

## Article

# Multi-Step Ahead Probabilistic Forecasting of Daily Streamflow Using Bayesian Deep Learning: A Multiple Case Study

Fatemeh Ghobadi  and Doosun Kang \*

Department of Civil Engineering, Kyung Hee University, 1732 Deogyong-daero, Giheung-gu, Yongin-si 17104, Korea

\* Correspondence: doosunkang@khu.ac.kr; Tel.: +82-31-201-2513

**Abstract:** In recent decades, natural calamities such as drought and flood have caused widespread economic and social damage. Climate change and rapid urbanization contribute to the occurrence of natural disasters. In addition, their destructive impact has been altered, posing significant challenges to the efficiency, equity, and sustainability of water resources allocation and management. Uncertainty estimation in hydrology is essential for water resources management. By quantifying the associated uncertainty of reliable hydrological forecasting, an efficient water resources management plan is obtained. Moreover, reliable forecasting provides significant future information to assist risk assessment. Currently, the majority of hydrological forecasts utilize deterministic approaches. Nevertheless, deterministic forecasting models cannot account for the intrinsic uncertainty of forecasted values. Using the Bayesian deep learning approach, this study developed a probabilistic forecasting model that covers the pertinent subproblem of univariate time series models for multi-step ahead daily streamflow forecasting to quantify epistemic and aleatory uncertainty. The new model implements Bayesian sampling in the Long short-term memory (LSTM) neural network by using variational inference to approximate the posterior distribution. The proposed method is verified with three case studies in the USA and three forecasting horizons. LSTM as a point forecasting neural network model and three probabilistic forecasting models, such as LSTM-BNN, BNN, and LSTM with Monte Carlo (MC) dropout (LSTM-MC), were applied for comparison with the proposed model. The results show that the proposed Bayesian long short-term memory (BLSTM) outperforms the other models in terms of forecasting reliability, sharpness, and overall performance. The results reveal that all probabilistic forecasting models outperformed the deterministic model with a lower RMSE value. Furthermore, the uncertainty estimation results show that BLSTM can handle data with higher variation and peak, particularly for long-term multi-step ahead streamflow forecasting, compared to other models.

**Keywords:** Bayesian neural network; forecasting uncertainty; multi-step ahead forecasting; probabilistic streamflow forecasting; variational inference



**Citation:** Ghobadi, F.; Kang, D. Multi-Step Ahead Probabilistic Forecasting of Daily Streamflow Using Bayesian Deep Learning: A Multiple Case Study. *Water* **2022**, *14*, 3672. <https://doi.org/10.3390/w14223672>

Academic Editors: Fi-John Chang, Li-Chiu Chang and Jui-Fa Chen

Received: 17 October 2022

Accepted: 11 November 2022

Published: 14 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Sustainable water resource management is an essential requirement worldwide, and streamflow forecasting is an essential component of an effective water resource management plan [1]. Accurate streamflow forecasting plays a critical role in many decision-making scenarios related to water resource management such as flood/drought control and mitigation, reservoir management, hydropower generation, sediment transport, and irrigation management [2]. Owing to the complex and nonlinear characteristics associated with streamflow [3], forecasting is challenging. Sustainable water resource management plans are used to meet the requirements of people today and in the future. To support risk-aware decision-making in water resource management, current streamflow forecasting approaches should be improved to estimate forecasting uncertainties and leverage large volumes of data with complex dependencies [4].

Deep learning (DL), a sophisticated and mathematically complex evolution of machine learning (ML) algorithms, has recently received huge attention from researchers and has gradually become the most widely used forecasting approach in hydrology [5–9]. The advantage of DL is its flexibility to learn massive data and the ease of incorporating exogenous covariates [10]. The advantages of DL techniques over traditional ML algorithms for streamflow forecasting have been discussed in many studies [1,11]. However, DL has not been extensively explored in forecasting uncertainty.

Only a few studies have been conducted on probabilistic approaches to streamflow forecasting. Thus, the existing uncertainty was not addressed by most DL approaches. However, deterministic approaches may not be as efficient as probabilistic methods and exhibit suboptimal performance. In general, deterministic forecasting is widely used in hydrology as an input for various water resource management plans. The transition from deterministic to probabilistic forecasting methods with uncertainty quantification is strongly favored in academia and industry. The primary issue in streamflow forecasting is the complex uncertainty rooted in the stochastic characteristics of streamflow time series. Furthermore, probabilistic forecasting has emerged to overcome the shortcomings of conventional deterministic methods and to deal with uncertainty more effectively. The probabilistic approach has recently gained importance because it can extract more valuable information from historical data and quantify the uncertainty of the future by forming a probability distribution over possible outcomes. The probabilistic approach extends beyond single-point forecasting for each time step and can provide a band of likely forecasting intervals above and below the mean forecasted value. Existing deterministic methods report the mean of possible outcomes, and they are unable to reflect the inherent uncertainty that exists in the real world.

Despite the fact that hydrological prediction can be most helpful when given in probabilistic form [12], the use of probabilistic modeling is still a relatively new concept in the field of hydrology [13]. Moreover, the probabilistic approach is crucial to optimal decision-making that reveals the upper and lower bounds between which the uncertain actual future values may exist. Occasionally, decision-making requires more than single-point forecasting; this is where distribution would be beneficial. To make reliable forecasts and to conduct a comprehensive performance evaluation, a probabilistic approach should be considered in streamflow forecasting. Most existing streamflow forecasting methods focus on deterministic forecasting. The application of various machine learning algorithms in deterministic prediction has been investigated in many studies [14–19]. Limited research has been conducted on multistep-ahead streamflow predictions [1,20–23]. Even though considerable efforts have been made to improve the performance of streamflow forecasting models from short- to long-term [24,25] and from single- to multi-step ahead [9,26,27], they are still limited by uncertainties [28–31].

An effective way to perform probability forecasting in the field of hydrology is to apply the Bayesian approach due to the benefits of uncertainty representation, understanding generalization, and reliable prediction through the lens of probability theory. The Bayesian approach can be classified into four primary groups: Bayesian model averaging (BMA), Bayesian model updating (BMU), Bayesian networks (BN), and Bayesian neural networks (BNN) [32]. A BNN is a type of stochastic artificial neural network that uses a BMU for training and updating the probabilistic distributions of network parameters. Furthermore, BMU and BN have become prevalent, and they have been implemented in various fields such as computer vision, natural language processing, medical diagnostics, autonomous driving, and flood hazard analysis [32]. Han and Coulibaly [33] presented a comprehensive review of Bayesian approaches applied to flood forecasting from 1999 to 2015. The results reveal that probabilistic flood forecasting can reduce uncertainty and provide more accurate and reliable forecasting. Moreover, they mentioned that only a limited number of river basins have been studied from the Bayesian perspective to date. Furthermore, we should determine if the Bayesian approaches are suitable for different watersheds with different sizes and physical and climatic characteristics. Costa and Fernandes [34] developed a

Bayesian framework to estimate the extreme flood quantile from a rainfall-runoff model of a dam in California. Xu et al. [35] developed a real-time probabilistic channel flood forecasting model by combining a hydraulic model with the Bayesian approach in the upstream reaches of the Three Gorges Dam on the Yangtze River, China. A state-of-the-art review was provided by Huang et al. [36] to summarize the application of Bayesian inference in system identification and damage assessment for civil infrastructure. Goodarzi et al. [37] proposed a decision-making model using BN to predict heavy precipitation in the Kan Basin. Bayesian neural networks are yet to be applied to probabilistic streamflow forecasting, as aforementioned.

Recent studies on probabilistic prediction in the field of hydrology are summarized in Table 1. As shown in Table 1, a few researchers trained a deterministic model and used the obtained deterministic result to obtain a probabilistic forecasting result to estimate the uncertainty [38]. However, in a few studies, a deterministic layer has been coupled with a probabilistic layer to achieve forecasting uncertainty [39]. Conversely, a few studies have focused on developing a probabilistic model by introducing stochastic components into the network by giving the network either stochastic activation or weights [40–42].

**Table 1.** Overview of recent probabilistic prediction studies in the field of hydrology.

Field	Probabilistic Method	Base Models	Posterior Approximation *		Evaluation Metrics		Ref.
			VI	MCM	Deterministic	Probabilistic	
Streamflow	LSTM-HetGP	ANN, HetGP, GLM, LSTM	-	-	NSE, RMSE, MRE, MSLE	percentage of coverage (POC) and the average interval width (AIW)	[39]
Flood	LSTM	ARIMA	-	-	RMSE, MAE		[40]
Streamflow	LSTM with multiparameter ensemble and dropout ensemble	-	-	✓	PBIAS, MARE, RMSE, NSE, KGE	POC, average width (AW), average interval score (AIS)	[41]
Streamflow	Variational Bayesian Long Short-Term Memory network (VB-LSTM)	Bayesian model Averaging (BMA)	✓	-	MAE	CRPS	[42]
Runoff	XGBoost (XGB) and Gaussian process regression (GPR) with Bayesian optimization algorithm (BOA)	GBR, LGB, CNN, LSTM, ANN, SVR, QR, GPR—combined with GPR	-	-	RMSE, MAPE, R <sup>2</sup>	Coverage probability, Mean width percentage, Suitability metric, CRPS	[38]
Runoff	B-spline quantile regression model combined with kernel density estimation	QR, QRNN	-	-	RMSE, R <sup>2</sup> , Q <sub>r</sub>	PICP, PINAW, CRPS	[43]
Streamflow	Bayesian LSTM model	physics-based hydrologic model (Precipitation-Runoff Modeling System)	-	✓	NSE, RMSE-observations standard deviation ratio (RSR)		[44]

\* The tick mark (✓) denotes the application of Posterior Approximation.

The application of probabilistic DL showed superior performance in various fields, including residential net load forecasting [45,46], short-term scheduling in power markets [47], photovoltaic power [48], load forecasting for buildings [49], and electricity consumption [50]. This indicates the wide range of the potential applicability of the probabilistic DL approaches. Univariate streamflow forecasting using conventional data-driven models has been investigated in the previous studies [51–54]. To the best of the authors' knowledge, the application of BLSTM in multi-step ahead probabilistic prediction using a retrospective univariate time series has not been applied to streamflow prediction yet. To address the aforementioned research gaps, this study proposed a framework for transforming a deterministic model into a probabilistic model with improved performance.

This study developed a Bayesian deep neural network framework to characterize the prognostic uncertainties for probabilistic streamflow forecasting, which investigated both epistemic and aleatoric uncertainties. The motivation of the framework was to transform existing deterministic prediction models into their probabilistic counterparts for better performance in water resources management and decision-making, and to cover newly emerged challenges that humankind encountered primarily due to climate change.

The primary contributions of the study are as follows: For the first time, in streamflow prediction, we introduced the Bayesian LSTM network's application for multi-step ahead probabilistic forecasting in water resource management. Bayesian theory and LSTM networks were combined to generate probabilistic streamflow forecasts to capture both epistemic and aleatoric uncertainties. This is the first study to exploit Bayesian deep learning for streamflow prediction. Moreover, a comprehensive comparison with a series of state-of-the-art probabilistic prediction methods is conducted. The superior performance of the proposed scheme was demonstrated with respect to both the deterministic and probabilistic forecasting results. Moreover, to demonstrate the superiority of probabilistic forecasting, particularly for water resource management, a comparative analysis was conducted for three case studies with different forecasting horizons and timescales.

The paper is organized as follows: In Section 2, the materials and methods are presented in subsections on Bayesian long-short-term memory (BLSTM) (2.1), experimental setup (2.2), and performance evaluation (2.3). In Section 3, the case study, study area (3.1), and experimental setup (3.2) are detailed. The results are presented in Section 4, with two subsections focusing on the probabilistic forecasting performance assessment (4.1) and the impact of the forecast horizon on probabilistic forecasting performance (4.2). Furthermore, the concluding remarks of this study with directions for future research are discussed in Section 5.

## 2. Materials and Methods

The proposed Bayesian deep-learning approach for probabilistic streamflow forecasting is presented in detail in the following sections.

### 2.1. Bayesian Long Short-Term Memory (BLSTM)

In this study, the Bayesian approach is employed, which is a well-established and thorough approach to fit probabilistic models that capture and distinguish different sources of uncertainties [55]. The BNN is a stochastic artificial neural network (ANN) trained using the Bayesian approach. Probability is defined in terms of the degree of belief in the Bayesian approach; the more likely an outcome is, the higher its degree of belief. The primary idea of the Bayesian approach in deep learning is to replace each weight with a distribution [56]. An LSTM network overcomes the long-term dependency issue of conventional RNNs through additional interactions in its various unit cells. Additionally, LSTM cells (memory cells) are composed of three gates (input, forget, and output) for short-term memory selection and a state vector transmission responsible for long-term memory. Information can be selectively passed during the learning procedure by manipulating the gate settings. The LSTM network is mathematically represented as follows [57]:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (1)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad (2)$$

$$h_t = o_t \times \tanh(C_t), \quad (3)$$

$$\tilde{C} = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \quad (4)$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tilde{C}, \quad (5)$$

$$\sigma(x) = \text{sigmoid}(x) = \frac{1}{1 + e^{-x}}, \quad (6)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \tag{7}$$

where at time step  $t$ ,  $x_t$  is the input vector,  $h_t$  is the LSTM output and hidden state (short-term memory), and  $i_t$ ,  $f_t$ , and  $o_t$  are the input, forget, and output gates, respectively.  $W$  and  $b$  are the weight matrix and bias, respectively.  $C_t$  is the current cell state (long-term memory), and  $\tilde{C}$  is the candidate cell state value.  $\sigma$  is a sigmoid activation function that uses  $h_{t-1}$  and  $x_t$  to make decisions regarding the input, forget, and output gates [57].

Given the input data,  $X_{train} = [x_1, \dots, x_{Train}]$  and their corresponding output labels  $Y_{train} = [y_1, \dots, y_{Train}]$ . The primary goal of the Bayesian approach is to identify the parameter  $W$  of a function  $y = f^W(x)$  that probably generates the desired output [58,59]. In this approach, a prior distribution that represents the prior belief about the neural network parameter distribution before observing the inputs is employed over  $W$  to capture epistemic uncertainty. The Bayesian neural network structure is illustrated in Figure 1. Rather than a single network, this method trains a set of networks in which the weight of each network is derived from a shared learning probability distribution [59].

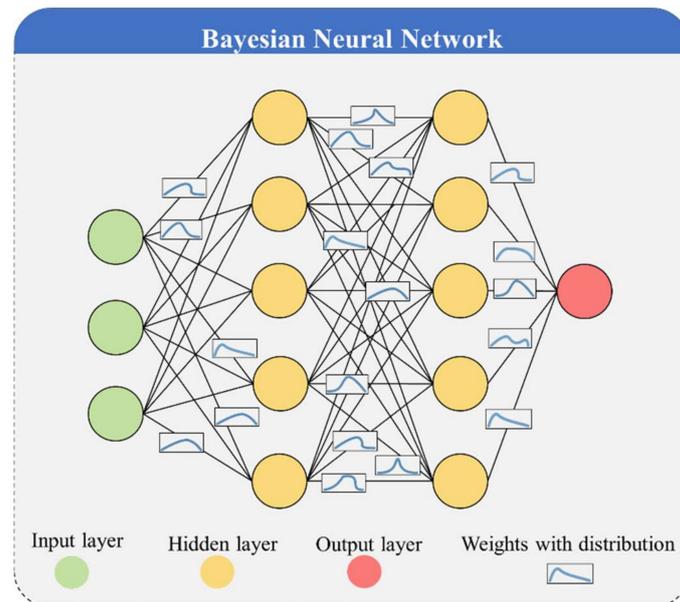


Figure 1. Structure of the Bayesian neural network.

Setting a standard normal distribution as a prior with zero mean, which can bring the benefit of regularization, has been demonstrated as one of the most effective solutions when the prior distribution is difficult to identify. After training the Bayesian deep neural network and observing data, the model likelihood distribution  $p(Y_{Train}|f^W)$  should be defined as a normal distribution  $\mathcal{N}(f^W(X_{Train}), \sigma^2)$  and observation noise  $\sigma$  to capture roughly suitable parameters. Based on the Bayesian rule, the posterior  $p(W|X_{Train}, Y_{Train})$  is employed over the weights to generate samples of predictions rather than the prior distribution. The posterior is calculated as follows [59]:

$$p(W|X_{Train}, Y_{Train}) = \frac{p(Y_{Train}|X_{Train}, W) \cdot p(W)}{p(Y_{Train}|X_{Train})} \tag{8}$$

where  $p(Y_{Train}|X_{Train})$  is the marginal likelihood probability that cannot be estimated, thereby, the posterior is not tractable without a variational inference to approximate it. With this distribution, suitable parameters given by the input data can be captured, and the output  $y$  can be predicted for a new input  $x$  by integration [58]:

$$p(y|x, X_{Train}, Y_{Train}) = \int p(y|x, W) p(W|X_{Train}, Y_{Train}) dW. \tag{9}$$

To evaluate the true posterior  $p(W|X_{Train}, Y_{Train})$ , an approximation variational distribution  $q_\theta(W)$ , which is parameterized by  $\theta$ , is required to ensure that the optimal distribution  $\tilde{q}_\theta(W)$  can represent the posterior by minimizing the Kullback–Leibler (KL) divergence [60] between the approximation variational and posterior distributions [61]:

$$KL(q_\theta(W) \parallel p(W|X_{Train}, Y_{Train})) = \int q_\theta(W) \log \frac{q_\theta(W)}{p(W|X_{Train}, Y_{Train})} dW. \tag{10}$$

Generally, two methods are available to approximate the posterior distribution: variational inference (VI) and Monte Carlo (MC) dropout [56,61,62]. The study employed the VI to solve the optimization issue analytically. Interested readers can refer to [62] for detailed information on the approximation method. The predictive distribution can be approximated by:

$$p(y|x, X_{Train}, Y_{Train}) = \int p(y|x, W)\tilde{q}_\theta(W)dW = \tilde{q}_\theta(y|x). \tag{11}$$

### 2.1.1. Epistemic Uncertainty

Mathematically, by simulating the model based on input  $x$ , the predictive mean can be estimated with an unbiased estimator, as follows [63]:

$$\tilde{\mathbb{E}}[y] := \frac{1}{T} \sum_{t=1}^T f^{\hat{W}_t}(x), \tag{12}$$

where  $\tilde{\mathbb{E}}[y]$  is the predictive mean,  $f^{\hat{W}_t}$  is the stochastic output of the prediction model,  $\hat{W}_t$  represents the sample weights, and  $T$  denotes the number of samples at time  $t$ . Similar to the estimation of the predictive mean, given that  $\hat{W}_t \sim \tilde{q}_\theta(W)$  and  $p(y|f^W(x)) = \mathcal{N}(y; f^w(x), \sigma^2)$  for  $\sigma > 0$ , the predictive variance can be estimated by an unbiased estimator as follows [63]:

$$\tilde{\mathbb{E}}[y^T y] := \frac{1}{T} \sum_{t=1}^T f^{\hat{W}_t}(x)^T f^{\hat{W}_t}(x) + \sigma^2. \tag{13}$$

The term  $\sigma^2$  corresponds to inherent noise in the input data. Afterward, the epistemic uncertainty, which represents the uncertainty of the model about its prediction outputs, is captured by the predictive variance that can be approximated as [63]:

$$\widetilde{Var}[y] = \tilde{\mathbb{E}}[y^T y] - \tilde{\mathbb{E}}[y]^T \tilde{\mathbb{E}}[y]. \tag{14}$$

### 2.1.2. Aleatoric Uncertainty

Aleatoric uncertainty can be divided into homoscedastic and heteroscedastic uncertainties. To capture the aleatoric uncertainty, parameter  $\sigma$  should be tuned. For each input  $x$ , in the homoscedastic uncertainty, the observation noise  $\sigma$  is assumed to be constant. In contrast, heteroscedastic uncertainty assumes that observation noise varies with the input. Heteroscedastic models are data-dependent and can be expressed as:

$$\mathcal{L}(\theta) = \frac{1}{T_{train}} \sum_{i=1}^{T_{train}} \frac{1}{2\sigma(x_i)^2} \| y_i - f(x_i) \|^2 + \frac{1}{2} \log(\sigma(x_i)^2). \tag{15}$$

Because the maximum posterior is performed to find a single value for parameter  $\theta$ , this approach does not capture the epistemic uncertainty since it is a property of the model, not the input data.

### 2.1.3. Combining Aleatoric and Epistemic Uncertainty

Abdar et al. [64] explained that an effective way to combine both uncertainties in a single model is to transform the heteroscedastic model into a Bayesian model by placing

a distribution over its weight and bias parameters. Thus, both the predictive mean and variance were derived from the developed prediction model.

$$[\hat{y}, \hat{\sigma}^2] = f_M^W(X), \tag{16}$$

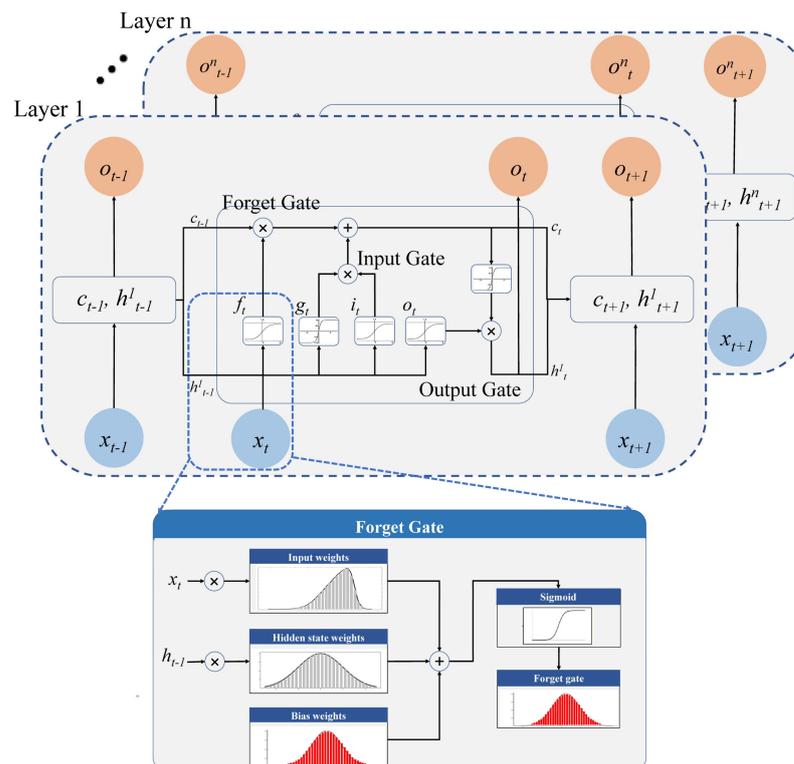
where  $f_M^W$  is the prediction model (BLSTM) used in this study, parameterized by the model weight  $\hat{W}$  [55,56]. The Gaussian likelihood is used to model the aleatoric uncertainty, and the final loss function of the prediction model can be expressed as [58]:

$$\mathcal{L}_M(\theta) = \frac{1}{T_{train}} \sum_{i=1}^{T_{train}} \frac{1}{2\sigma(x_i)^2} \|y_i - \hat{y}_i\|^2 + \frac{1}{2} \log \hat{\sigma}_i^2. \tag{17}$$

Finally, the predictive uncertainty of the prediction model, consisting of both aleatoric and epistemic uncertainties, can be approximated as

$$\widetilde{\text{Var}}[y] = \frac{1}{T_{sample}} \sum_{t=1}^{T_{sample}} \hat{y}_t^2 - \left( \frac{1}{T_{sample}} \sum_{t=1}^{T_{sample}} \hat{y}_t \right)^2 + \frac{1}{T_{sample}} \sum_{t=1}^{T_{sample}} \hat{\sigma}_t^2, \tag{18}$$

where  $T_{sample}$  denotes the number of training samples. An example of the Bayesian LSTM cell of the proposed BLSTM network is shown in Figure 2, with a zoomed-in plot of the forget gate at time step  $t$  in the first layer.



**Figure 2.** Example of the proposed BLSTM network with a zoomed-in plot of the forget gate at time step  $t$  in the first layer.

### 2.2. Performance Evaluation Metrics

To assess the performance of the prediction models, this study adopted the root mean square error (RMSE) metric for deterministic prediction and three metrics for probabilistic prediction: continuous ranked probability score (CRPS), prediction interval coverage probability (PICP), and mean prediction interval width (MPIW), which are formulated as follows.

### 1. Metric for Deterministic Forecasting

To evaluate the accuracy of the deterministic forecasting results, the root-mean-square error (RMSE) was selected as a commonly used hydrological evaluation indicator. The RMSE is defined as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}, \quad (19)$$

where  $\hat{Y}_i$  is the predicted variable,  $Y_i$  is the observed value, and  $n$  is the number of samples.

### 2. Metrics for Probabilistic Forecasting

A useful metric to assess the accuracy of probabilistic prediction models is the CRPS. The CRPS expresses the distance between the probabilistic forecast  $p$  and the observed value  $Y_i$  and is defined as

$$CRPS = \int_{-\infty}^{+\infty} [P(\hat{Y}_i) - H(\hat{Y}_i - Y_i)]^2 d\hat{Y}_i, \quad (20)$$

$$P(\hat{Y}_i) = \int_{-\infty}^{\hat{Y}_i} p(x) dx, \quad (21)$$

$$H(\hat{Y}_i - Y_i) = \begin{cases} 0 & \text{for } \hat{Y}_i < Y_i \\ 1 & \text{for } \hat{Y}_i \geq Y_i \end{cases}, \quad (22)$$

where  $p(x)$  is the probability density function (PDF),  $P(\hat{Y}_i)$  is the prediction cumulative distribution function (CDF), and  $H$  is the Heaviside step function, which equals 0 if  $\hat{Y}_i < Y_i$  and equals 1 otherwise.

The mean prediction interval width (MPIW) is an effective representation of sharpness in probabilistic predictions. This metric is defined as

$$MPIW = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i^u - \hat{Y}_i^l), \quad (23)$$

where  $n$  is the size of the test set, and  $\hat{Y}_i^u$  and  $\hat{Y}_i^l$  denote the upper and lower bounds of the 95% prediction interval, respectively.

Prediction interval coverage probability (PICP) or (PI) is the probability that the target lies within the interval provided by the prediction model. PICP is defined as:

$$PICP = \frac{1}{n} \sum_{i=1}^n c_i, \quad c_i = \begin{cases} 1, & \text{if } Y_i \in [\hat{Y}_i^l, \hat{Y}_i^u] \\ 0, & \text{if } Y_i \notin [\hat{Y}_i^l, \hat{Y}_i^u] \end{cases}. \quad (24)$$

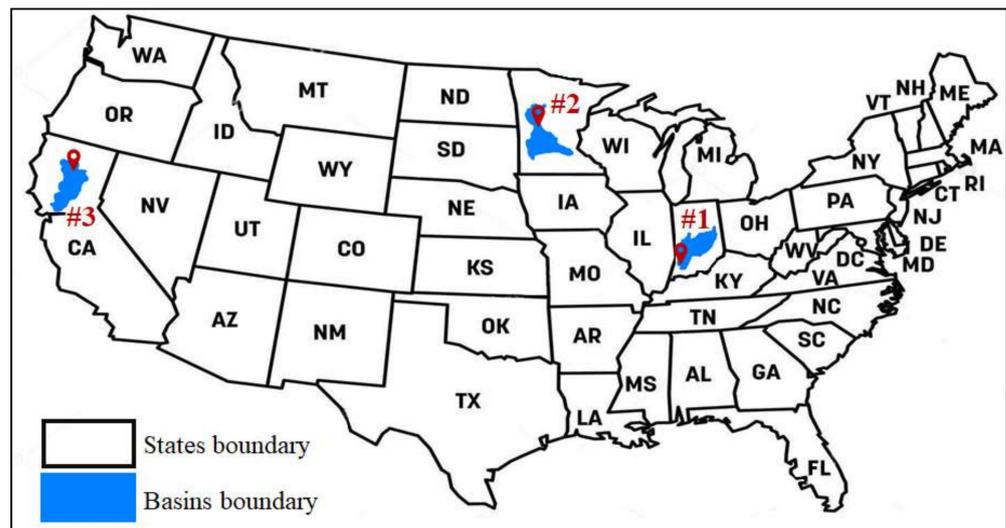
Thus, PICP indicates the frequency with which the prediction interval (PI) captures the observed value, ranging from 0 if all predicted values lie outside the PI and 1 if all predicted values lie inside the PI.

## 3. Case Study

To evaluate the performance of the probabilistic data-driven models under different conditions, three basins in the United States with different hydroclimatic conditions and drainage areas were selected as study areas, as described in the following section.

### 3.1. Study Area

The study basins were located in different climate regions of three states across the United States, i.e., IN (Indiana), MN (Minnesota), and CA (California). Figure 3 shows the locations of the three basins. The first case study was conducted in Bartholomew County, IN, the second was conducted in Koochiching County, MN, and third was conducted in Shasta County, CA. The drainage area of the river basins was approximately 1560–4420 km<sup>2</sup>.



**Figure 3.** Location of 3 study basins in different climate regions across the United States.

Based on the USGS statewide streamflow–water year 2021 report, the annual mean streamflow was ranked by state from 1 to 92, indicating the maximum and minimum annual flow for all years analyzed. Streamflow rankings were grouped into categories of much below normal, below normal, normal, above normal, and much above normal based on percentiles of flow (<10%, 10–24%, 25–75%, 76–90%, and >90%, respectively) [65]. Much-below-normal streamflow with a rank 84–91 is reported in CA and below-normal streamflow with a rank 70–83 is reported in MN. The annual mean streamflow rank for IN was reported to be normal, with a rank 24–50. Daily historical streamflow data for the three selected case studies were obtained from the United States Geological Survey (USGS) website, (<https://waterdata.usgs.gov/nwis>) (accessed on 1 February 2022).

The descriptive information of the daily streamflow in the three case studies is presented in Table 2. Details of the case studies, including gauge ID, gauge name, and drainage area, are presented in Table 3. For all catchments, streamflow with a 30-day lag was considered owing to the cross-correlation results.

**Table 2.** Descriptive information of daily historical streamflow data for three case studies.

Criteria	Case Study 1	Case Study 2	Case Study 3
No. Samples	27,146	34,205	22,645
Mean (m <sup>3</sup> /s)	58	30	31
Std (m <sup>3</sup> /s)	90	51	42
Min (m <sup>3</sup> /s)	2.5	0.6	3
25% (m <sup>3</sup> /s)	13	4	9
50% (m <sup>3</sup> /s)	31	11	20
75% (m <sup>3</sup> /s)	64	32	36
Max (m <sup>3</sup> /s)	1654	702	1274

**Table 3.** Details of the selected case studies.

Case Study. No.	Station ID	G-Name	Elev. (m)	Drainage Area (km <sup>2</sup> )	Lon. (°)	Lat. (°)	Period
1	03364000	EAST FORK WHITE RIVER AT COLUMBUS, IN	183.8	4421	85°55'32"	39°12'00"	1948–2022
2	05131500	LITTLE FORK RIVER AT LITTLEFORK, MN	330.3	4403	93°32'57"	48°23'45"	1928–2022
3	11368000	MC CLOUD R AB SHASTA LK CA	335.3	1564	122°13'07"	40°57'30"	1945–2007

### 3.2. Experiment Setup

Before using the data to train the model, data preprocessing began with min-max normalization and log transfer as the initial phase of model development. The input time

step was then derived from an autocorrelation analysis of the transformed-streamflow time series. Using a threshold of more than 0.5, which represents a moderate relationship, the past 30 days were selected as input. The autocorrelation analysis results for three case studies are given in the Supplementary File. The datasets for the three case studies were split into three sets: the first set accounted for 60% of the data, and it was used for model training; the second set was used for model validation (20%), and the remaining 20% was used for test purposes. Subsequently, the sliding window technique with a window size of 30 days was used. To demonstrate the superior performance of the Bayesian forecasting approach, probabilistic methods that have been widely used in the literature were employed for comparison. More specifically, LSTM-BNN [66], LSTM Monte-Carlo Dropout (LSTM-MC) [62,67], BNN [68], and deterministic LSTM were implemented in this study. Monte Carlo dropout is a straightforward epistemic uncertainty extension to the neural network. In general, dropout is a technique used to avoid overfitting by randomly dropping units during training. This can be considered the application of random noise in training. When this dropout was performed multiple times, multiple results were obtained. The distribution of the samples represents the uncertainty of the prediction model. The structure of the prediction models along with their graphical scheme are given in the Supplementary File.

The prediction models were developed in Python 3.6.9 with the Keras [69], TensorFlow [70], and PyTorch [71] libraries. The prediction model was implemented by an NVIDIA® GeForce® RTX 2070 SUPER and an Intel® Core i9-10920X central processing unit at 3.5 GHz utilizing 128 GB random access memory. For a fair comparison among the prediction models, a grid search for hyperparameter tuning was used to ensure identical evaluation.

#### 4. Result and Discussion

To better clarify the forecasting performance of BLSTM method, the study compares the BLSTM to the LSTM-BNN, BNN, and LSTM-MC in terms of prediction interval uncertainty, sharpness, prediction reliability, and multi-step ahead probabilistic prediction performance. Moreover, LSTM is used as a deterministic model to evaluate the performance of all probabilistic prediction models against the deterministic model. In this section, the predictive ability of the four probabilistic models for 1 day (Scenario I), 7 days (Scenario II), and 30 days (Scenario III) ahead of streamflow prediction is investigated.

##### 4.1. Probabilistic Prediction Performance Assessment

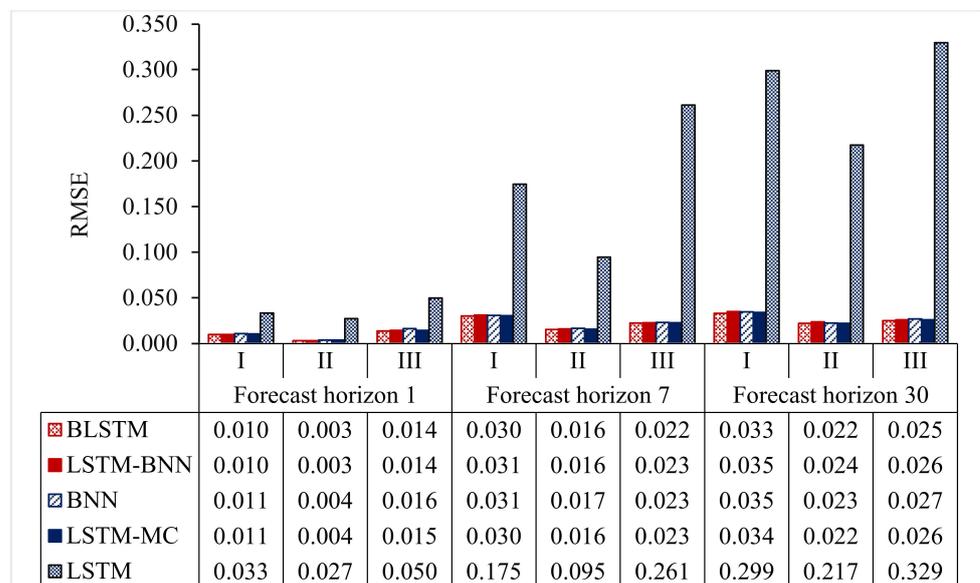
The PICP, MPIW, and CRPS values for the four models in the three case studies during the test period are listed in Table 4. The length of the bar represents the value of the evaluation metrics. In terms of PICP, the higher the value, the better and longer the bar, and vice versa for the other measures. Three major aspects must be considered simultaneously to evaluate probabilistic forecasting performance. PICP refers to the reliability of a model, MPIW refers to the model's sharpness, and CRPS indicates overall performance. In Scenario I, case study I was considered as an example because the models used the same mechanism to quantify the forecast uncertainty, and the PICP values of the four models were relatively close. Note that the larger the PICP and the smaller the MPIW and CRPS, the better the model performance. We observed that BLSTM showed better performance in handling datasets with high Std and peak streamflow. Case study III had 22,645 samples, which was ~17% and ~34% less than that of case studies I and II, with 27,146 and 34,205 samples, respectively. This difference did not lead to a particular change in the prediction performance of all the models for the first scenario. In this case study, from Scenarios I to III in BLSTM models, PICP decreased ~2% and 4%, respectively. While for case study I, PICP decreased ~25% and 50%, respectively, and for case study II, PICP decreased ~1% and 2%, respectively. Therefore, from the obtained results, we inferred that for single-step ahead prediction, the results were promising for all case studies, and the number of samples and peak streamflow did not affect the prediction performance. This

made BNNs extremely data-efficient because they could learn from even a small dataset without overfitting. Furthermore, we predict that more uncertainty is associated with the results, particularly for the case study with a higher peak of streamflow, leading to wider prediction intervals and lower coverage. As expected, all models exhibited better predictive performance during shorter lead times (1–7 days) than during the longer horizon (30 days). Therefore, from the obtained results, we can infer that the probabilistic forecasting model can lead to higher uncertainty and lower accuracy over a longer forecasting horizon.

**Table 4.** Summary of prediction performance results for three case studies and three scenarios.

Model	Metric	Forecast Horizon = 1			Forecast Horizon = 7			Forecast Horizon = 30		
		Case I	Case II	Case III	Case I	Case II	Case III	Case I	Case II	Case III
BLSTM	PICP	0.950	0.964	0.956	0.709	0.958	0.941	0.477	0.943	0.921
	MPIW	0.021	0.006	0.016	0.023	0.024	0.023	0.032	0.042	0.028
	CRPS	0.087	0.035	0.066	0.375	0.212	0.214	0.576	0.437	0.337
LSTM-BNN	PICP	0.957	0.967	0.971	0.591	0.962	0.956	0.371	0.953	0.942
	MPIW	0.023	0.008	0.018	0.035	0.037	0.032	0.039	0.076	0.040
	CRPS	0.086	0.034	0.070	0.400	0.226	0.237	0.615	0.457	0.367
BNN	PICP	0.955	0.953	0.961	0.496	0.779	0.870	0.281	0.630	0.865
	MPIW	0.027	0.009	0.022	0.045	0.050	0.040	0.053	0.141	0.047
	CRPS	0.101	0.041	0.083	0.412	0.240	0.262	0.634	0.461	0.371
LSTM-MC	PICP	0.973	0.994	0.981	0.454	0.948	0.913	0.268	0.945	0.909
	MPIW	0.046	0.040	0.027	0.049	0.050	0.032	0.057	0.076	0.039
	CRPS	0.109	0.071	0.106	0.414	0.251	0.248	0.636	0.545	0.376

To further evaluate the results of the four probabilistic models for all scenarios in the three case studies, LSTM as a well-known deterministic model was used to make a comparison in terms of RMSE, as shown in Figure 4. All probabilistic models in all horizons performed well and provided more accurate prediction performance in terms of RMSE than the deterministic LSTM, indicating the superiority of all probabilistic models in comparison with the conventional deterministic model. The range of RMSE indicated that all models were fairly trained, and they showed promising predictability performance in terms of RMSE.



**Figure 4.** Comparison among prediction models in terms of RMSE.

As shown in Figure 5, the probability that the prediction lies within the prediction interval by the LSTM-MC model is higher, followed by the LSTM-BNN, BNN, and BLSTM in the first scenario. The values of PICP indicate the percentage of the observed streamflow data lies within their 95% predictive intervals. In Scenario I, LSTM-MC outperformed BNN in terms of PICP. Moreover, for a longer horizon (7 days and 30 days ahead), due to the gradient vanishing of LSTM and BNNs as non-sequential models, both showed the lowest coverage and the points falling within the interval decreased by increasing the uncertainty in comparison with BLSTM and LSTM-BNN.

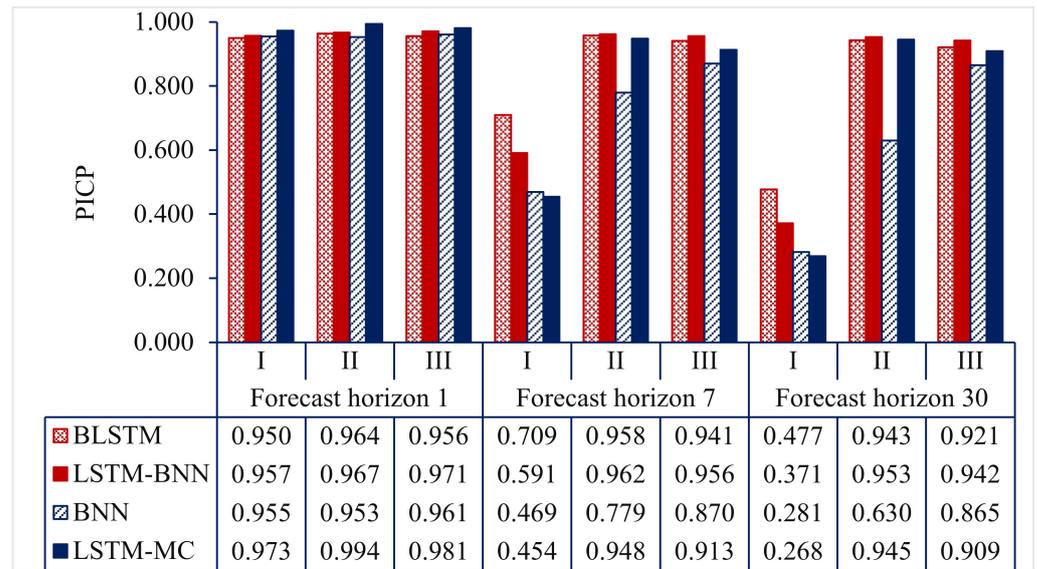


Figure 5. Comparison among prediction models in terms of PICP.

The MPIW values of the four models during the test period are shown in Figure 6a. The MPIW was considered an effective representation of sharpness in probabilistic predictions, and referred to the concentration of the predictive distributions. The more concentrated the predictive distributions, the lower the MPIW, the sharper the prediction, and consequently the better the predictive performance. As shown in Figure 6a, BLSTM has the lowest MPIW, which indicates that it is the sharpest predictive model among the other models in all scenarios. The stand-alone BNN and LSTM-MC were slightly different, whereas the LSTM-MC obtained the highest MPIW value among the other models, particularly by increasing the horizon. Compared with the BLSTM model with the narrowest MPIW, LSTM-MC had the worst prediction sharpness over all horizons. The fact that BLSTM presents the best predictive capability indicates the significance of capturing both epistemic and aleatoric uncertainties.

To comprehensively evaluate the probability prediction accuracy and reliability, a comparison among all prediction models in terms of the CRPS is shown in Figure 6b. Overall, BLSTM and LSTM-BNN competed with each other in all case studies and scenarios. However, in the longer horizon, BLSTM outperformed other models and proved its superiority by keeping more points falling within its forecasting interval while keeping the interval as narrow as possible while also increasing the uncertainty for a longer horizon. Therefore, the proposed BLSTM model outperformed the other models in terms of RMSE, MPIW, and CRPS, demonstrating the forecasting accuracy, sharpness, and overall performance of the model.

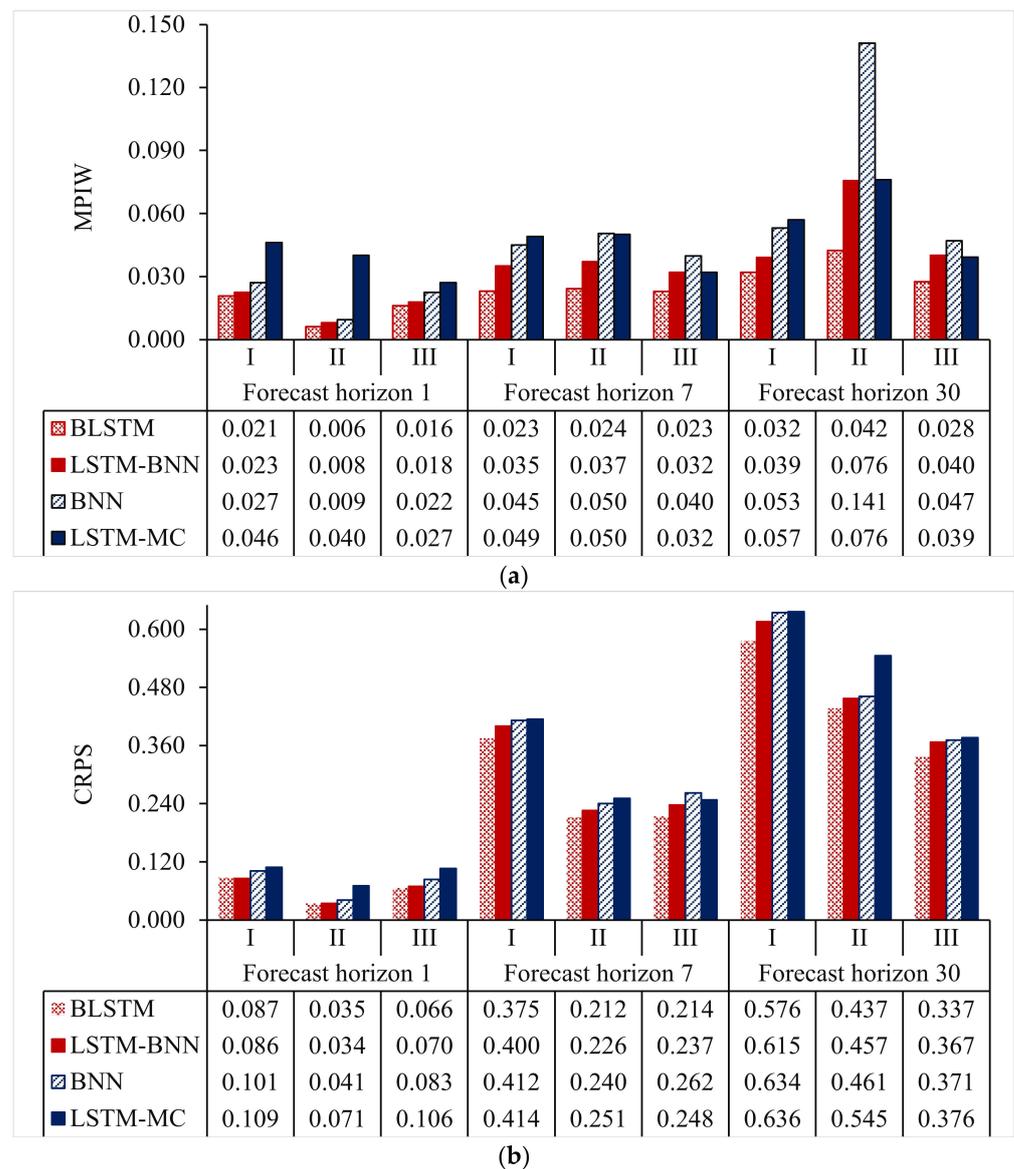
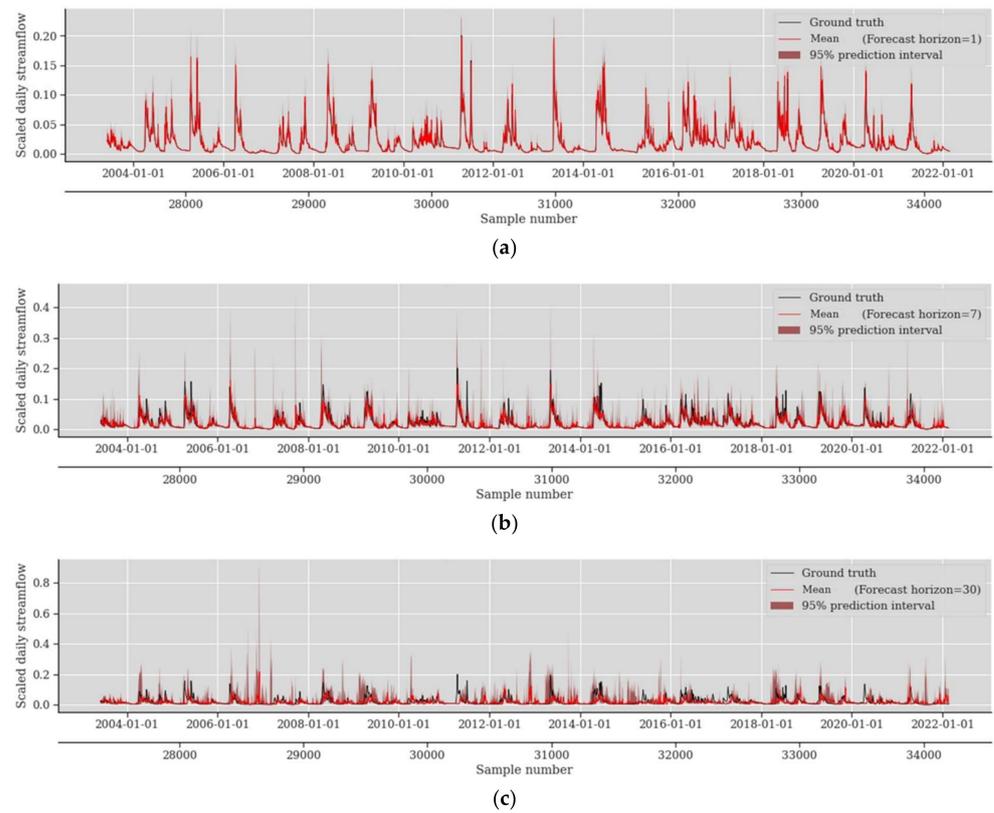


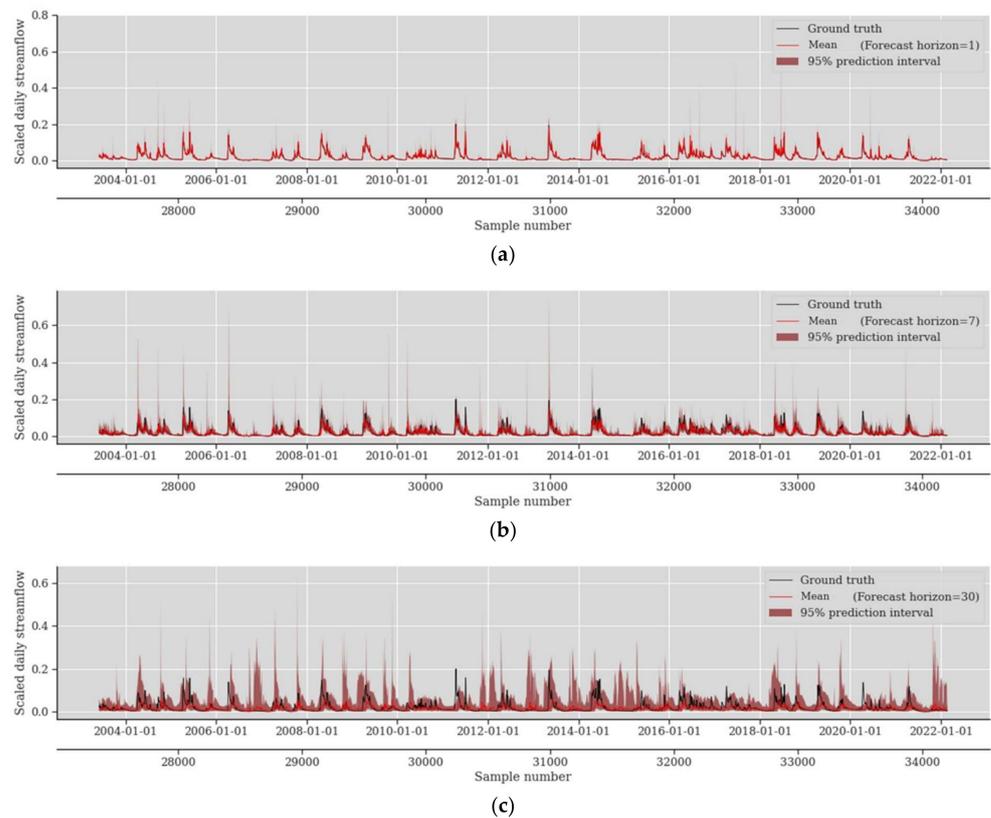
Figure 6. Comparison among prediction models in terms of (a) MPIW and (b) CRPS.

4.2. Impact of Forecast Horizon in Probabilistic Prediction Performance

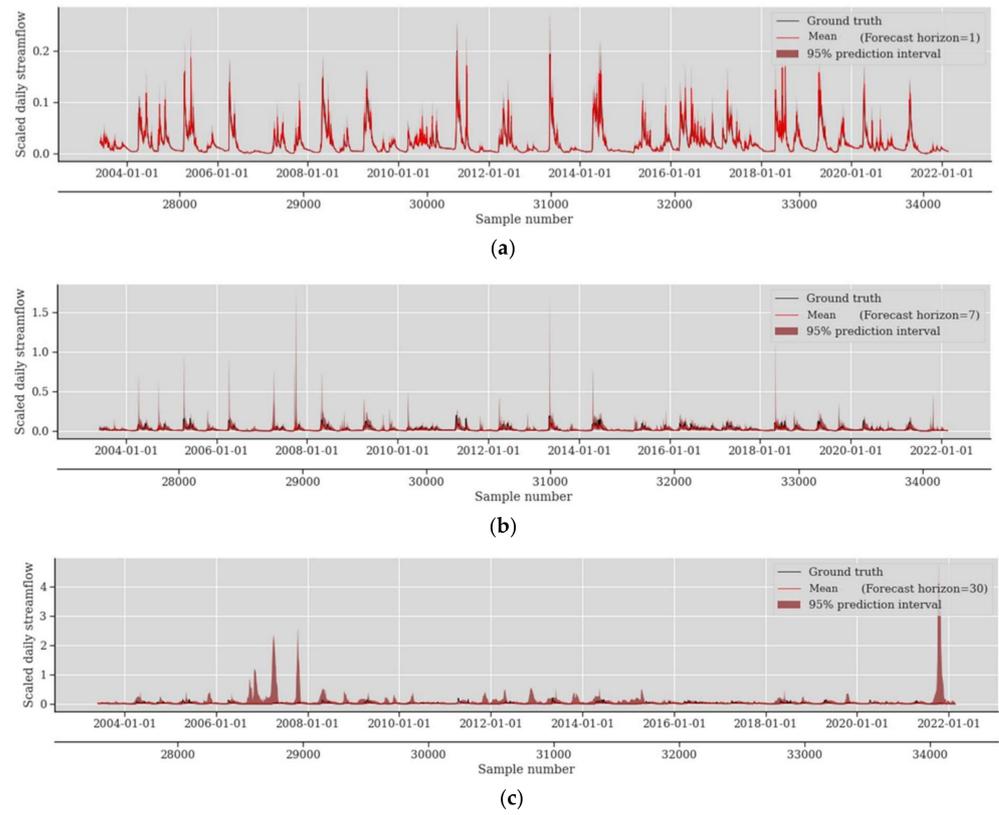
The prediction results of all models are compared graphically in Figures 7–10 in the form of time series for the entire test set for case study II for all scenarios (forecasting horizon). Considering space limitations, we have avoided adding the results of all case studies and scenarios in the main text, and only the results of case study II are presented. The actual streamflow and forecast value for the test period are represented by black and red curves, respectively. The red band represents the prediction interval, with a 95% confidence level. The probabilistic forecasts generated with the BLSTM model presented the benefits of high prediction coverage of observed streamflow data (PICP) with a tighter prediction width (MPIW) and better overall performance (CRPS), corresponding to reliability, sharpness, and resolution. Furthermore, accurate peak prediction, which is a crucial factor for disaster prevention and water resources management, can be predicted with reasonable magnitudes with the proposed BLSTM. Additionally, with increasing the forecast horizon, BLSTM still showed reliable performance, while other models were incapable of handling such a situation. In forecast horizon 30, massive fluctuations in the prediction results occurred for all the models and case studies. However, most of the prediction results were covered by the 95% interval in Scenario I, followed by Scenario II.



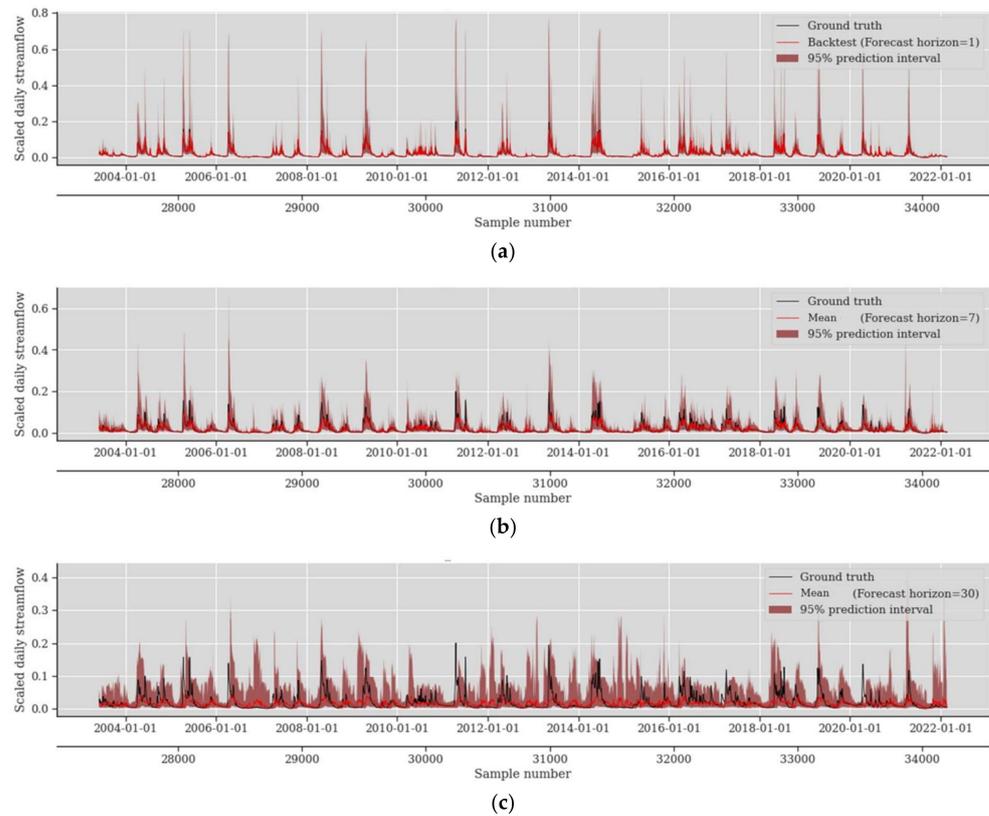
**Figure 7.** Probabilistic streamflow forecasting results obtained by BLSTM for case study II for (a) forecast horizon 1, (b) forecast horizon 7, and (c) forecast horizon 30.



**Figure 8.** Probabilistic streamflow forecasting results obtained by LSTM-BNN for case study II for (a) forecast horizon 1, (b) forecast horizon 7, and (c) forecast horizon 30.

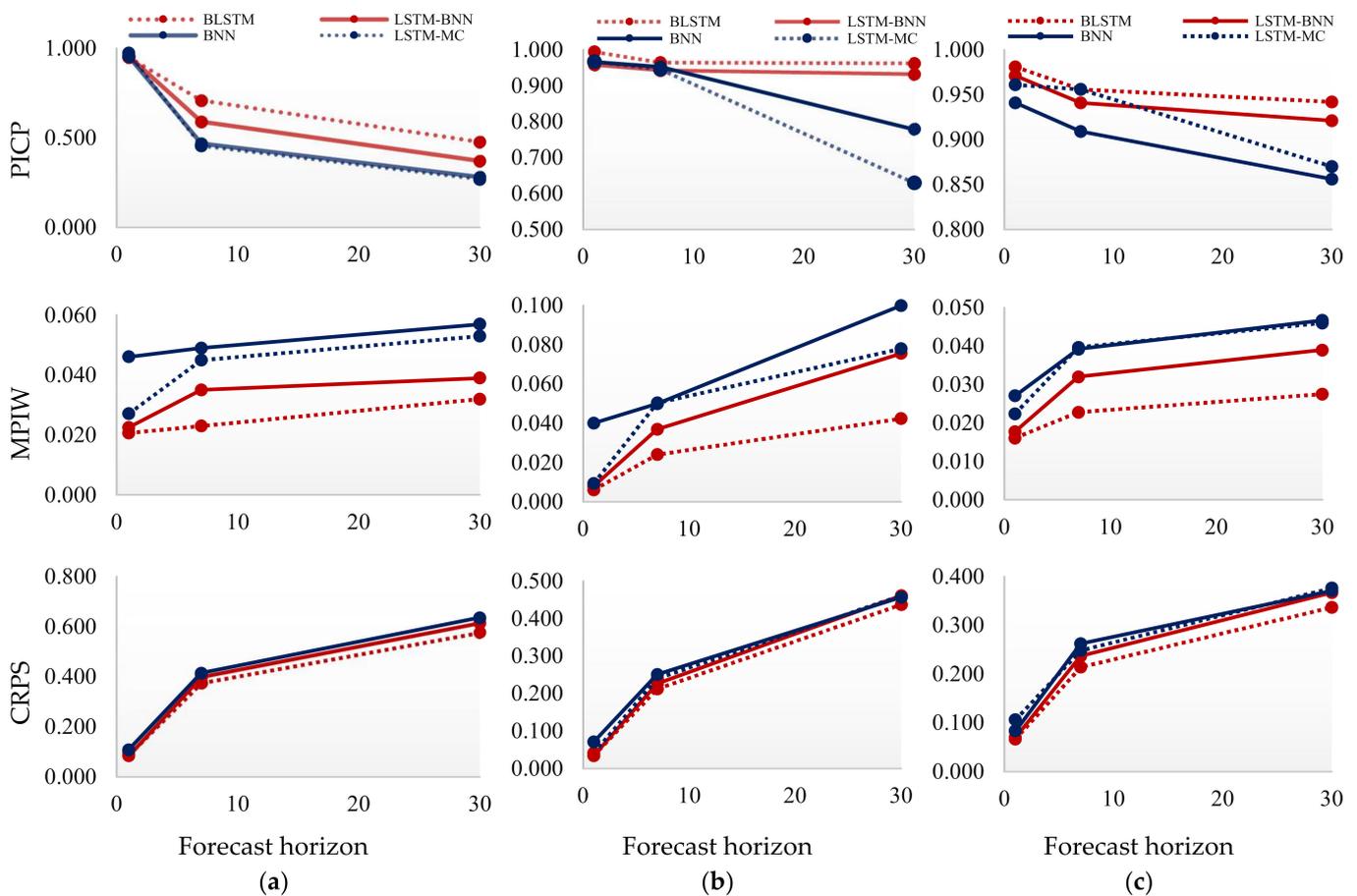


**Figure 9.** Probabilistic streamflow forecasting results obtained by BNN for case study II for (a) forecast horizon 1, (b) forecast horizon 7, and (c) forecast horizon 30.



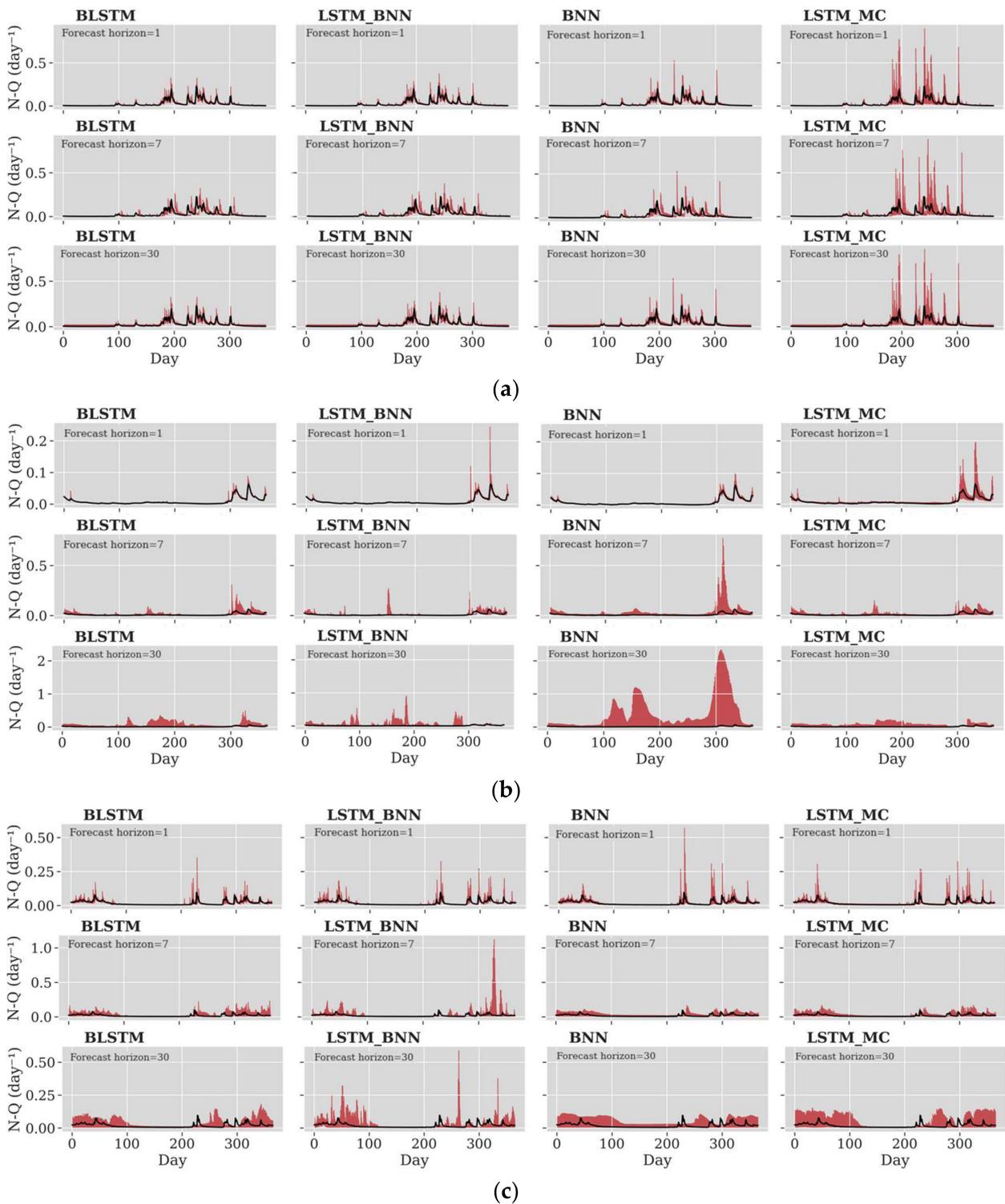
**Figure 10.** Probabilistic streamflow forecasting results obtained by LSTM-MC for case study II (a) forecast horizon 1, (b) forecast horizon 7, and (c) forecast horizon 30.

As shown in Figure 11, with an increase in the forecasting horizon from 1 to 30 days, the MPIW and CRPS values increase, and the PICP decreases. This indicates that the accuracy of the prediction models decreases with an increase in the forecasting horizon. The prediction accuracy over longer horizons decreased mainly as a result of the accumulative error issue in multi-step ahead recursive models and the gradient vanishing issue in long sequence time-series forecasting. Nevertheless, the predictive mean values of probabilistic streamflow from the BLSTM model matched the observations better than the other three models. The performance of all models gradually worsened with increasing lead times for the three case studies. As shown in Figure 11, when the overall prediction accuracy was low, MPIW was smaller. The interval width of the forecasting with LSTM-MC increased rapidly with the prediction horizon. The average interval width of the proposed BLSTM was much smaller than that of the other models. Simultaneously, the overall performance in terms of CRPS was higher, proving the superiority of the proposed method for the probability forecasting of daily streamflow, particularly for longer prediction horizons.



**Figure 11.** Change in probabilistic streamflow forecasting results by increasing horizon for (a) case study I, (b) case study II, and (c) case study III.

For a better and more vivid comparison of all model performances, the time series of all models for all case studies in the three scenarios are shown in Figure 12a–c for the first year of the test period (365 days). We observed that BNN and LSTM-MC underestimated the peak flows with a misleading trend in the first 365 days. In the first scenario, a 95% PI was relatively narrow and constant for all models, indicating that models captured both low and high flow values appropriately with low uncertainty. However, longer horizons in Scenarios II and III were associated with a wider 95% PI, indicating greater model uncertainty.

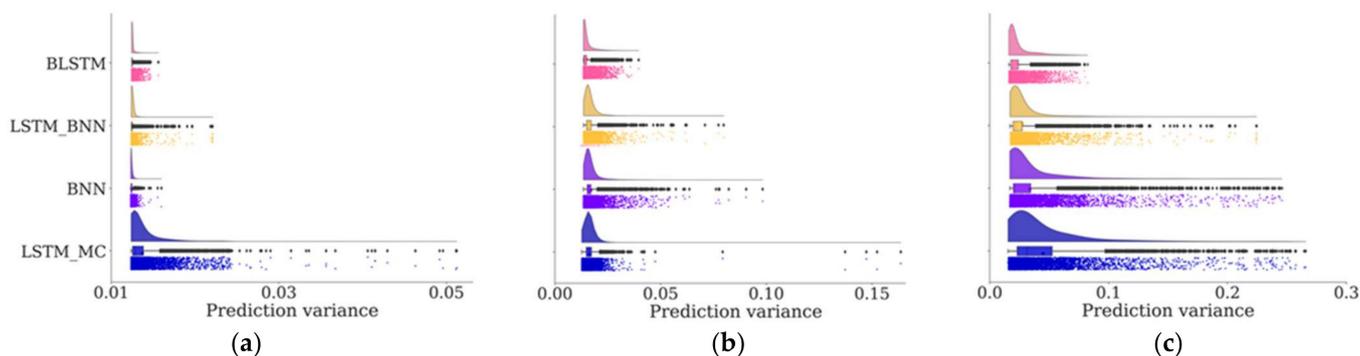


**Figure 12.** Prediction results of all models with 1, 7, and 30 days ahead forecasting for (a) case study I, (b) case study II, and (c) case study III.

Furthermore, we observed that all case studies can be effectively covered by the PI. Furthermore, for case study II, which had the lowest peak, BLSTM achieved the best results in all three scenarios. In contrast, case study I, with the highest peak at 1654 m<sup>3</sup>/s and Std. of 90 m<sup>3</sup>/s, achieved the worst prediction results for all scenarios. From Scenarios I to II in case study I for BLSTM, LSTM-BNN, BNN, and LSTM-MC, PICP decreased by approximately 25, 38, 48, and 53%, respectively, indicating the best performance of BLSTM

in maintaining its predictability in case study I, with the highest peak and Std in the extended horizon prediction. Moreover, by increasing the horizon, prediction performance for case study I dramatically decreased, whereas, in terms of PICP for BLSTM, case studies II and III from Scenarios I to III decreased by approximately 1–2% and 2–4%, respectively. Furthermore, LSTM-MC and BNN achieved the worst overall prediction performance for all the scenarios. The catchment area of case study I was relatively large, and heavy rain was the primary source of streamflow. These two characteristics cause the seasonal and annual variations in streamflow to be greater than those in the other two case studies. In this case study, the streamflow was very stable and small during the dry season, whereas in the rainy season, the streamflow increased steeply and then decreased. This made forecasting challenging and resulted in a higher uncertainty than that in the other case studies. Therefore, for this type of catchment, using more in-situ meteorological predictors, such as precipitation and temperature, along with available high-resolution large-scale hydroclimate data, can improve forecasting accuracy.

The kernel density estimation plots of the daily streamflow prediction of all models for case study II are displayed in Figure 13a–c. As depicted, the kernel density estimation plots are on the top with boxplots, and the data points of the prediction are underneath. The boxes represent the inner quartiles, the vertical lines within the box indicate the median, and the diamonds represent the outliers in each model. As shown in Figure 13, the prediction variance of BLSTM is lower in comparison with the other models, in particular for the Scenario III which is forecasting horizon 30. Moreover, the inter-quartile range of BLSTM is smaller which indicates that the BLSTM prediction results has less dispersion while LSTM-MC has the highest dispersion. The results of this study indicate that BLSTM shows the best overall probabilistic prediction performance.



**Figure 13.** Kernel density estimation plots of daily streamflow prediction of all models in case study II for (a) forecast horizon 1, (b) forecast horizon 7, and (c) forecast horizon 30.

## 5. Conclusions

This study proposes BLSTM as a probabilistic prediction model to estimate streamflow uncertainty. For comparison, three probabilistic models and one deterministic model, including LSTM-BNN, BNN, LSTM-MC, and LSTM, are developed under three scenarios: 1 day, 7 days, and 30 days ahead daily streamflow forecasting. The results are compared in terms of reliability, sharpness, and overall performance for three different case studies in the USA. Reliability is measured by computing the PICP, sharpness is measured by computing the MPIW, and overall performance is measured by CRPS. The results show that all probabilistic models outperformed the deterministic model (LSTM). Moreover, among the probabilistic models, BLSTM is superior. The Bayesian LSTM achieves better results with less computing time and is easier to train than those of LSTM-BNN and BNN. The results reveal that the BLSTM network with variational inference achieves the highest accuracy. The fact that BLSTM shows the best predictive performance indicates the significance of capturing temporal dependencies by considering both uncertainties. Moreover, taking advantage of the long- and short-term dependencies and capturing the inherent uncertainty that is inevitable in hydrology provides better prediction results. For

longer forecast horizons, models such as the BNN and LSTM-MC perform poorly due to the fact that the former is not an autoregressive model, and both have the gradient-vanishing problem in the long sequence time series. In addition, the issue of cumulative error in multi-step ahead recursive model is inevitable. Future research will investigate an enhanced network structure with a large prediction capacity, such as attention-based and parallel-feed architectures, to handle the long sequence time-series forecasting. In addition to the recursive models, other multi-step ahead prediction strategies, such as direct and hybrid, can be studied to minimize the accumulated error issue for longer horizons forecasting. Moreover, in the relevant future work, meteorological parameters such as precipitation, temperature, and humidity will be included as input to allow the model to detect the complexity necessary to enhance the accuracy of prediction, particularly for the longer horizon, and to evaluate the effect of multivariate input on model uncertainty.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/w14223672/s1>, Figure S1: Autocorrelation function plots of transformed-streamflow timeseries; Figure S2: The visualizations of network execution graphs and traces for BLSTM models' outputs; Figure S3: The visualizations of network execution graphs and traces for LSTM-BNN models' outputs; Figure S4: The visualizations of network execution graphs and traces for BNN models' outputs; Figure S5: The visualizations of network execution graphs and traces for LSTM-MC models' outputs; Figure S6: Prediction results of all models for case study II; Table S1: General structures of deep neural networks.

**Author Contributions:** F.G.: conceptualization, methodology, investigation, software, validation, formal analysis, data curation, writing—original draft, writing—review and editing, visualization. D.K.: supervision, validation, writing—review and editing, resources, and funding acquisition. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by (1) the Korea Ministry of Environment (MOE) as “Graduate School specialized in Climate Change” and (2) the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Korea Ministry of Trade, Industry & Energy (MOTIE) grant number [2022400000260].

**Data Availability Statement:** Data will be made available on request. Daily historical streamflow data for the three selected case studies were obtained from the United States Geological Survey (USGS) website, accessed on 1 February 2022 (<https://waterdata.usgs.gov/nwis>).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Ghobadi, F.; Kang, D. Improving Long-Term Streamflow Prediction in a Poorly Gauged Basin Using Geo-Spatiotemporal Mesoscale Data and Attention-Based Deep Learning: A Comparative Study. *J. Hydrol.* **2022**, *615*, 128608. [[CrossRef](#)]
2. Wang, Y.; Liu, J.; Li, R.; Suo, X.; Lu, E.H. Medium and Long-Term Precipitation Prediction Using Wavelet Decomposition-Prediction-Reconstruction Model. *Water Resour. Manag.* **2022**, *36*, 971–987. [[CrossRef](#)]
3. Dikshit, A.; Pradhan, B.; Santosh, M. Artificial Neural Networks in Drought Prediction in the 21st Century—A Scientometric Analysis. *Appl. Soft Comput.* **2022**, *114*, 108080. [[CrossRef](#)]
4. Van den Hurk, B.J.J.M.; Bouwer, L.M.; Buontempo, C.; Döschner, R.; Ercin, E.; Hananel, C.; Hunink, J.E.; Kjellström, E.; Klein, B.; Manez, M.; et al. Improving Predictions and Management of Hydrological Extremes through Climate Services: Wwww.Imprex.Eu. *Clim. Serv.* **2016**, *1*, 6–11. [[CrossRef](#)]
5. Lange, H.; Sippel, S. *Machine Learning Applications in Hydrology BT—Forest-Water Interactions*; Levia, D.F., Carlyle-Moses, D.E., Iida, S., Michalzik, B., Nanko, K., Tischer, A., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 233–257. ISBN 978-3-030-26086-6.
6. Lin, Y.; Wang, D.; Wang, G.; Qiu, J.; Long, K.; Du, Y.; Xie, H.; Wei, Z.; Shangguan, W.; Dai, Y. A Hybrid Deep Learning Algorithm and Its Application to Streamflow Prediction. *J. Hydrol.* **2021**, *601*, 126636. [[CrossRef](#)]
7. Hagen, J.S.; Leblais, E.; Lawrence, D.; Solomatine, D.; Sorteberg, A. Identifying Major Drivers of Daily Streamflow from Large-Scale Atmospheric Circulation with Machine Learning. *J. Hydrol.* **2021**, *596*, 126086. [[CrossRef](#)]
8. Ren, K.; Wang, X.; Shi, X.; Qu, J.; Fang, W. Examination and Comparison of Binary Metaheuristic Wrapper-Based Input Variable Selection for Local and Global Climate Information-Driven One-Step Monthly Streamflow Forecasting. *J. Hydrol.* **2021**, *597*, 126152. [[CrossRef](#)]

9. Tayerani Charmchi, A.S.; Ifaei, P.; Yoo, C.K. Smart Supply-Side Management of Optimal Hydro Reservoirs Using the Water/Energy Nexus Concept: A Hydropower Pinch Analysis. *Appl. Energy* **2021**, *281*, 116136. [[CrossRef](#)]
10. Papacharalampous, G.; Tyralis, H. A Review of Machine Learning Concepts and Methods for Addressing Challenges in Probabilistic Hydrological Post-Processing and Forecasting. *arXiv* **2022**. [[CrossRef](#)]
11. Ghimire, S.; Yaseen, Z.M.; Farooque, A.A.; Deo, R.C.; Zhang, J.; Tao, X. Streamflow Prediction Using an Integrated Methodology Based on Convolutional Neural Network and Long Short-Term Memory Networks. *Sci. Rep.* **2021**, *11*, 17497. [[CrossRef](#)]
12. Klotz, D.; Kratzert, F.; Gauch, M.; Keefe Sampson, A.; Brandstetter, J.; Klambauer, G.; Hochreiter, S.; Nearing, G. Uncertainty Estimation with Deep Learning for Rainfall-Runoff Modeling. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 1673–1693. [[CrossRef](#)]
13. Papacharalampous, G.; Tyralis, H.; Langousis, A.; Jayawardena, A.W.; Sivakumar, B.; Mamassis, N.; Montanari, A.; Koutsoyianis, D. Probabilistic Hydrological Post-Processing at Scale: Why and How to Apply Machine-Learning Quantile Regression Algorithms. *Water* **2019**, *11*, 2126. [[CrossRef](#)]
14. Adnan, R.M.; Liang, Z.; Parmar, K.S.; Soni, K.; Kisi, O. Modeling Monthly Streamflow in Mountainous Basin by MARS, GMDH-NN and DENFIS Using Hydroclimatic Data. *Neural Comput. Appl.* **2021**, *33*, 2853–2871. [[CrossRef](#)]
15. Apaydin, H.; Taghi Sattari, M.; Falsafian, K.; Prasad, R. Artificial Intelligence Modelling Integrated with Singular Spectral Analysis and Seasonal-Trend Decomposition Using Loess Approaches for Streamflow Predictions. *J. Hydrol.* **2021**, *600*, 126506. [[CrossRef](#)]
16. Mehdizadeh, S.; Fathian, F.; Safari, M.J.S.; Adamowski, J.F. Comparative Assessment of Time Series and Artificial Intelligence Models to Estimate Monthly Streamflow: A Local and External Data Analysis Approach. *J. Hydrol.* **2019**, *579*, 124225. [[CrossRef](#)]
17. Nanda, T.; Sahoo, B.; Chatterjee, C. Enhancing Real-Time Streamflow Forecasts with Wavelet-Neural Network Based Error-Updating Schemes and ECMWF Meteorological Predictions in Variable Infiltration Capacity Model. *J. Hydrol.* **2019**, *575*, 890–910. [[CrossRef](#)]
18. Khosravi, K.; Golkarian, A.; Tiefenbacher, J.P. Using Optimized Deep Learning to Predict Daily Streamflow: A Comparison to Common Machine Learning Algorithms. *Water Resour. Manag.* **2022**, *36*, 699–716. [[CrossRef](#)]
19. Xu, W.; Chen, J.; Zhang, X.J. Scale Effects of the Monthly Streamflow Prediction Using a State-of-the-Art Deep Learning Model. *Water Resour. Manag.* **2022**, *36*, 3609–3625. [[CrossRef](#)]
20. Cheng, M.; Fang, F.; Kinouchi, T.; Navon, I.M.; Pain, C.C. Long Lead-Time Daily and Monthly Streamflow Forecasting Using Machine Learning Methods. *J. Hydrol.* **2020**, *590*, 125376. [[CrossRef](#)]
21. Cui, Z.; Zhou, Y.; Guo, S.; Wang, J.; Xu, C.Y. Effective Improvement of Multi-Step-Ahead Flood Forecasting Accuracy through Encoder-Decoder with an Exogenous Input Structure. *J. Hydrol.* **2022**, *609*, 127764. [[CrossRef](#)]
22. Yin, H.; Zhang, X.; Wang, F.; Zhang, Y.; Xia, R.; Jin, J. Rainfall-Runoff Modeling Using LSTM-Based Multi-State-Vector Sequence-to-Sequence Model. *J. Hydrol.* **2021**, *598*, 126378. [[CrossRef](#)]
23. Kao, I.F.; Zhou, Y.; Chang, L.C.; Chang, F.J. Exploring a Long Short-Term Memory Based Encoder-Decoder Framework for Multi-Step-Ahead Flood Forecasting. *J. Hydrol.* **2020**, *583*, 124631. [[CrossRef](#)]
24. Babaeian, E.; Paheding, S.; Siddique, N.; Devabhaktuni, V.K.; Tuller, M. Short- and Mid-Term Forecasts of Actual Evapotranspiration with Deep Learning. *J. Hydrol.* **2022**, *612*, 128078. [[CrossRef](#)]
25. Xiang, Z.; Yan, J.; Demir, I. A Rainfall-Runoff Model With LSTM-Based Sequence-to-Sequence Learning. *Water Resour. Res.* **2020**, *56*, e2019WR025326. [[CrossRef](#)]
26. Ferreira, L.B.; da Cunha, F.F. Multi-Step Ahead Forecasting of Daily Reference Evapotranspiration Using Deep Learning. *Comput. Electron. Agric.* **2020**, *178*, 105728. [[CrossRef](#)]
27. Granata, F.; Di Nunno, F.; de Marinis, G. Stacked Machine Learning Algorithms and Bidirectional Long Short-Term Memory Networks for Multi-Step Ahead Streamflow Forecasting: A Comparative Study. *J. Hydrol.* **2022**, *613*, 128431. [[CrossRef](#)]
28. Masrur Ahmed, A.A.; Deo, R.C.; Feng, Q.; Ghahramani, A.; Raj, N.; Yin, Z.; Yang, L. Deep Learning Hybrid Model with Boruta-Random Forest Optimiser Algorithm for Streamflow Forecasting with Climate Mode Indices, Rainfall, and Periodicity. *J. Hydrol.* **2021**, *599*, 126350. [[CrossRef](#)]
29. Rahimzad, M.; Moghaddam Nia, A.; Zolfonoon, H.; Soltani, J.; Danandeh Mehr, A.; Kwon, H.H. Performance Comparison of an LSTM-Based Deep Learning Model versus Conventional Machine Learning Algorithms for Streamflow Forecasting. *Water Resour. Manag.* **2021**, *35*, 4167–4187. [[CrossRef](#)]
30. Barzegar, R.; Aalami, M.T.; Adamowski, J. Coupling a Hybrid CNN-LSTM Deep Learning Model with a Boundary Corrected Maximal Overlap Discrete Wavelet Transform for Multiscale Lake Water Level Forecasting. *J. Hydrol.* **2021**, *598*, 126196. [[CrossRef](#)]
31. Granata, F.; Di Nunno, F. Forecasting Evapotranspiration in Different Climates Using Ensembles of Recurrent Neural Networks. *Agric. Water Manag.* **2021**, *255*, 107040. [[CrossRef](#)]
32. Zheng, Y.; Xie, Y.; Long, X. A Comprehensive Review of Bayesian Statistics in Natural Hazards Engineering. *Nat. Hazards* **2021**, *108*, 63–91. [[CrossRef](#)]
33. Han, S.; Coulibaly, P. Bayesian Flood Forecasting Methods: A Review. *J. Hydrol.* **2017**, *551*, 340–351. [[CrossRef](#)]
34. Costa, V.; Fernandes, W. Bayesian Estimation of Extreme Flood Quantiles Using a Rainfall-Runoff Model and a Stochastic Daily Rainfall Generator. *J. Hydrol.* **2017**, *554*, 137–154. [[CrossRef](#)]
35. Xu, X.; Zhang, X.; Fang, H.; Lai, R.; Zhang, Y.; Huang, L.; Liu, X. A Real-Time Probabilistic Channel Flood-Forecasting Model Based on the Bayesian Particle Filter Approach. *Environ. Model. Softw.* **2017**, *88*, 151–167. [[CrossRef](#)]

36. Huang, Y.; Shao, C.; Wu, B.; Beck, J.L.; Li, H. State-of-the-Art Review on Bayesian Inference in Structural System Identification and Damage Assessment. *Adv. Struct. Eng.* **2019**, *22*, 1329–1351. [[CrossRef](#)]
37. Goodarzi, L.; Banihabib, M.E.; Roozbahani, A.; Dietrich, J. Bayesian Network Model for Flood Forecasting Based on Atmospheric Ensemble Forecasts. *Nat. Hazards Earth Syst. Sci.* **2019**, *19*, 2513–2524. [[CrossRef](#)]
38. Bai, H.; Li, G.; Liu, C.; Li, B.; Zhang, Z.; Qin, H. Hydrological Probabilistic Forecasting Based on Deep Learning and Bayesian Optimization Algorithm. *Hydrol. Res.* **2021**, *52*, 927–943. [[CrossRef](#)]
39. Zhu, S.; Luo, X.; Yuan, X.; Xu, Z. An Improved Long Short-Term Memory Network for Streamflow Forecasting in the Upper Yangtze River. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 1313–1329. [[CrossRef](#)]
40. Gude, V.; Corns, S.; Long, S. Flood Prediction and Uncertainty Estimation Using Deep Learning. *Water* **2020**, *12*, 884. [[CrossRef](#)]
41. Althoff, D.; Rodrigues, L.N.; Bazame, H.C. Uncertainty Quantification for Hydrological Models Based on Neural Networks: The Dropout Ensemble. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 1051–1067. [[CrossRef](#)]
42. Li, D.; Marshall, L.; Liang, Z.; Sharma, A. Hydrologic Multi-Model Ensemble Predictions Using Variational Bayesian Deep Learning. *J. Hydrol.* **2022**, *604*, 127221. [[CrossRef](#)]
43. He, Y.; Fan, H.; Lei, X.; Wan, J. A Runoff Probability Density Prediction Method Based on B-Spline Quantile Regression and Kernel Density Estimation. *Appl. Math. Model.* **2021**, *93*, 852–867. [[CrossRef](#)]
44. Lu, D.; Konapala, G.; Painter, S.L.; Kao, S.C.; Gangrade, S. Streamflow Simulation in Data-Scarce Basins Using Bayesian and Physics-Informed Machine Learning Models. *J. Hydrometeorol.* **2021**, *22*, 1421–1438. [[CrossRef](#)]
45. Sun, M.; Zhang, T.; Wang, Y.; Strbac, G.; Kang, C. Using Bayesian Deep Learning to Capture Uncertainty for Residential Net Load Forecasting. *IEEE Trans. Power Syst.* **2020**, *35*, 188–201. [[CrossRef](#)]
46. Wang, Y.; Gan, D.; Sun, M.; Zhang, N.; Lu, Z.; Kang, C. Probabilistic Individual Load Forecasting Using Pinball Loss Guided LSTM. *Appl. Energy* **2019**, *235*, 10–20. [[CrossRef](#)]
47. Toubeau, J.F.; Bottieau, J.; Vallee, F.; De Greve, Z. Deep Learning-Based Multivariate Probabilistic Forecasting for Short-Term Scheduling in Power Markets. *IEEE Trans. Power Syst.* **2019**, *34*, 1203–1215. [[CrossRef](#)]
48. Wang, H.; Yi, H.; Peng, J.; Wang, G.; Liu, Y.; Jiang, H.; Liu, W. Deterministic and Probabilistic Forecasting of Photovoltaic Power Based on Deep Convolutional Neural Network. *Energy Convers. Manag.* **2017**, *153*, 409–422. [[CrossRef](#)]
49. Xu, L.; Hu, M.; Fan, C. Probabilistic Electrical Load Forecasting for Buildings Using Bayesian Deep Neural Networks. *J. Build. Eng.* **2022**, *46*, 103853. [[CrossRef](#)]
50. Van der Meer, D.W.; Shepero, M.; Svensson, A.; Widén, J.; Munkhammar, J. Probabilistic Forecasting of Electricity Consumption, Photovoltaic Power Generation and Net Demand of an Individual Building Using Gaussian Processes. *Appl. Energy* **2018**, *213*, 195–207. [[CrossRef](#)]
51. Zhang, Z.; Zhang, Q.; Singh, V.P. Univariate Streamflow Forecasting Using Commonly Used Data-Driven Models: Literature Review and Case Study. *Hydrol. Sci. J.* **2018**, *63*, 1091–1111. [[CrossRef](#)]
52. Lange, H.; Sippel, S. Machine Learning Applications in Hydrology. In *Forest-Water Interactions; Ecological Studies*, Volume 240; Springer: Cham, Switzerland, 2020; pp. 233–257.
53. Abdul Kareem, B.; Zubaidi, S.L.; Ridha, H.M.; Al-Ansari, N.; Al-Bdairi, N.S.S. Applicability of ANN Model and CPSOCGSA Algorithm for Multi-Time Step Ahead River Streamflow Forecasting. *Hydrology* **2022**, *9*, 171. [[CrossRef](#)]
54. Wegayehu, E.B.; Muluneh, F.B. Short-Term Daily Univariate Streamflow Forecasting Using Deep Learning Models. *Adv. Meteorol.* **2022**, *2022*, 1860460. [[CrossRef](#)]
55. Kendall, A.; Gal, Y. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? *Adv. Neural Inf. Process. Syst.* **2017**, *2017*, 5575–5585. [[CrossRef](#)]
56. Blundell, C.; Cornebise, J.; Kavukcuoglu, K.; Wierstra, D. Weight Uncertainty in Neural Networks. *arXiv* **2015**. [[CrossRef](#)]
57. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
58. Li, G.; Yang, L.; Lee, C.G.; Wang, X.; Rong, M. A Bayesian Deep Learning RUL Framework Integrating Epistemic and Aleatoric Uncertainties. *IEEE Trans. Ind. Electron.* **2021**, *68*, 8829–8841. [[CrossRef](#)]
59. Bernardo, J.M.; Smith, A.F.M. *Bayesian Theory*; Wiley Blackwell: Hoboken, NJ, USA, 2008; ISBN 9780470316870.
60. Runnalls, A.R. Kullback-Leibler Approach to Gaussian Mixture Reduction. *IEEE Trans. Aerosp. Electron. Syst.* **2007**, *43*, 989–999. [[CrossRef](#)]
61. Jospin, L.V.; Buntine, W.; Boussaid, F.; Laga, H.; Bennamoun, M. Hands-on Bayesian Neural Networks—A Tutorial for Deep Learning Users. *IEEE Comput. Intell. Mag.* **2020**, *17*, 29–48. [[CrossRef](#)]
62. Gal, Y.; Ghahramani, Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *33rd Int. Conf. Mach. Learn. ICML* **2016**, *3*, 1651–1660. [[CrossRef](#)]
63. Gal, Y. Uncertainty in Deep Learning. Ph.D. Thesis, University of Cambridge, Cambridge, UK, 2016.
64. Abdar, M.; Pourpanah, F.; Hussain, S.; Rezazadegan, D.; Liu, L.; Ghavamzadeh, M.; Fieguth, P.; Cao, X.; Khosravi, A.; Acharya, U.R.; et al. A Review of Uncertainty Quantification in Deep Learning: Techniques, Applications and Challenges. *Inf. Fusion* **2021**, *76*, 243–297. [[CrossRef](#)]
65. Jian, X.; Wolock, D.M.; Lins, H.F.; Henderson, R.J.; Brady, S.J. *Streamflow—Water Year 2021: U.S. Geological Survey Fact Sheet 2022–3072*; USGS: Reston, VA, USA, 2022.

66. Chen, R.; Cao, J.; Zhang, D. Probabilistic Prediction of Photovoltaic Power Using Bayesian Neural Network-LSTM Model. In Proceedings of the 2021 IEEE 4th International Conference on Renewable Energy and Power Engineering (REPE), Beijing, China, 9–11 October 2021; pp. 294–299. [[CrossRef](#)]
67. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
68. Fortunato, M.; Blundell, C.; Vinyals, O. Bayesian Recurrent Neural Networks. *arXiv* **2019**. [[CrossRef](#)]
69. Ketkar, N. Introduction to Keras. In *Deep Learning with Python*; Ketkar, N., Ed.; Apress: Berkeley, CA, USA, 2017; pp. 97–111. ISBN 978-1-4842-2766-4.
70. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A System for Large-Scale Machine Learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
71. PyTorch Documentation—PyTorch 1.13 Documentation. Available online: <https://pytorch.org/docs/stable/index.html> (accessed on 7 November 2022).