

29 April 2026

Accepted 2 May 2026

Keywords:

Uncertainty quantification

Physics-guided neural network

Deep probabilistic learning

Hydrodynamic model

Canal system

ABSTRACT

Reliable prediction of water supply dynamics in large-scale canal systems is critical for water allocation and operational decision-making in inter-basin water transfer projects. Uncertainty in lateral offtake discharges evolves over time and often exhibits multi-peaked distributions due to real-time hydraulic states and unplanned gate operations. However, reliably quantifying and interpreting the evolving uncertainty remains difficult under such dynamically changing and small-sample conditions. Here we show that a physics-guided mixture density network (PgMDN) can effectively characterize this uncertainty while remaining physically consistent. In the proposed PgMDN, physical knowledge is incorporated into the loss function through local mass balance and a consistency constraint between predictions and their associated uncertainty, while long short-term memory layers are employed to model temporal dependencies and multi-factor influences. In addition, Shapley additive explanation analysis is used to identify the dominant hydraulic inputs contributing to predictive uncertainty. Tested on real-world canal datasets, the proposed PgMDN outperforms the standard mixture density network, achieving over a 25% reduction in both mean absolute error and root mean square error, together with improved reliability, as measured by the R-index (increasing from 0.45 to 0.82), and stronger generalization. The results further reveal that water level fluctuations and boundary inflow are key drivers of predictive uncertainty, supporting the physical interpretability of the proposed model. Overall, this study provides a scalable and interpretable tool for real-time modeling of environmental infrastructure and the operational management of large-scale water diversion systems.

© 2026 The Authors. Published by Elsevier B.V. on behalf of Chinese Society for Environmental Sciences, Harbin Institute of Technology, Chinese Research Academy of Environmental Sciences. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Inter-basin water transfers play a vital role in redistributing water resources across spatial and temporal scales, supporting groundwater recovery, water security, and long-term ecological resilience under climatic and socio-economic change [1–3]. Reliable hydrodynamic predictions are essential for securing water supply and providing early risk warnings [4,5].

Compared with natural rivers or watershed systems, canal systems are influenced by both natural processes and human

operations, with flow regimes shaped by gate operations and evolving hydraulic conditions. In such systems, lateral offtake discharges frequently deviate from planned supply targets. These deviations, driven by interactions with real-time system states and unplanned gate operations, produce multimodal distributions with multiple distinct peaks under given conditions [6,7]. This uncertainty degrades the accuracy of water-level predictions and may lead to ineffective operational decisions and delayed emergency responses [8]. Therefore, robust uncertainty quantification (UQ) of real-time water offtake discharges is urgently needed to support adaptive and resilient water distribution under dynamic and uncertain operational conditions.

The Monte Carlo (MC) method, a traditional numerical approach for UQ, relies on repeated random sampling to

* Corresponding author.

E-mail address: GGH@whu.edu.cn (G. Guan).

approximate the probability distributions of model predictions [9]. It is commonly applied to physics-based models, such as hydrologic simulations, by sampling inputs or parameters [10]. To improve efficiency when sampling from complex distributions, several acceleration strategies have been developed. Among these, Markov chain MC has been used to quantify prediction intervals for nonlinear flow behaviors [11–13], and Camacho et al. [14] demonstrated its effectiveness in both one- and multi-dimensional hydrodynamic modeling. Additionally, Latin hypercube sampling improves coverage of parameter spaces and reduces redundant simulations [15,16], while sequential MC enhances exploration efficiency in complex water systems [17].

Despite these advances, MC-based UQ requires carefully defined sampling ranges and remains computationally expensive. Moreover, such methods are ill-suited for predicting uncertainty in water-offtake discharges, as the absence of explicitly physical models makes it difficult to define input variables, often leading to unrealistic sampling.

As a more efficient alternative, statistical modeling approaches estimate predictive distributions directly from observed data. Parametric methods—assuming Gaussian or non-Gaussian distributions—have been widely applied in hydrodynamic modeling, including flood forecasting [18], stage-discharge relationships [19], and extreme precipitation analysis [20]. To increase flexibility, non-parametric methods that avoid strict distributional assumptions have also been explored. For example, data-driven classification approaches have been used to generate probabilistic flood forecasts from historical model errors [21], generalized likelihood uncertainty estimation has been linked with Bayesian inference in hydrological modeling [22], and bootstrapping techniques have been combined with hydrodynamic models for flood warning systems [23]. However, because these methods primarily rely on historical patterns and neglect dynamic interactions within the system, they struggle to capture the multimodal characteristics of future offtake discharges driven by real-time hydraulic conditions.

With the rapid adoption of deep learning for modeling complex, implicit input–output relationships, Bayesian neural networks (BNNs) and other stochastic, data-driven approaches have emerged as powerful tools for UQ [24,25]. Bayesian long short-term memory (LSTM) networks have been applied to simulate uncertainty in streamflow for data-scarce basins [26] and in soil moisture dynamics [27]. As an approximate Bayesian inference technique, MC dropout simulates an ensemble of networks by randomly deactivating neurons, enabling estimation of posterior distributions in hydrological models [28,29] and water distribution failure analysis [30]. To better represent real-world distributions, variational Bayesian inference has been used to model uncertainty in LSTM weights [31], while Gaussian mixture models have been integrated with LSTM networks for short-term forecasting [32].

McLachlan et al. [33] demonstrated that finite mixtures of Gaussian distributions can approximate any continuous probability distribution. Building on this principle, the mixture density network (MDN), originally proposed by Bishop [34], has been increasingly applied to improve UQ in hydrodynamic modeling. Due to their flexibility in representing multi-modal distributions, MDNs efficiently have been successfully used in applications such as stage-discharge curve estimation [35], hydraulic-geological modeling [36], and streamflow prediction [37,38]. Furthermore, comparative studies have shown that MDNs outperform MC dropout-based LSTM models in capturing complex and arbitrary distributions [39].

However, conventional MDN and BNN models often lack physical constraints and depend heavily on large, high-quality

datasets, limiting their reliability in real-time hydrodynamic prediction. Although recent studies in other domains have incorporated physical knowledge into probabilistic neural networks to improve the reliability and interpretability [40,41], these approaches are typically designed for static tasks with single-variable inputs. In canal systems, uncertainty in offtake discharges arises from dynamic, multivariate hydraulic interactions, presenting additional challenges compared to natural hydrological systems. Consequently, existing approaches remain insufficient for modeling uncertainty in real-time offtake discharge prediction.

To address these limitations, this study proposes a physics-guided mixture density network (PgMDN) that integrates physical knowledge into probabilistic learning, enabling reliable real-time uncertainty modeling in canal systems. The proposed model captures dynamic, multivariate interactions while maintaining physically plausible predictions under data-limited conditions. In addition, the Shapley additive explanations (SHAP) analysis is employed to identify key hydraulic drivers of predictive uncertainty. The performance of the PgMDN is evaluated using real-world canal system data, demonstrating improvements in accuracy, reliability, and generalization.

2. Methodology

2.1. Uncertainty in the offtake discharge prediction

The hydrodynamic behavior of canal systems is governed by unsteady discharges through the upstream control gate (Q_u), the downstream control gate (Q_d), and lateral offtakes (q) [42]. To minimize water-level fluctuations caused by water withdrawal, offtakes are typically located near the downstream end of canal reaches [43]. In practice, Q_u and Q_d are often predetermined to meet water-supply or flood-control objectives. However, the offtake discharge q exhibits a multimodal distribution due to the influence of real-time hydraulic conditions and unplanned gate operations. Considering flow time travel delays and uncertainty in system information, the offtake discharge q in real-world systems can be expressed as follows:

$$\tilde{q}_t = Q_{u,(t-t_d)} - Q_{d,t} + \varepsilon_{q,t} \quad (1)$$

where t is the time step; t_d represents the delay time step associated with water flow from the upstream end to the downstream end; $\varepsilon_{q,t}$ denotes the residual error in the predictive offtake discharge at time step t ; and $\tilde{q}_t$ is a stochastic variable ($\text{m}^3 \text{s}^{-1}$).

To quantify the inherent uncertainty in the predictive q , a cumulative distribution function (CDF) is used to characterize its probability distribution, obtained by integrating the corresponding probability density function. This probabilistic formulation enables representation of the stochastic and potentially multimodal nature of q . At each time step t , conditioned on a predictor vector (denoted $\mathbf{x}_t$), the model outputs a conditional distribution $p(q_t|\mathbf{x}_t)$ with nonlinear, time-varying statistical characteristics (e.g., mean and variance).

2.2. Physics-guided mixture density network

2.2.1. Standard MDN architecture

The MDN architecture employed in this study consists of four layers: one input layer, two hidden layers, and one output layer. Based on the availability of real-time sensor measurements, the input layer comprises water-level variations at the upstream and downstream (ΔH_u and ΔH_d); planned discharges at the upstream and downstream (Q_u and Q_d); the downstream water level (H_d);

and a nonlinear interaction term ($H_d\Delta H_d$).

These variables collectively represent the hydraulic state of the system and the scheduled control actions, providing information relevant to subsequent offtake discharge. In particular, H_d and $H_d\Delta H_d$ characterize downstream hydraulic conditions, which directly influence offtake flow, as offtakes are typically located near the downstream end. The selected variables enable the model to capture both natural and operational influences embedded in the hydraulic state.

Because water-level states are derived from past observations, whereas discharge plans are specified in advance, the predictor vector at time step t includes variables from different time steps. Specifically, ΔH_u , ΔH_d , H_d , and $H_d\Delta H_d$ are taken from the time step $t-1$, whereas Q_u and Q_d correspond to time step t . Accordingly, the predictor vector $\mathbf{x}_t$ is constructed using variables from both t and $t-1$, as expressed in equation (2):

$$\mathbf{x}_t = \{\Delta H_{u,t-1}, \Delta H_{d,t-1}, Q_{u,t}, Q_{d,t}, H_{d,t-1}, (H_d\Delta H_d)_{t-1}\} \quad (2)$$

where $\Delta H_{u,t-1}$ and $\Delta H_{d,t-1}$ denote changes in upstream and downstream water levels at time step $t-1$ (m), respectively; $Q_{u,t}$ and $Q_{d,t}$ denote planned discharges through the upstream and downstream control gates at time step t ($\text{m}^3 \text{s}^{-1}$), respectively; and $H_{d,t-1}$ denotes the downstream water level at time step $t-1$ (m).

As a flexible probabilistic framework, the MDN incorporates two stacked LSTM layers to capture temporal dependencies in the input sequence. The output of the second LSTM layer is mapped to the parameters of the mixture distribution. Each LSTM processes a historical input sequence spanning T time steps ($\mathbf{x}_{t-T}, \dots, \mathbf{x}_t$), and produces a multimodal predictive distribution for q at the t th time step (q_t) in a sequence-to-one architecture. Key hyperparameters—including the number of neurons per layer, number of training epochs, learning rate, and batch size—are tuned during model development [39]. The MDN outputs an estimated CDF represented as a mixture of component distributions with different weights [44]. The conditional distribution $p(q_t|\mathbf{x}_t)$ can be expressed as equations (3) and (4):

$$(q_t|\mathbf{x}_t) = \sum_{i=1}^M \omega_{i,t}(\mathbf{x}_t) \phi(q_t|\theta_{i,t}(\mathbf{x}_t)) \quad (3)$$

$$\sum_{i=1}^M \omega_{i,t}(\mathbf{x}_t) = 1, 0 \leq \omega_{i,t}(\mathbf{x}_t) \leq 1 \quad (4)$$

where M denotes the number of mixture components and $\omega_{i,t}(\mathbf{x}_t)$ is the mixing weight of the i th kernel $\phi(\bullet)$ at time step t . The kernel function $\phi(\bullet)$ can be any type of parametric density distribution and is governed by a vector of parameters $\theta_{i,t}(\mathbf{x}_t)$ (e.g., mean μ and standard deviation σ).

In this study, the number of mixture components (M) is set to 3. The kernel function $\phi(\bullet)$ follows a Gaussian distribution due to its flexibility and effectiveness in probabilistic modeling [45]. $\theta_{i,t}(\mathbf{x}_t)$ can be further expressed as $(\mu_{i,t}, \sigma_{i,t}^2)$, where $\mu_{i,t}(\mathbf{x}_t)$ and $\sigma_{i,t}^2(\mathbf{x}_t)$ indicate the mean and variance of the i th Gaussian component, respectively. Accordingly, outputs of the MDN contain a set of time-varying parameters $(\omega_{i,t}, \mu_{i,t}, \sigma_{i,t}^2)$. To ensure valid parameterization, activation functions are applied in the output layer: A Softmax function enforces that the mixture weights $\omega_{i,t}$ sum to 1, while a Softplus function ensures positivity of the variances $\sigma_{i,t}^2$; the means $\mu_{i,t}$ retain their original values [34]. Model training is performed using the negative log-likelihood (NLL) function based on the observed data:

$$N_{\text{data}} = - \sum_{i=1}^N \ln \left[\sum_{i=1}^M \omega_{i,t}(\mathbf{x}_t) g(q_{o,t}|\mu_{i,t}(\mathbf{x}_t), \sigma_{i,t}^2(\mathbf{x}_t)) \right] \quad (5)$$

where N_{data} denotes the NLL term based on observed data; $g(\bullet)$ denotes the Gaussian distribution; q_o represents observed values of q in $\text{m}^3 \text{s}^{-1}$; and N is the total number of time steps.

The trained model is then used to predict the distribution parameters of the conditional distribution $p(q_t|\mathbf{x}_t)$. Based on these parameters, quantiles of the mixture distribution can be computed to quantify predictive uncertainty. The mixture-weighted mean (μ) represents the expected value of q , while the mixture-weighted variance (σ^2) reflects uncertainty. Because the inverse CDF of the mixture distribution lacks a closed-form solution, the 90% prediction interval is approximated using MC sampling with 10,000 draws from the fitted Gaussian mixture.

2.2.2. MDN with physical constraints

To improve real-time prediction of offtake discharge under limited data conditions, we incorporate local flow continuity constraints and a coupling relationship between predictive mean variation and uncertainty into the MDN framework, resulting in the proposed PgMDN model. While retaining most components of the original MDN architecture, the PgMDN introduces two physics-based constraint terms into a unified loss function to guide model training. These additional terms influence the back-propagation process by adjusting internal network parameters during forward propagation, thereby enhancing the physical consistency of the model predictions (Fig. 1); further details are provided below.

- (1) A physical NLL term based on the mass balance equation at the offtake

Unlike the standard MDN loss function, which minimizes only the discrepancy between predicted means and field-observed

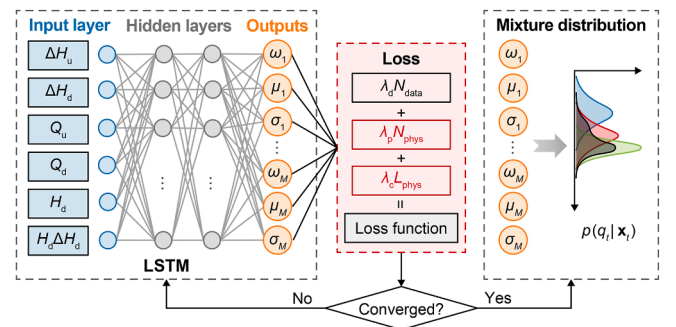


Fig. 1. Architecture of the PgMDN, showing the input variables, LSTM–MDN structure, physics-guided loss function, and conditional output distribution. The network outputs the mixture weights, means, and standard deviations of the Gaussian components to parameterize the conditional probability distribution $p(q_t|\mathbf{x}_t)$. PgMDN, physics-guided mixture density network; LSTM, long short-term memory; MDN, mixture density network; ΔH_u , upstream water-level variation; ΔH_d , downstream water-level variation; Q_u , upstream discharge; Q_d , downstream discharge; H_d , downstream water level; $H_d\Delta H_d$, interaction term between downstream water level and downstream water-level variation; ω , mixture weight; μ , mean; σ , standard deviation; M , number of Gaussian components; N_{data} , the NLL term based on observed data; N_{phys} , the physical NLL term derived from local mass balance; NLL, negative log-likelihood; L_{phys} , physics-guided consistency term between mean variation and predictive uncertainty; λ_d , λ_p , and λ_c , the weighting coefficients of the data NLL term, the physical NLL term, and the physics-guided term, respectively.

values, the proposed physical NLL term incorporates a governing physical equation as an additional training objective. Based on the mass balance principle, the predicted mean values of q from the PgMDN are encouraged to be consistent with the difference between inflow (Q_{in}) and outflow (Q_{out}) within the computational grid containing the offtake.

The values of Q_{in} and Q_{out} are generated using a physics-based model derived from one-dimensional (1D) hydrodynamics. This model is based on simplified hydraulic assumptions and is driven by observed data from a real-world canal system, including boundary discharges, initial water surface profile, and offtake discharges during the training period. Notably, observed water levels capture complex hydraulic effects present in real canal systems, including unmodeled losses, and are therefore used as key inputs for offtake discharge prediction. Based on these inputs, Q_{in} and Q_{out} are simulated by solving the 1D discrete Saint-Venant equations using the four-point Preissmann scheme [46].

This physical constraint is incorporated into the loss function via an NLL formulation similar to equation (5), with the observed-value term replaced by the mass-balance term. Rather than enforcing strict equality, this formulation promotes consistency between predicted mean values and underlying physical laws, as expressed in equation (6):

$$N_{phys} = - \sum_{i=1}^N \ln \left[\sum_{i=1}^M \omega_{i,t}(\mathbf{x}_t) g \left((Q_{in} - Q_{out})_t | \mu_{i,t}(\mathbf{x}_t), \sigma_{i,t}^2(\mathbf{x}_t) \right) \right] \quad (6)$$

where N_{phys} denotes the physical NLL term derived from local mass balance; Q_{in} and Q_{out} denote the inflow and outflow discharges within the computational grid containing the offtake ($m^3 s^{-1}$), respectively.

Although this constraint is explicitly applied only to the predicted mean, it also influences the variance due to the internal coupling between mean and variance within the joint output layer of the PgMDN.

- (2) A physics-guided term constraining the relationship between changes in predicted means ($\Delta\mu$) and the corresponding standard deviation (σ)

Significant variations in the predicted mean values of q often reflect rapid changes in the underlying hydraulic state of canal systems [6]. Such fluctuations are typically associated with dynamic transitions and potential unplanned or abrupt gate operations. Consequently, large changes in predicted mean values can be interpreted as indicators of reduced system stability, under which predictive uncertainty is expected to increase. Based on this physical insight, a positive correlation is assumed between the magnitude of changes in predicted discharge (q) and the corresponding uncertainty.

To incorporate this assumption into model training, a physics-guided term is added to the loss function to enforce consistency across predictions. This term penalizes the model when the evolution of predicted means and predictive uncertainty follows opposing trends. Specifically, increases (or decreases) in mean fluctuations should correspond to wider (or narrower) uncertainty intervals. This is implemented by encouraging directional consistency between changes in predicted means ($\Delta\mu$) and the standard deviation (σ), such that the loss approaches zero only when both vary in the same direction.

In addition, a parameter α is introduced to enable adaptive learning by imposing stronger constraints in regions that exhibit sharper variation during training. This mechanism allows the

model to focus on dynamically unstable regions rather than uniformly inflating uncertainty and producing overly conservative predictions. The physics-guided term is defined in equations (7) and (8), where μ and σ denote the mixture-weighted values of μ_i and σ_i , respectively:

$$L_{phys} = \sum_{t=2}^N \alpha_t (E_{\Delta\mu,t} \cdot E_{\sigma,t} - |E_{\Delta\mu,t}| \cdot |E_{\sigma,t}|)^2 \quad (7)$$

$$\alpha_t = \frac{e^{|E_{\Delta\mu,t}|}}{\sum_{t=2}^N e^{|E_{\Delta\mu,t}|}}, E_{\Delta\mu,t} = \Delta\mu_t - \Delta\mu_{t-1}, \Delta\mu_t = |\mu_t - \mu_{t-1}|, E_{\sigma,t} = \sigma_t - \sigma_{t-1} \quad (8)$$

where L_{phys} denotes physics-guided consistency term between mean variation and predictive uncertainty; α_t regulates the strength of the constraint; e denotes the base of the natural logarithm; $E_{\Delta\mu,t}$ and $E_{\sigma,t}$ represent the change trends of predicted mean variation and predicted variance, respectively; and $\Delta\mu_t$ denotes the absolute difference between adjacent predicted means.

The final loss function for PgMDN training is formulated as a sum of the data NLL term, the physical NLL term, and the physics-guided term, as shown in equation (9). In this study, all loss components are assigned equal weights to ensure comparable contributions during training. The back-propagation algorithm, therefore, seeks to minimize both data-fitting loss and physics-based constraint losses simultaneously.

$$L = \lambda_d N_{data} + \lambda_p N_{phys} + \lambda_c L_{phys} \quad (9)$$

where λ_d , λ_p , and λ_c denote the weighting coefficients of the data NLL term, the physical NLL term, and the physics-guided term, respectively. In this study, $\lambda_d = \lambda_p = \lambda_c = 1$.

2.3. SHAP analysis

Identifying key contributors to model uncertainty is essential for understanding the behavior of the PgMDN and improving the credibility of its predictions. To this end, SHAP is employed to interpret the influence of each input variable on the predictive uncertainty interval, thereby enhancing the interpretability of the trained model [47].

Based on cooperative game theory, SHAP computes Shapley values to quantify each feature's average marginal contribution to model predictions [48]. This approach evaluates model behavior across all possible feature combinations, both including and excluding the feature of interest.

In this study, SHAP analysis is implemented for the mixture-weighted variance (σ^2) predicted by the PgMDN. Shapley values quantify the relative importance of each input variable in influencing predictive uncertainty. By examining the consistency of feature importance across days, the dominant hydraulic factors governing uncertainty in water-offtake discharge can be identified.

2.4. Evaluation metrics

The MDN and PgMDN models are evaluated based on predictions of q from two perspectives: accuracy and uncertainty. Predictive accuracy is assessed using the mean absolute error (MAE) and the root mean square error (RMSE). In addition, the reliability index (R-index) and skill score (S-index) are used to evaluate model performance in uncertainty forecasting [49]. All metrics are computed across all prediction time steps, yielding a single evaluation value for each model.

MAE and RMSE quantify the average absolute and squared

differences between predictions and observations, respectively, with lower values indicating better predictive performance. These metrics are defined in equations (10) and (11):

$$E_{MAE} = \frac{1}{N_p} \sum_{t=1}^{N_p} |q_{o,t} - q_{p,t}| \quad (10)$$

$$E_{RMSE} = \sqrt{\frac{1}{N_p} \sum_{t=1}^{N_p} (q_{o,t} - q_{p,t})^2} \quad (11)$$

where the N_p denotes the total number of prediction samples, and q_o and q_p represent observed and predicted values, respectively.

The R -index measures the proportion of observed data points that fall within the predicted uncertainty interval. A value closer to the predetermined confidence level indicates better agreement with the theoretical expectation; positive deviations are considered favorable, whereas negative deviations indicate under-coverage. Given sufficient reliability, a narrower uncertainty interval reflects improved predictive performance.

The S -index provides a comprehensive evaluation of both reliability and interval width, with values closer to zero indicating better overall performance. A positive S -index corresponds to confident (narrow) predictions, whereas a negative value indicates conservative (overly wide) intervals. Thus, values near zero reflect an optimal balance between reliability and sharpness [49].

Given a confidence level τ , the R -index R_τ and the S -index S_τ are defined as follows:

$$\delta_\tau = \begin{cases} 1, & \text{if } q_{o,t} \leq \hat{q}_{\tau,t} \\ 0, & \text{otherwise} \end{cases}, R_\tau = \sum_{t=1}^{N_p} \frac{\delta_\tau}{N_p} \quad (12)$$

$$S_\tau = \frac{1}{N_p} \sum_{t=1}^{N_p} (\delta_\tau - \tau) (q_{o,t} - \hat{q}_{\tau,t}) \quad (13)$$

where τ is the prescribed confidence level (0.9 in this study); $\hat{q}_{\tau,t}$ denotes the bounds of the uncertainty interval at a confidence level τ for time step t ($\text{m}^3 \text{s}^{-1}$); and δ_τ is a binary indicator specifying whether the observed value falls within the predicted interval.

3. Case study

In this study, small synthetic datasets generated by analytical functions are used to evaluate the MDN's baseline capability to learn diverse distribution patterns under limited data conditions. These datasets reflect representative statistical characteristics of real systems. In addition, real-world canal data are used to compare the MDN and PgMDN models, demonstrating the practical performance improvements achieved by the proposed approach.

3.1. Synthetic data

We generated three types of synthetic datasets.

- (1) Case 1, a linear function with heteroscedastic Gaussian error;
- (2) Case 2, a nonlinear function with a non-Gaussian, multimodal error distribution; and
- (3) Case 3, a nonstationary time series with abrupt changes and multimodal noise.

The equations for Cases 1 and 2 are adopted from Li et al. [38], while Case 3 is constructed using piecewise functions incorporating

regime switching and jump terms. Input variables are randomly generated within specified ranges: [0, 1] for Case 1 and [-1, 1] for Cases 2 and 3. The number of training and testing samples is set to match the number of samples in the real canal system data used for real-time prediction, as described in Section 3.3.

3.2. Real-world data

The proposed models are evaluated using observational data from two canal reaches (Reach A and Reach B) along the Middle Route of the South-to-North Water Diversion Project of China (Fig. 2). These two reaches are located in the Beijing–Shijiazhuang section of the project in Hebei Province, North China, and are representative of regulated reaches within this large canal system. Although both reach the same system, they differ in geometric configuration and operational characteristics and are therefore treated as distinct test cases.

Reach A is characterized by relatively large offtake discharges with modest short-term variability, representing a substantial proportion of the main canal flow. A continuous nine-day dataset collected under typical operating conditions is used for detailed model comparison. In contrast, Reach B is represented by a longer 27-day dataset collected in a different season. This dataset exhibits greater variability, including abrupt changes and pronounced fluctuations associated with non-routine operational events. Additionally, the offtake discharge in Reach B constitutes a smaller proportion of the main canal flow compared to Reach A. The extended dataset is primarily used to evaluate model performance and robustness across a wider range of operating scenarios. The geometric characteristics, flow ranges, and data durations for both reaches are summarized (Table 1).

Both case studies share several common features: (1) trapezoidal cross-sections with wide and shallow profiles; (2) offtakes located near the downstream end; and (3) mild, nearly uniform bed slopes of 0.00004. Due to the limited sensor availability and monitoring data, this study excludes the effects of evaporation, leakage, and other loss processes in predicting offtake discharge.

3.3. Simulation setup and model training

For real-time prediction, the output time step of both the MDN

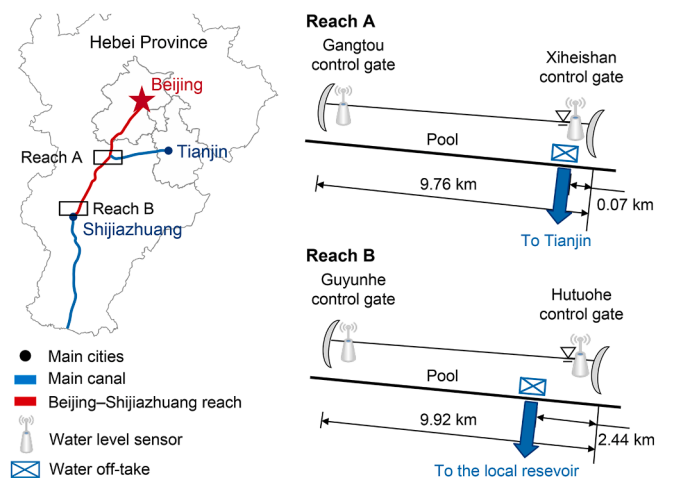


Fig. 2. Location and schematic configuration of the two study reaches (Reach A and Reach B) in the Beijing–Shijiazhuang section of the Middle Route of China's South-to-North Water Diversion Project. The figure identifies the upstream and downstream control gates, water-level sensor locations, offtake positions and directions, and key reach lengths.

Table 1
Geometric and hydraulic characteristics of Reach A and Reach B.

Reach	Bottom width (m)	Side slope	Bed slope	Q_u ($\text{m}^3 \text{ s}^{-1}$)	Q_d ($\text{m}^3 \text{ s}^{-1}$)	H_u (m)	H_d (m)	q ($\text{m}^3 \text{ s}^{-1}$)
Reach A	19.5	2.5	0.00004	79–88	45–54	3.6–3.9	4.2–4.5	30–35
Reach B	10.0	3.0	0.00004	153–178	130–175	4.8–5.2	5.0–5.4	5–27

Note: Q_u , future discharge plans at the upstream end; Q_d , future discharge plans at the downstream end; H_u , upstream water level; H_d , downstream water level; q , offtake discharge.

and PgMDN models is set to 20 min, consistent with the data acquisition frequency. The LSTM layers use historical input sequences spanning six time steps ($T = 6$), corresponding to a 2-h window.

We adopted a rolling-horizon prediction scheme, in which fixed-length training, validation, and test sets are sequentially updated to simulate real-time operation. An initial three-day training period is used to initialize the models. For Reach A, this is followed by a six-day rolling prediction period based on the nine-day dataset, using fixed-length training (72 h), validation (12 h, sixfold), and test (24 h) sets. For Reach B, the same approach is applied, resulting in a 24-day rolling prediction period enabled by the extended dataset. Under this framework, geophysical parameters are assumed to remain approximately constant within each analyzed time window.

Both MDN and PgMDN models are trained using the adaptive moment estimation (Adam) optimizer [50]. The learning rate, first-moment decay rate, and second-moment decay rate are set to 0.01, 0.9, and 0.999, respectively, based on established practice and validation results. Hyperparameters for the LSTM layers are determined through grid search within ranges informed by prior studies [51]. Specifically, the number of hidden units is tested at {40, 60, 80, 100}, the number of epochs at {100, 150, 200, 300}, and the L2 regularization weight at $\{10^{-4}, 5 \times 10^{-4}, 10^{-3}, 5 \times 10^{-3}\}$.

Hyperparameter selection is conducted using sixfold cross-validation within each 72-h training window, constituting an automated tuning procedure. Due to the relatively high dimensionality of the output, a larger number of hidden units is required to ensure sufficient model capacity. The final configuration includes 80 units per LSTM layer, 200 training epochs, a full batch size of 216 samples, and an L2 regularization weight of 0.001.

All input features are normalized to the range [0,1] prior to training. No early stopping or learning rate decay is applied. These settings are used consistently for both the MDN and PgMDN models. All numerical experiments are conducted in MATLAB R2019b on a Windows system equipped with an Intel Core i7-9700 CPU (3.00 GHz) and 16 GB of random access memory.

4. Results

4.1. Application of MDN to synthetic data

The MDN model exhibits distinct predictive performance and associated uncertainty intervals across the three synthetic data patterns (Fig. 3). Across the three cases, the MDN model performs better in Cases 1 and 2, which are characterized by simpler functional relationships, as reflected by higher reliability and skill scores closer to zero. In contrast, Case 3, which represents a more complex nonstationary time series, exhibits higher MAE and RMSE values along with lower reliability.

These results indicate that while the MDN can capture arbitrary and non-standard distributions, its predictive accuracy and uncertainty calibration degrade as data complexity and temporal variability increase.

Although these synthetic experiments demonstrate the MDN's ability to learn complex conditional distributions, they are conducted under simplified conditions. We therefore next examined the performance of both MDN and PgMDN in real-world systems (Section 4.2).

4.2. Comparison of MDN and PgMDN using real system data

4.2.1. Prediction performance on the short-term dataset

Following the training setup described in Section 3.3, this section presents model performance over a continuous six-day prediction period.

The predicted q values and their associated uncertainty intervals evolve over time, along with the distribution of predictive σ values across different variation phases of q (Fig. 4). The PgMDN model more accurately captures short-term fluctuations in q than the MDN model, particularly during the 24–48 h and 72–144 h periods. In addition, PgMDN provides more consistent and realistic uncertainty intervals compared to MDN.

For example, MDN produces abrupt, overly narrow uncertainty intervals between 72 h and 96 h, while generating excessively wide intervals during the other periods (Fig. 4a). This behavior is further illustrated in the corresponding box plots. The MDN exhibits relatively concentrated distributions of predicted σ values, with no clear relationship to variations in q . Notably, it produces the smallest spread in predicted σ during periods of high variability (Fig. 4a), indicating limited responsiveness of uncertainty estimates to dynamic system conditions.

In contrast, the PgMDN exhibits progressively wider σ distributions as the variability of q increases (Fig. 4b). This behavior suggests that the model appropriately adjusts predictive uncertainty in response to dynamic system changes. These results demonstrate that incorporating physical constraints enables the PgMDN to better capture uncertainty under rapidly evolving hydraulic conditions.

MDN and PgMDN are compared in terms of both accuracy and uncertainty metrics (Fig. 5). The PgMDN achieves significantly improved prediction accuracy ($MAE = 0.20$, $RMSE = 0.26$; Fig. 5a), reducing the MDN's corresponding errors by 28% and 26%, respectively.

From the perspective of uncertainty quantification, PgMDN also demonstrates substantial improvements. It achieves a higher R -index and a lower absolute S -index compared to MDN. At the 90% confidence level, reliability increases from 0.45 to 0.82. Furthermore, the negative S -index value, which is close to zero, indicates that PgMDN produces more conservative yet well-balanced uncertainty intervals relative to MDN.

The scatter points represent daily performance across six forecasting windows (Fig. 5). PgMDN exhibits less day-to-day variation in MAE and RMSE than MDN, with even greater stability observed in the R -index and S -index. Overall, consistent improvements across all evaluation metrics demonstrate that PgMDN outperforms MDN in terms of accuracy, reliability, and robustness.

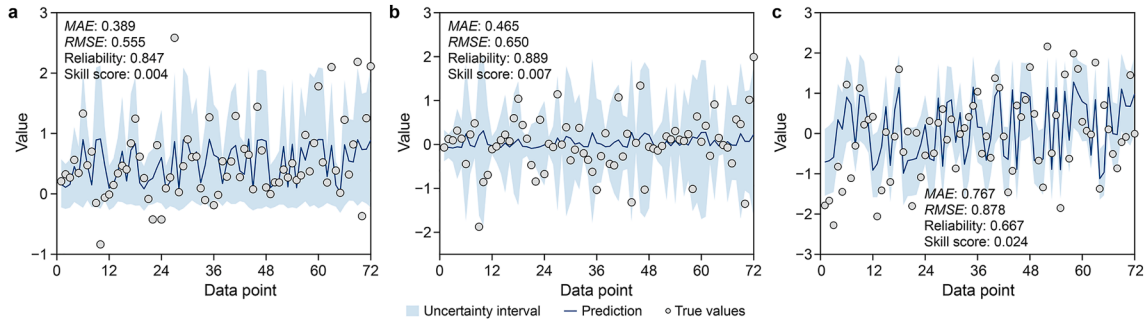


Fig. 3. Predictive performance of the MDN model for Case 1 (a), Case 2 (b), and Case 3 (c), together with the associated 90% predictive uncertainty intervals. MDN, mixture density network; MAE, mean absolute error; RMSE, root mean square error.

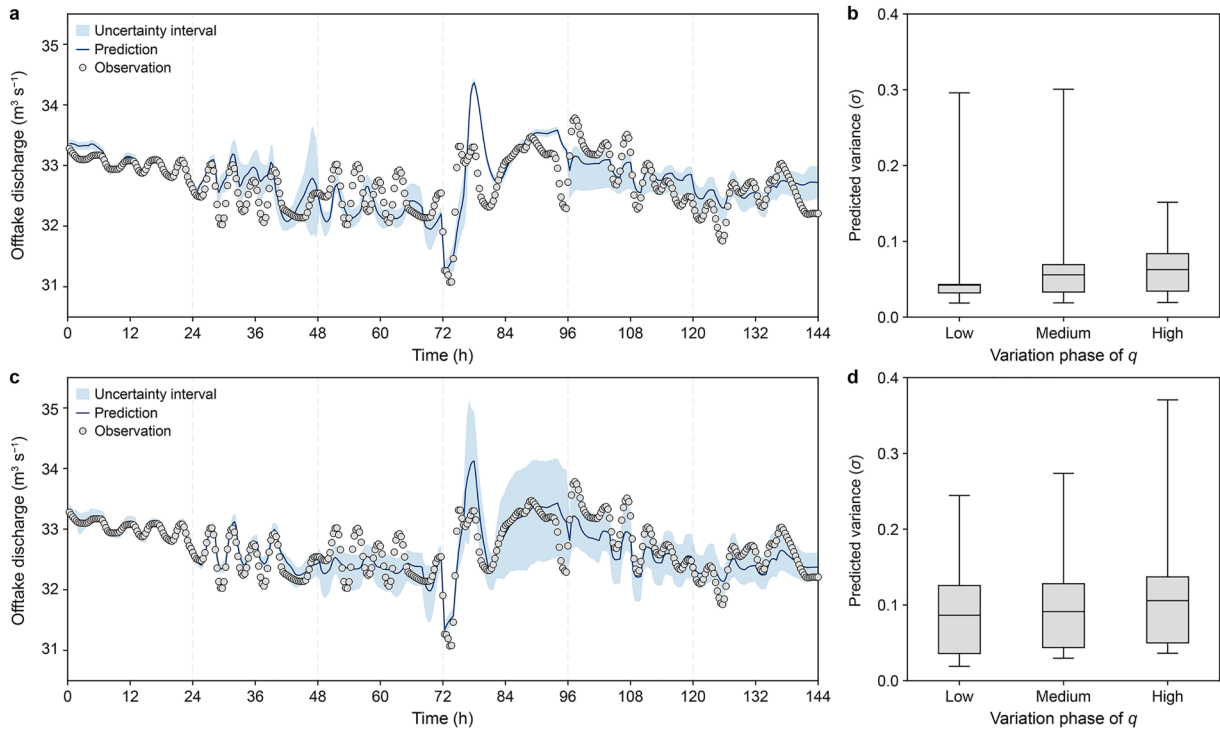


Fig. 4. Comparison of MDN and PgMDN predictions for Reach A. **a,c**, Time-series predictions from the MDN (a) and PgMDN (c), shown together with observations and 90% predictive uncertainty intervals across 432 test samples. **b,d**, Distributions of the predicted variances (σ) from the MDN (b) and PgMDN (d) for low, medium, and high variation phases of the offtake discharge (q). Variation phases were defined by the rate of change of q : low, $0.00\text{--}0.06\text{ m}^3\text{ s}^{-1}\text{ h}^{-1}$ (188 samples); medium, $0.06\text{--}0.30\text{ m}^3\text{ s}^{-1}\text{ h}^{-1}$ (172 samples); and high, $0.30\text{--}2.00\text{ m}^3\text{ s}^{-1}\text{ h}^{-1}$ (72 samples). MDN, mixture density network; PgMDN, physics-guided mixture density network.

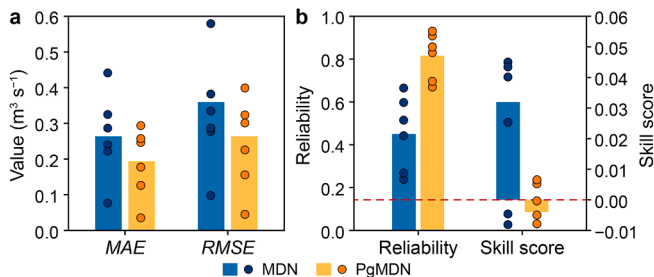


Fig. 5. Performance comparison between MDN and PgMDN for Reach A. **a**, Accuracy metrics: MAE and RMSE. **b**, Uncertainty metrics: reliability and skill score. Bars represent global metrics across the 6-day prediction period (432 samples), while scattered dots represent daily evaluation values (72 samples). The red dashed line indicates the zero reference for the skill score. MDN, mixture density network; PgMDN, physics-guided mixture density network. MAE, mean absolute error; RMSE, root mean square error.

4.2.2. Performance validation under extended conditions

The prediction performance of MDN and PgMDN over the 24-day forecasting period in Reach B is compared (Fig. 6). The PgMDN achieves substantially lower prediction errors, reducing MAE and RMSE from 0.99 and $1.48\text{ m}^3\text{ s}^{-1}$ (MDN) to 0.54 and $0.75\text{ m}^3\text{ s}^{-1}$, respectively. In addition, PgMDN provides more reliable uncertainty estimates, increasing reliability from 0.64 to 0.78 and reducing the skill score from 0.12 to 0.03 . The predicted uncertainty intervals produced by PgMDN adapt more effectively to abrupt fluctuations, particularly during periods of significant operational change. This indicates improved stability and more accurate characterization of uncertainty over extended time horizons.

To further assess model robustness under varying operating conditions, the 24-day dataset is categorized into low, medium, and high variability levels based on the rate of change in offtake discharge. PgMDN consistently achieves higher reliability and

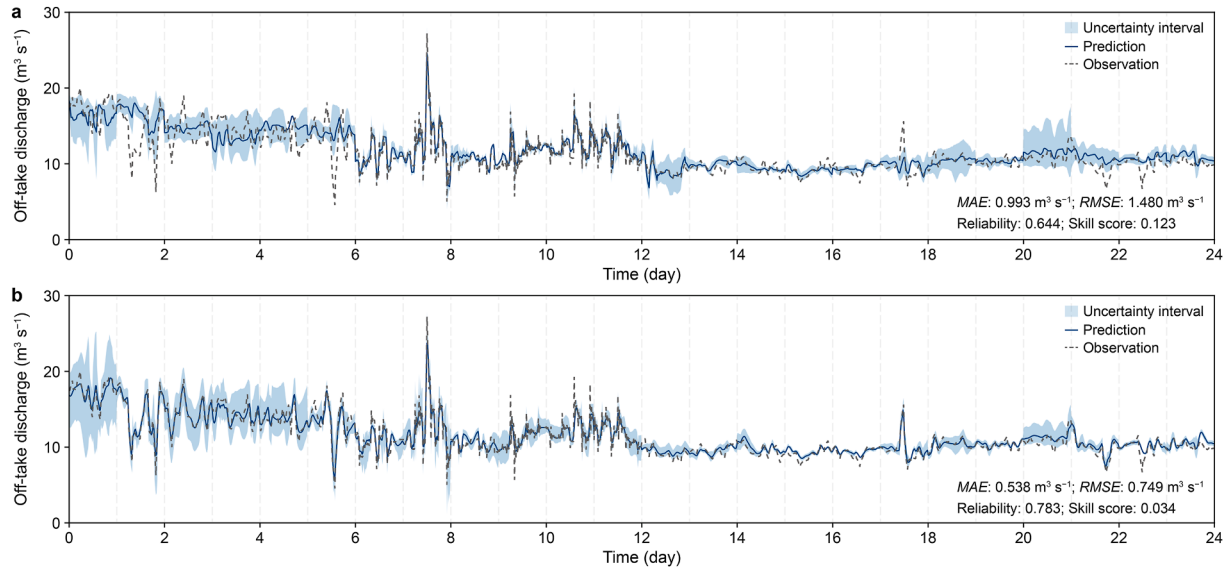


Fig. 6. Comparison of MDN and PgMDN predictions for Reach B over the 24-day test period. Predictions from the MDN (a) and PgMDN (b) compared with observations, with 90% predictive uncertainty intervals, over the 24-day test period (1728 samples). MDN, mixture density network; PgMDN, physics-guided mixture density network. MAE, mean absolute error; RMSE, root mean square error.

improved uncertainty performance than MDN across all variability levels (Fig. 7). Specifically, PgMDN achieves reliability values of approximately 0.74 under low-level variability and up to 0.84 under high variability. In contrast, MDN exhibits lower and more variable coverage across different conditions. In addition, PgMDN produces skill scores closer to zero, with the best performance observed under high-variability conditions. These findings demonstrate that the proposed PgMDN model improves uncertainty quantification under both stable and highly dynamic operating conditions.

4.2.3. Model generalization ability

We conducted two tests to evaluate model generalization. In the first test, both MDN and PgMDN are trained on data from Reach A during Day 1 (0–24 h), which represent relatively stable conditions, and then tested on Day 4 (72–96 h), characterized by sharp fluctuations. In the second test, both models are trained on data from Days 3–5 in Reach B under typical regular conditions and evaluated on Day 8, which exhibits pronounced variability and abrupt changes. These tests assess the robustness of PgMDN relative to MDN under unfamiliar and rapidly changing operational dynamics.

Prediction errors increase for both models under generalization scenarios (Fig. 8), reflecting the increased difficulty of forecasting under unseen conditions. This effect is more pronounced for Reach

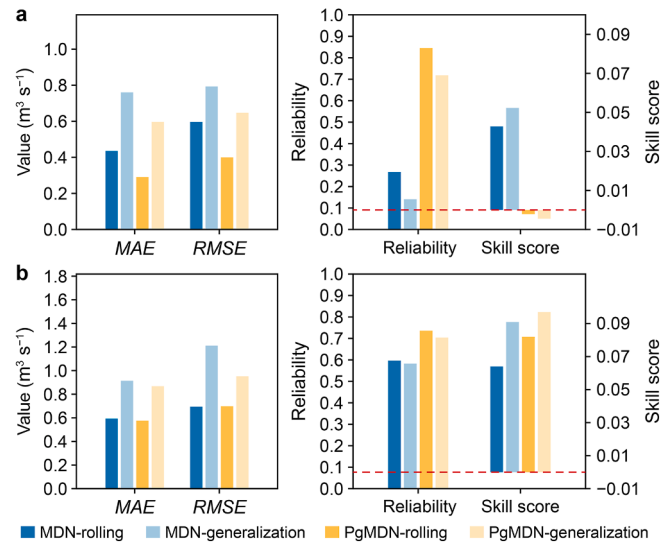


Fig. 8. Comparison of MDN and PgMDN performance under rolling and generalization tests in Reach A (a) and Reach B (b). The red dashed line indicates the zero reference for the skill score. MDN, mixture density network; PgMDN, physics-guided mixture density network. MAE, mean absolute error; RMSE, root mean square error.

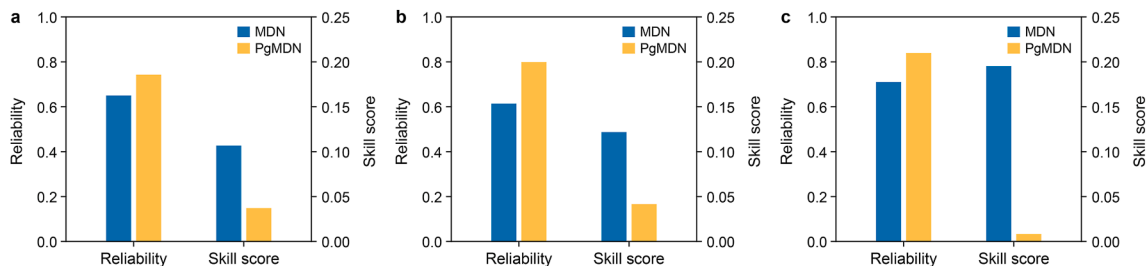


Fig. 7. Comparison of MDN and PgMDN performance in Reach B across different variation phases of the offtake discharge (q). Uncertainty metrics (reliability and skill score) for low (a), medium (b), and high (c) variation phases. Variation phases were defined by the rate of change of q : low, $0.0\text{--}0.2\text{ m}^3\text{ s}^{-1}\text{ h}^{-1}$ (864 samples); medium, $0.2\text{--}1.0\text{ m}^3\text{ s}^{-1}\text{ h}^{-1}$ (604 samples); and high, $1.0\text{--}9.0\text{ m}^3\text{ s}^{-1}\text{ h}^{-1}$ (260 samples). MDN, mixture density network; PgMDN, physics-guided mixture density network.

B, which exhibits more frequent and intense fluctuations, making it a more challenging test case.

Despite these challenges, PgMDN consistently achieves lower MAE and RMSE than MDN across both reaches and test scenarios. In terms of uncertainty quantification, PgMDN maintains stable reliability and skill score values. Even under highly dynamic conditions in Reach B, PgMDN preserves reliability levels of approximately 0.7–0.8 and skill scores close to zero, whereas MDN exhibits more pronounced degradation. These results indicate that incorporating physics-guided constraints enhances model robustness and generalization, particularly under abrupt, previously unseen operating conditions.

4.3. Analysis of input contributions

Shapley values are computed for each test data point to quantify the contribution of input variables to predictive uncertainty, and these values are averaged daily. To facilitate comparison, the average Shapley values are expressed as percentages, indicating the relative contribution of each variable on each day for Reach A (Fig. 9), with overall contributions across the six-day period summarized in panel a and daily contributions in panels b–g.

The global results (Fig. 9a) indicate that variables related to water-level variation—namely, X1, X2, and X6—are the primary drivers of predictive uncertainty. Daily results (Fig. 9b–g) further show that a nonlinear term X6 ($H_d\Delta H_d$) plays a dominant role on most days, contributing between 19.6% and 53.6%.

Additionally, X1 (ΔH_u) and X2 (ΔH_d) generally exert greater influence on uncertainty than X3 (Q_u) and X4 (Q_d), particularly on Days 2, 5, and 6, where their contributions exceed 20% (Fig. 9c–f,g). However, this pattern varies across days. For example, on Day 3, only X6 exhibits a significant contribution, while the influence of other variables is minimal (Fig. 9d). This may help explain the reduced model performance observed on Day 3 (Fig. 4), where the model struggles to capture the trend of actual offtake discharge.

On Day 4, X3 (Q_u) becomes the dominant contributor, accounting for 46% of total importance, indicating that inflow discharge plays a key role in driving uncertainty during periods of rapid change. In contrast, boundary discharge variables (X3 and X4) have relatively minor contributions under stable conditions, typically below 12%. Meanwhile, X5 (H_d) shows a consistently small but highly variable contribution across all days, ranging from 0.5% to 10.5% (Fig. 9e).

Overall, these results suggest that predictive uncertainty is primarily driven by variables related to hydraulic change (X1 and X2) and the nonlinear downstream response (X6). However, under rapidly changing conditions, inflow discharge (X3) becomes a more dominant factor influencing uncertainty.

The SHAP results indicate that dominant drivers of predictive uncertainty vary with operating conditions. Specifically, X1 and X2 exceed 20% on Days 2, 5, and 6, whereas X3 accounts for 46% of total importance on Day 4, enabling operators to adjust their monitoring focus accordingly. During periods of pronounced variability in offtake discharge, inflow discharge becomes a key variable for operational monitoring and decision support. Under relatively stable conditions, variations in downstream water levels provide more relevant information for routine system management.

5. Discussion

5.1. Sensitivity and robustness analysis

To evaluate PgMDN's sensitivity to the weights of the loss terms, we conducted a comparative analysis using the nine-day

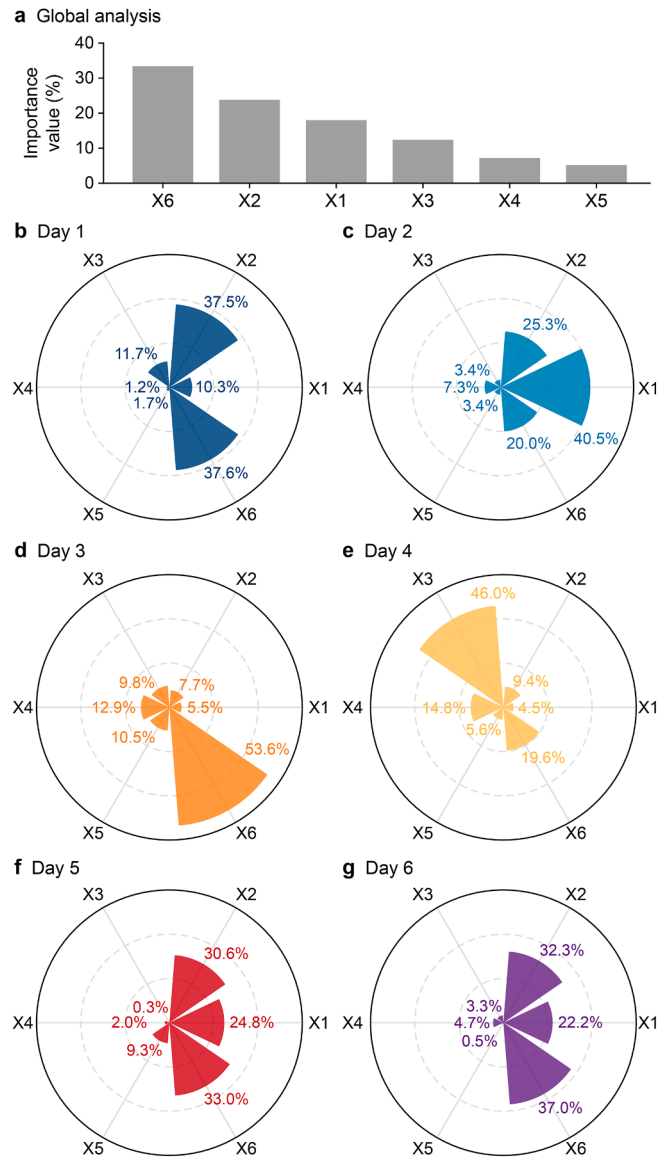


Fig. 9. Global and day-specific importance of input variables over the 6-day prediction period. a, Global importance ranking based on all 432 samples. b, Day-specific importance distributions for days 1–6, based on 72 samples per day. Input variables are defined as X1, ΔH_u , upstream water-level variation; X2, ΔH_d , downstream water-level variation; X3, Q_u , upstream discharge; X4, Q_d , downstream discharge; X5, H_d , downstream water level; X6, $H_d\Delta H_d$, interaction term between downstream water level and downstream water-level variation.

dataset from Reach A. The data loss weight is fixed at $\lambda_d = 1$, while the two physics-related terms are varied around their default values. Five configurations of are considered: $(\lambda_d, \lambda_p, \lambda_c) = (1, 1, 1)$, $(1, 0.5, 1)$, $(1, 1, 0.5)$, $(1, 5, 1)$, and $(1, 1, 5)$. All other training settings are kept identical.

The performance of PgMDN under these configurations is presented (Fig. 10a). MAE and RMSE remain relatively stable under moderate perturbations of the physics-related weights, indicating that the model is not overly sensitive to their selection. However, excessively increasing the physical NLL term $(1, 5, 1)$ degrades both accuracy and uncertainty performance, suggesting that overly strong physical constraints may restrict model flexibility and limit the learning of data patterns. In contrast, increasing the weight of the physics-guided term $(1, 1, 5)$ improves both accuracy and reliability, underscoring its greater benefit. Overall, PgMDN

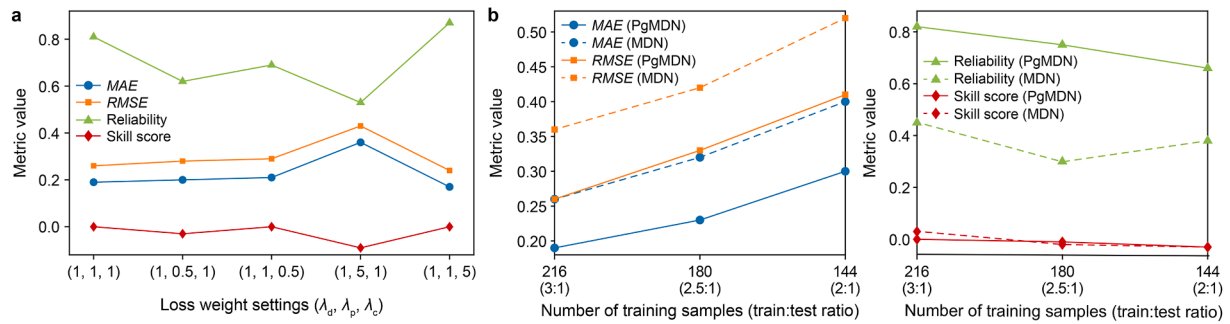


Fig. 10. Robustness analysis of PgMDN under different loss-weight settings and training sample sizes. **a**, PgMDN performance under different combinations of λ_d , λ_p , and λ_c . **b–g**, Comparison of MDN and PgMDN performance under different training sample sizes at Day 1 (**b**), Day 2 (**c**), Day 3 (**d**), Day 4 (**e**), Day 5 (**f**), and Day 6 (**g**). MDN, mixture density network; PgMDN, physics-guided mixture density network. MAE, mean absolute error; RMSE, root mean square error; λ_d , λ_p , and λ_c , the weighting coefficients of the data NLL term, the physical NLL term, and the physics-guided term, respectively; NLL, negative log-likelihood.

demonstrates robust performance and does not depend on a narrowly tuned weighting configuration.

To further assess robustness under data-scarce conditions, the forecasting horizon is fixed at 72 samples, while the amount of training data is progressively reduced. For each training size, MDN and PgMDN are trained and evaluated on the same test period under identical experimental settings.

Both models show degraded performance as the training data decrease, as reflected by increasing MAE and RMSE and declining reliability (Fig. 10b). However, PgMDN consistently achieves lower prediction errors and substantially higher reliability than MDN across all data sizes. Moreover, the degradation in PgMDN performance is more gradual, whereas MDN exhibits a sharper decline when training data becomes limited. The skill score of PgMDN remains close to zero under different data sizes, indicating stable probabilistic behavior, while MDN shows larger deviations under small-sample conditions.

5.2. Comparison with previous studies

As discussed in the Introduction, UQ in canal prediction has been approached through ensemble hydrodynamic models, stochastic operational frameworks, and data-driven probabilistic methods. Physics-based methods ensure physical consistency but often incur high computational costs, whereas purely data-driven models typically rely on scenario perturbations and may lack robustness under changing conditions.

In contrast, PgMDN models uncertainty by directly learning the conditional probability distribution of canal states, allowing it to capture complex, input-dependent uncertainty arising from unobserved operational disturbances without relying on predefined scenarios. Physical knowledge is incorporated into this probabilistic framework to constrain the learned distributions, ensuring consistency with fundamental hydraulic principles.

The superior accuracy and reliability of PgMDN predictions can be attributed to two primary factors. First, the MDN architecture with LSTM layers effectively captures implicit multimodal patterns in the data [38,39] (Fig. 3). Second, incorporating physical constraints enhances sensitivity to hydrodynamic laws, leading to more consistent predictions. This effect is reflected in the improved performance metrics (Figs. 6 and 9) and aligns with findings from previous work [41].

Furthermore, the SHAP analysis (Fig. 8) suggests that the model captures physically meaningful relationships. Uncertainty is primarily driven by water-level variations under normal conditions, whereas inflow variations dominate during periods of rapid operational change. This behavior is consistent with the

established understanding of canal system dynamics, supporting the interpretability of the model [52].

Compared to prior applications of MDN models [27,37,38], the proposed PgMDN demonstrates improved performance on small and real-world datasets by embedding physical constraints directly into the loss function. For variables such as offtake discharge, which are governed by short-term hydrodynamic processes rather than long-term hydrologic trends [38,39], PgMDN provides reliable and robust multimodal predictions.

This physics-guided framework enables the model to generate physically consistent, state-dependent conditional distributions while preserving the flexibility of deep learning. Such capability is particularly valuable for canal system operations, especially under short-horizon monitoring conditions. The resulting probabilistic forecasts of offtake discharge can be integrated into hydrodynamic models to estimate potential water level ranges under varying scenarios [8].

Accordingly, the PgMDN model is well-suited for real-time canal management, where predictive intervals can support adaptive scheduling and risk-aware control [53]. These probabilistic outputs enable more informed operational decisions, such as adjusting safety margins or optimizing the timing of control actions, rather than relying solely on deterministic forecasts. In this sense, PgMDN provides a practical framework for bridging physical hydrodynamic modeling, data-driven UQ, and operational decision support.

5.3. Limitations and future work

Despite its demonstrated performance, several limitations remain. First, the mass balance formulation relies on inflow and outflow discharges derived from a simplified 1D hydraulic model, which may not fully capture complex hydraulic interactions in real canal systems. As a result, local deviations from strict mass balance may occur under certain operating conditions, particularly when additional physical uncertainties are present.

In this study, such complexities are implicitly accounted for through real-time water level observations, which serve as key inputs for offtake discharge prediction. While this assumption is reasonable in practice, it remains an approximation. The multi-term loss formulation introduces an inherent trade-off between local physical consistency and data-driven flexibility. Because the physical constraint is applied locally, residual discrepancies may be absorbed by the data likelihood term. Further improvements could involve refining the hydraulic model or incorporating additional observational data.

Second, the PgMDN employs relatively soft physical constraints

compared to physics-informed neural networks [54,55]. While this enhances flexibility, it may reduce stability under certain input conditions. Incorporating stronger or more explicit physical constraints may further improve the model's ability to capture hydrodynamic behavior.

Additionally, although relatively long datasets are used in this study, the evaluation does not cover a full seasonal cycle. Future work could also extend the framework to multi-step forecasting, enabling better representation of long-term temporal dependencies and seasonal dynamics. Over longer horizons, changes in environmental and geophysical conditions may become increasingly important. Moreover, model validation may be influenced by measurement noise, simplified physical assumptions, and limited temporal coverage, which can introduce variability in performance metrics under certain conditions.

Finally, the current approach relies on a predefined set of hydraulic input features, which may not fully capture higher-order interactions within the system. Future research could explore end-to-end learning approaches, such as attention-based architectures, to implicitly encode physical relationships [56]. Such methods may also help improve the identification of high-risk conditions while maintaining interpretability and real-time applicability. Additionally, broader comparisons with alternative UQ methods remain an important direction for future work.

6. Conclusion

This study proposes a novel PgMDN model that integrates physical constraints into a data-driven probabilistic framework to quantify uncertainty in water-offtake discharge predictions. By embedding both a local mass-balance constraint and a coupling between system-state variation and uncertainty into the loss function, the model produces more robust and physically consistent conditional distributions.

The baseline MDN is first validated using small-sample synthetic datasets. The proposed PgMDN model is then evaluated using real-world canal system data across multiple time scales and operating conditions. Compared to the standard MDN, PgMDN demonstrates consistently superior performance in both point prediction and uncertainty quantification. The main findings are as follows.

- (1) Compared to the standard MDN, which performs well on synthetic datasets, PgMDN achieves more accurate predictions on small real-world datasets, reducing MAE and RMSE by 28% and 26%, respectively.
- (2) PgMDN improves prediction reliability from 0.45 to 0.82 and provides more conservative yet well-balanced uncertainty estimates while maintaining stable generalization under transfer testing and reduced training data conditions.
- (3) SHAP analysis shows predictive uncertainty is primarily driven by water level variations under normal conditions and by inflow discharge during rapid changes, supporting the physical interpretability of the model.

In summary, this study demonstrates that incorporating physical knowledge into neural network architectures can significantly enhance their robustness, generalization capability, and interpretability. This approach provides a promising pathway for integrating data-driven modeling with physical understanding, enabling more resilient and adaptive management of water resources and environmental systems.

CRediT authorship contribution statement

Wangjiayi Liu: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Guanghua Guan:** Writing – review & editing, Supervision, Software, Funding acquisition, Conceptualization. **Xiaonan Chen:** Resources, Data curation. **Liangsheng Shi:** Validation, Project administration. **Guangtao Fu:** Writing – review & editing, Supervision, Conceptualization. **Dragan Savic:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is funded by the National Key Research and Development Program of China [Grant No. 2024YFC3211800] and the China Scholarship Council (CSC) [Grant No. 202406270118]. The authors would also like to acknowledge the Construction and Administration Bureau of the Middle-Route of the South-to-North Water Diversion Project of China for supporting the data collection.

Data availability

All data, models, or codes that support the findings of this study are available from the corresponding author upon reasonable request.

References

- [1] J. Dong, X. Chen, Y. Li, M. Gao, L. Wei, N. Tangdamrongsu, W. Crow, Inter-basin water transfer effectively compensates for regional unsustainable water use, *Water Resour. Res.* 59 (12) (2023), <https://doi.org/10.1029/2023wr035129> e2023WR035129.
- [2] L. Wang, F. He, Y. Zhao, J. Wang, M. Hao, P. Lu, Y. Jia, K. Liu, H. Deng, Quantitative assessment and analysis of the impact of inter-basin water transfer on regional water resource stress, *J. Hydrol.* 657 (2025), <https://doi.org/10.1016/j.jhydrol.2025.133090>.
- [3] Q. Jiang, J. Li, Y. Sun, J. Huang, R. Zou, W. Ma, H. Guo, Z. Wang, Y. Liu, Deep-reinforcement-learning-based water diversion strategy, *Environ. Sci. Ecotechnol.* 17 (2024) 100298, <https://doi.org/10.1016/j.ese.2023.100298>.
- [4] W. Li, A. Sankarasubramanian, R. Ranjithan, E. Brill, Improved regional water management utilizing climate forecasts: an interbasin transfer model with a risk management framework, *Water Resour. Res.* 50 (8) (2014) 6810–6827, <https://doi.org/10.1002/2013wr015248>.
- [5] M. Yassin, A. Asfaw, V. Speight, J. Shucksmith, Evaluation of data-driven and process-based real-time flow forecasting techniques for informing operation of surface water abstraction, *J. Water Resour. Plann. Manag.* 147 (7) (2021), [https://doi.org/10.1061/\(asce\)jwr.1943-5452.0001397](https://doi.org/10.1061/(asce)jwr.1943-5452.0001397).
- [6] M. Shahrokhnia, M. Javan, M. Shahrokhnia, Influence of lateral canal water depth on offtake and cross-regulator discharge, *Irrig. Drain.* 58 (5) (2008) 561–568, <https://doi.org/10.1002/ird.461>.
- [7] S. Wang, P. Zhong, F. Zhu, B. Xu, C. Xu, L. Yang, M. Ben, Multi-objective optimization operation of multiple water sources under inflow-water demand forecast dual uncertainties, *J. Hydrol.* 630 (2024), <https://doi.org/10.1016/j.jhydrol.2024.130679>.
- [8] M. Bessar, F. Ancil, P. Matte, Uncertainty propagation within a water level ensemble prediction system, *J. Hydrol.* 603 (2021) 127193, <https://doi.org/10.1016/j.jhydrol.2021.127193>.
- [9] C. Papadopoulos, H. Yeung, Uncertainty estimation and monte carlo simulation method, *Flow Meas. Instrum.* 12 (4) (2001) 291–298, [https://doi.org/10.1016/S0955-5986\(01\)00015-2](https://doi.org/10.1016/S0955-5986(01)00015-2).
- [10] J. Knighton, E. White, E. Lennon, R. Rajan, Development of probability distributions for urban hydrologic model parameters and a monte carlo analysis of model sensitivity, *Hydrol. Process.* 28 (19) (2013) 5131–5139, <https://doi.org/10.1002/hyp.10009>.
- [11] R. Moua, N. Lesparre, J. Girard, B. Belfort, F. Lehmann, A. Younes, Coupled

- hydrogeophysical inversion of an artificial infiltration experiment monitored with ground-penetrating radar: synthetic demonstration, *Hydrol. Earth Syst. Sci.* 27 (23) (2023) 4317–4334, <https://doi.org/10.5194/hess-27-4317-2023>.
- [12] J. Zhang, Y. Li, G. Huang, X. Chen, A. Bao, Assessment of parameter uncertainty in hydrological model using a Markov-Chain-Monte-Carlo-based multilevel-factorial-analysis method, *J. Hydrol.* 538 (2016) 471–486, <https://doi.org/10.1016/j.jhydrol.2016.04.044>.
 - [13] J. Vrugt, C. Braak, C. Diks, G. Schoups, Hydrologic data assimilation using particle markov chain monte carlo simulation: theory, concepts and applications, *Adv. Water Resour.* 51 (2013) 457–478, <https://doi.org/10.1016/j.advwatres.2012.04.002>.
 - [14] R. Camacho, J. Martin, W. McAnally, J. Díaz-Ramírez, H. Rodríguez, P. Sucsý, S. Zhang, A comparison of Bayesian methods for uncertainty analysis in hydraulic and hydrodynamic modeling, *J. Am. Water Resour. Assoc.* 51 (5) (2015) 1372–1393, <https://doi.org/10.1111/1752-1688.12319>.
 - [15] P. Dey, S. Roy, J. Bathi, A. Mishra, Evaluation of uncertainty in stream flow prediction using Monte Carlo simulation for watershed-scale hydrological modeling, *J. Hydrol. Eng.* 29 (1) (2024), <https://doi.org/10.1061/jhyeff.Heeng-6008>.
 - [16] H. Razmkhah, A. AkhoundAli, F. Radmanesh, Correlated parameters uncertainty propagation in a rainfall-runoff model, considering 2-Copula; Case Study: karoon III river basin, *Environ. Model. Assess.* 22 (5) (2017) 503–521, <https://doi.org/10.1007/s10666-017-9569-z>.
 - [17] E. Jeremiah, S. Sisson, L. Marshall, R. Mehrotra, A. Sharma, Bayesian calibration and uncertainty analysis of hydrological models: a comparison of adaptive Metropolis and sequential Monte Carlo samplers, *Water Resour. Res.* 47 (7) (2011) W07547, <https://doi.org/10.1029/2010wr010217>.
 - [18] M. Li, Q. Wang, J. Bennett, D. Robertson, Error reduction and representation in stages (ERRIS) in hydrological modelling for ensemble streamflow forecasting, *Hydrol. Earth Syst. Sci.* 20 (9) (2016) 3561–3579, <https://doi.org/10.5194/hess-20-3561-2016>.
 - [19] Q. Shao, D. Dutta, F. Karim, C. Petheram, A method for extending stage-discharge relationships using a hydrodynamic model and quantifying the associated uncertainty, *J. Hydrol.* 556 (2018) 154–172, <https://doi.org/10.1016/j.jhydrol.2017.11.012>.
 - [20] Y. Zhang, V. Singh, A. Byrd, Basin-scale statistical method for probable maximum precipitation with uncertainty analysis, *J. Hydrol. Eng.* 24 (2) (2019) 04018067, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001759](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001759).
 - [21] N. Van Steenbergen, J. Ronsyn, P. Willems, A non-parametric data-based approach for probabilistic flood forecasting in support of uncertainty communication, *Environ. Model. Software* 33 (2012) 92–105, <https://doi.org/10.1016/j.envsoft.2012.01.013>.
 - [22] D. Nott, L. Marshall, J. Brown, Generalized likelihood uncertainty estimation (GLUE) and approximate Bayesian computation: what's the connection? *Water Resour. Res.* 48 (12) (2012) W12602, <https://doi.org/10.1029/2011wr011128>.
 - [23] M. Faghih, M. Mirzaei, J. Adamowski, J. Lee, A. El-Shafie, Uncertainty estimation in flood inundation mapping: an application of non-parametric bootstrapping, *River Res. Appl.* 33 (4) (2017) 611–619, <https://doi.org/10.1002/rra.3108>.
 - [24] X. Zhang, F. Liang, R. Srinivasan, M. Liew, Estimating uncertainty of streamflow simulation using Bayesian neural networks, *Water Resour. Res.* 45 (2) (2009) W02403, <https://doi.org/10.1029/2008wr007030>.
 - [25] J. Quilty, J. Adamowski, M. Boucher, A stochastic data-driven ensemble forecasting framework for water resources: a case Study using ensemble members derived from a database of deterministic wavelet-based models, *Water Resour. Res.* 55 (1) (2019) 175–202, <https://doi.org/10.1029/2018wr023205>.
 - [26] D. Lu, G. Konapala, S. Painter, S. Kao, S. Gangrade, Streamflow simulation in data-scarce basins using Bayesian and physics-informed machine learning models, *J. Hydrometeorol.* 22 (2021) 1421–1438, <https://doi.org/10.1175/jhm-d-20-0082.1>.
 - [27] Y. Li, S. Yan, J. Gong, Quantifying uncertainty in soil moisture retrieval using a Bayesian neural network framework, *Comput. Electron. Agric.* 215 (2023) 108414, <https://doi.org/10.1016/j.compag.2023.108414>.
 - [28] D. Althoff, L. Rodrigues, H. Bazame, Uncertainty quantification for hydrological models based on neural networks: the dropout ensemble, *Stoch. Environ. Res. Risk Assess.* 35 (5) (2021) 1051–1067, <https://doi.org/10.1007/s00477-021-01980-8>.
 - [29] K. Fang, D. Kifer, K. Lawson, C. Shen, Evaluating the potential and challenges of an uncertainty quantification method for Long short-term Memory models for soil moisture predictions, *Water Resour. Res.* 56 (12) (2020), <https://doi.org/10.1029/2020wr028095> e2020WR028095.
 - [30] X. Fan, X. Zhang, X. Yu, Uncertainty quantification of a deep learning model for failure rate prediction of water distribution networks, *Reliab. Eng. Syst. Saf.* 236 (2023) 109088, <https://doi.org/10.1016/j.ress.2023.109088>.
 - [31] D. Li, L. Marshall, Z. Liang, A. Sharma, Y. Zhou, Bayesian LSTM with stochastic variational inference for estimating model uncertainty in process-based hydrological models, *Water Resour. Res.* 57 (9) (2021), <https://doi.org/10.1029/2021wr029772> e2021WR029772.
 - [32] J. Zhang, J. Yan, D. Infield, Y. Liu, F. Lien, Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model, *Appl. Energy* 241 (2019) 229–244, <https://doi.org/10.1016/j.apenergy.2019.03.044>.
 - [33] G. McLachlan, S. Lee, S. Rathnayake, Finite mixture models, *Annual Review of Statistics and Its Application* 6 (1) (2019) 355–378, <https://doi.org/10.1146/annurev-statistics-031017-100325>.
 - [34] C. Bishop, *Mixture Density Networks* (Technical Report), Aston University, 1994.
 - [35] C. Yooj, Park, A mixture-density-network based approach for finding rating curves: facing multi-modality and unbalanced data distribution, *KSCE J. Civ. Eng.* 14 (2) (2010) 243–250, <https://doi.org/10.1007/s12205-010-0243-0>.
 - [36] S. Han, H. Li, M. Li, J. Zhang, R. Guo, J. Ma, W. Zhao, Deep learning-based stochastic modelling and uncertainty analysis of fault networks, *Bull. Eng. Geol. Environ.* 81 (6) (2022) 242, <https://doi.org/10.1007/s10064-022-02735-7>.
 - [37] Y. Zhang, A. Ye, B. Analui, P. Nguyen, S. Sorooshian, K. Hsu, Y. Wang, Comparing quantile regression forest and mixture density long short-term memory models for probabilistic post-processing of satellite precipitation-driven streamflow simulations, *Hydrol. Earth Syst. Sci.* 27 (24) (2023) 4529–4550, <https://doi.org/10.5194/hess-27-4529-2023>.
 - [38] D. Li, L. Marshall, Y. Zhou, A. Sharma, L. Yang, Z. Liang, Y. Yao, Enhancing probabilistic hydrological predictions with mixture density Networks: accounting for heteroscedasticity and Non-Gaussianity, *J. Hydrol.* 641 (2024) 131737, <https://doi.org/10.1016/j.jhydrol.2024.131737>.
 - [39] D. Klotz, F. Kratzert, M. Gauch, A. Keefe Sampson, J. Brandstetter, G. Klambauer, S. Hochreiter, G. Nearing, Uncertainty estimation with deep learning for rainfall-runoff modeling, *Hydrol. Earth Syst. Sci.* 26 (6) (2022) 1673–1693, <https://doi.org/10.5194/hess-26-1673-2022>.
 - [40] J. ChenY. Liu, Probabilistic physics-guided machine learning for fatigue data analysis, *Expert Syst. Appl.* 168 (2021) 114316, <https://doi.org/10.1016/j.eswa.2020.114316>.
 - [41] J. Chen, Y. Yu, Y. Liu, Physics-guided mixture density networks for uncertainty quantification, *Reliab. Eng. Syst. Saf.* 228 (2022) 108823, <https://doi.org/10.1016/j.ress.2022.108823>.
 - [42] A. Clemmens, T. Kacerek, B. Grawitz, W. Schuurmans, Test cases for canal control algorithms, *J. Irrigat. Drain. Eng.* 124 (1) (1998) 23–30.
 - [43] Y. HassaniS, Hashemy Shahdany, implementing agricultural water pricing policy in irrigation districts without a market mechanism: comparing the conventional and automatic water distribution systems, *Comput. Electron. Agric.* 185 (2021) 106121, <https://doi.org/10.1016/j.compag.2021.106121>.
 - [44] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
 - [45] A. Sharma, D. Tarboton, U. Lall, Streamflow simulation: a nonparametric approach, *Water Resour. Res.* 33 (2) (1997) 291–308, <https://doi.org/10.1029/96wr02839>.
 - [46] R. Szymkiewicz, Numerical stability of implicit four-point scheme applied to inverse linear flow routing, *J. Hydrol.* 176 (1–4) (1996) 13–23, [https://doi.org/10.1016/0022-1694\(95\)02785-8](https://doi.org/10.1016/0022-1694(95)02785-8).
 - [47] S. LundbergS. Lee, A unified approach to interpreting model predictions, in: *31st Conference on Neural Information Processing Systems (NIPS 2017)*, 2017, Long Beach, CA, USA.
 - [48] S. Lundberg, G. Erion, H. Chen, A. DeGrave, J. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S. Lee, From local explanations to global understanding with explainable AI for trees, *Nat. Mach. Intell.* 2 (1) (2020) 56–67, <https://doi.org/10.1038/s42256-019-0138-9>.
 - [49] P. Pinson, H. Nielsen, J. Møller, H. Madsen, G. Kariniotakis, Non-parametric probabilistic forecasts of wind power: required properties and evaluation, *Wind Energy* 10 (6) (2007) 497–516, <https://doi.org/10.1002/we.230>.
 - [50] Z. Chang, Y. Zhang, W. Chen, Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform, *Energy* 187 (2019) 115804, <https://doi.org/10.1016/j.energy.2019.07.134>.
 - [51] D. Stathakis, How many hidden layers and nodes? *Int. J. Rem. Sens.* 30 (8) (2009) 2133–2147, <https://doi.org/10.1080/01431160802549278>.
 - [52] A. Janon, M. Nodet, C. Prieur, C. Prieur, Global sensitivity analysis for the boundary control of an open channel, *Math. Control, Signals, Syst.* 28 (1) (2015), <https://doi.org/10.1007/s00498-015-0151-4>.
 - [53] S. Ye, C. Wang, Y. Wang, X. Lei, X. Wang, G. Yang, Real-time model predictive control study of run-of-river hydropower plants with data-driven and physics-based coupled model, *J. Hydrol.* (2022) 128942, <https://doi.org/10.1016/j.jhydrol.2022.128942>.
 - [54] M. Raissi, P. Perdikaris, G. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.* 378 (2019) 686–707, <https://doi.org/10.1016/j.jcp.2018.10.045>.
 - [55] J. Niu, W. Xu, H. Qiu, S. Li, F. Dong, 1-D coupled surface flow and transport equations revisited via the physics-informed neural network approach, *J. Hydrol.* 625 (2023) 130048, <https://doi.org/10.1016/j.jhydrol.2023.130048>.
 - [56] M. Kordani, M. Nikoo, M. Fooladi, I. Ahmadianfar, R. Nazari, A. Gandomi, Improving long-term flood forecasting accuracy using ensemble deep learning models and an attention mechanism, *J. Hydrol. Eng.* 29 (6) (2024), <https://doi.org/10.1061/jhyeff.Heeng-6262>.