Earth's Future

RESEARCH ARTICLE
10.1029/2022EF002958

Special Section:
Advancing flood characterization, modeling, and communication

A Bayesian Hierarchical Model Combination Framework for Real-Time Daily Ensemble Streamflow Forecasting Across a Rainfed River Basin

Álvaro Ossandón[1,2], Balaji Rajagopalan[1,3], Amar Deep Tiwari[4], Thomas Thomas[5], and Vimal Mishra[4]

[1]Department of Civil, Environmental and Architectural Engineering, University of Colorado, Boulder, CO, USA, [2]Departamento de Obras Civiles, Universidad Técnica Federico Santa María, Valparaíso, Chile, [3]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA, [4]Civil Engineering, Indian Institute of Technology, Gandhinagar, India, [5]National Institute of Hydrology, Central India Hydrology Regional Centre, WALMI Campus, Gehun Kheda, India

Key Points:

- We develop a Bayesian model combination framework for real-time streamflow forecasting
- The framework can be easily implemented in other major rainfed basins worldwide
- The framework can provide reliable streamflow forecast for short lead times

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

Á. Ossandón,
alvaro.ossandon@colorado.edu

Citation:

Ossandón, Á., Rajagopalan, B., Tiwari, A. D., Thomas, T., & Mishra, V. (2022). A Bayesian hierarchical model combination framework for real-time daily ensemble streamflow forecasting across a rainfed river basin. *Earth's Future*, *10*, e2022EF002958. https://doi.org/10.1029/2022EF002958

Received 3 JUN 2022
Accepted 29 NOV 2022

Author Contributions:

Conceptualization: Álvaro Ossandón, Balaji Rajagopalan, Vimal Mishra
Data curation: Amar Deep Tiwari, Thomas Thomas
Formal analysis: Álvaro Ossandón
Funding acquisition: Álvaro Ossandón, Balaji Rajagopalan, Vimal Mishra
Investigation: Álvaro Ossandón, Balaji Rajagopalan, Amar Deep Tiwari

Abstract The frequent occurrence of floods during the rainy season is one of the threats in rainfed river basins, especially in river basins of India. This study implemented a Bayesian hierarchical model combination (BHMC) framework to generate skillful and reliable real-time daily ensemble streamflow forecast and peak flow and demonstrates its utility in the Narmada River basin in Central India for the peak monsoon season (July–August). The framework incorporates information from multiple sources (e.g., deterministic hydrological forecast, meteorological forecast, and observed data) as predictors. The forecasts were validated with a leave-1-year-out cross-validation using accuracy metrics such as BIAS and Pearson correlation coefficient (*R*) and probabilistic metrics such as continuous ranked probability skill score, probability integral transform (PIT) plots, and the average width of the 95% confidence intervals (AWCI) plots. The results show that the BHMC framework can increase the forecast skill by 40% and reduce absolute bias by at least 28% compared to the raw deterministic forecast from a physical model, the Variable Infiltration Capacity model. In addition, PIT and AWCI show that the framework can provide sharp and reliable streamflow forecast ensembles for short lead times (1–3-day lead time) and provide useful skills beyond up to 5-day lead time. These will be of immense help in emergency and disaster preparedness.

Plain Language Summary This study implemented a Bayesian hierarchical based probabilistic forecast modeling framework to generate skillful and reliable real-time daily ensemble streamflow forecast and peak flow at multiple locations on a river network. This framework was demonstrated in the Narmada River basin network in Central India for the peak monsoon season (July–August). The framework incorporates information from multiple sources (e.g., deterministic hydrological forecast, meteorological forecast, and observed data) as predictors. The results show that compared to the raw deterministic forecast, the Bayesian framework can increase the forecast skill by 40%, reduce absolute bias by at least 28%, and provide sharp and reliable streamflow forecast ensembles at short lead times (1–3-day lead time) and provide useful skills beyond up to 5-day lead time. These will be of immense help in emergency and disaster preparedness.

1. Introduction

Among natural hazards, riverine floods are one of the most devastating, leading every year to severe infrastructural damages and human life losses across the world (Marcolongo et al., 2022; Wallemacq & House, 2018). In addition, they could become more frequent in the future, caused by an increase in frequency and intensity of extreme precipitation events projected due to anthropogenic climate change (Ali & Mishra, 2018; Papalexiou & Montanari, 2019; Wasko & Sharma, 2017). Floods are a significant concern in monsoonal regions since most precipitation is concentrated just in a few months, such as in India, which receives over the 80% of the total annual rainfall during the summer monsoon (June–September) season. These rainfall events are caused by synoptic-scale cyclonic depressions (Hunt & Fletcher, 2019; Hunt et al., 2016; Nanditha et al., 2022), leading to several floods in different rainfed basins across the country during the monsoon season. One of these basins is the Narmada River basin (NRB), located in the Central India region, covering four states (Madhya Pradesh, Gujarat, Maharashtra, and Chhattisgarh). In this region, there are high population centers and several reservoirs that are not well managed. In addition to that, India's current operational flood forecasting system provides poor predictive skill

Methodology: Álvaro Ossandón, Balaji Rajagopalan, Vimal Mishra
Project Administration: Balaji Rajagopalan, Vimal Mishra
Resources: Thomas Thomas
Software: Álvaro Ossandón, Amar Deep Tiwari
Supervision: Balaji Rajagopalan
Validation: Álvaro Ossandón, Amar Deep Tiwari
Visualization: Álvaro Ossandón
Writing – original draft: Álvaro Ossandón
Writing – review & editing: Álvaro Ossandón, Balaji Rajagopalan, Amar Deep Tiwari, Thomas Thomas, Vimal Mishra

since it works based on statistical models with gauge-to-gauge discharge correlations and coaxial correlations that use discharge and rainfall for most locations (CWC, 2015). These antecedents point to a critical need for skillful daily streamflow forecasts in the basin that could help mitigate the flood effects.

Recently, Ossandón et al. (2021) proposed a Bayesian hierarchical network model (BHNM) for daily streamflow ensemble forecasting and implemented it for this basin using observed streamflow and precipitation as predictors. BHNM provided promising results, but it was implemented only for a 1-day lead time and based on historical data. Motivated by these encouraging results, Ossandón et al. (2022) adapted the framework and applied it to "postprocess" simulated streamflow from the Variable Infiltration Capacity (VIC) model, including observed precipitation in the NRB. Their results showed that the model achieved homoscedasticity for the simulated residuals and significantly improved the skill of postprocessed simulated streamflow, which surged the hopes for applying this framework to real-time forecasting. Tiwari et al. (2022) developed a daily deterministic forecast for the Narmada basin at different lead times using the VIC model and meteorological forecast as forcing to evaluate the role of a bias correction method such as Quantile Mapping (QM) in postprocessing streamflow forecast. They found that the VIC model did not capture the high events well at the majority of the locations in the basin, but QM could significantly improve the forecast at 1–3-day lead time in terms of bias reduction. However, some scholars have stated that QM cannot ensure the reliability and coherence of forecasts since it does not consider the correlation between raw forecasts and observations (Zhao et al., 2017).

Past studies have demonstrated that multimodel combinations—such as Bayesian Model Averaging, to name one of them—tend to outperform a single model approach by combining information from multiple sources (e.g., Georgakakos et al., 2004; Krishnamurti et al., 2000; Mendoza et al., 2014; Rajagopalan et al., 2002). All these approaches are based on combining the information from multiple models with appropriate weights for each. Motivated by these results and the need for skillful long-lead daily flow forecast, we propose to adapt the Bayesian hierarchical framework proposed by Ossandón et al. (2021, 2022) to develop a Bayesian hierarchical model combination (BHMC) framework that incorporates multiple sources of information and provides real-time daily ensemble streamflow forecasting for the peak monsoon season (July–August). We demonstrate it by its application to the NRB. In this study, we apply this framework in this basin to demonstrate that it can (a) enhance the skill of real-time forecasts from hydrological models at different lead times and (b) deal with the heteroscedasticity of their residuals.

2. Study Area and Methods

2.1. Study Area

This study was conducted in the NRB in Central India. The flow direction in the basin is from east to west. It is the largest river in the country that drains into the Arabian Sea in the West (Figure 1). The NRB has an area of 97,882 km^2 and receives an average annual rainfall of 1,120 mm, and most of it arrives during the summer monsoon season (June–September). The upper parts of the basin receive higher precipitation than the lower basin. Several reservoirs have been built in the NRB for irrigation during the nonmonsoon season (few others have been designed for flood control) since it corresponds to an essential water source for the populous states of Madhya Pradesh and Gujarat. In the basin, most flood events occur during the July–August period; thus, this study will focus on that period.

2.2. Data

2.2.1. Historical Data Sets

This study used 0.25° gridded daily precipitation, maximum and minimum temperature, and wind data sets for 1951–2018, developed by the India Meteorology Department (IMD). The daily gridded precipitation data set was obtained through interpolating observations from 6,995 rain gauge stations across India using an inverse distance weighting scheme (see Pai et al., 2014). The daily maximum and minimum temperature fields for the 0.25° horizontal resolution grid were generated by interpolating observations from 395 stations across India using the Synergraphic Mapping approach (Maurer et al., 2002). The 2.5° gridded daily wind speed was obtained from the NCEP–NCAR reanalysis data set (Kalnay et al., 1996; Kistler et al., 2001) and regridded to a 0.25° grid to maintain consistency with other variables. Besides, from 1978 to 2018, we obtained from the India Water

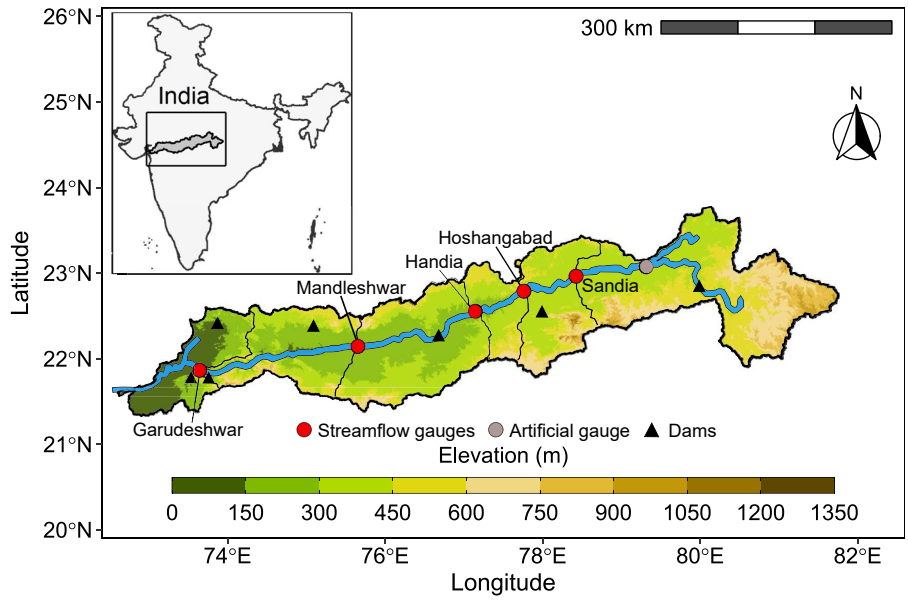

Figure 1. Map of the Narmada River basin (NRB), including five subbasin outlets (red circles), some of the major dams in the basin: Bargi, Tawa, Indirasagar, Jobat, and Sardar Sarovar (from upstream to downstream direction), and 150-m elevation bands. The gray circle represents an artificial gauge created for the Bayesian hierarchical model combination (BHMC).

Resource Information System observed daily streamflow and water storage at five gauges and reservoirs in the NRB (Figure 1).

2.2.2. Meteorological Forecast

We used a 0.25° gridded single trace meteorological forecast from the Global Forecast System (GFS), which was developed at the Indian Institute of Tropical Meteorology, as forcing for the hydrological forecast. The GFS provides precipitation, wind, and maximum and minimum temperature at a 3-hr temporal resolution. We used the meteorological forecast from the GFS for the monsoon season for the 2000–2018 period. Different studies have reported good performance of the GFS for the Indian region during the monsoon season (Kumar et al., 2019; Mukhopadhyay et al., 2019; Sridevi et al., 2020).

2.3. Methods: Bayesian Hierarchical Model Combination Framework

Considering observed information and meteorological forecasts are available at different lead times, the general BHMC framework comprises two components. The first component corresponds to deterministic streamflow forecasts obtained from z hydrological model at different lead times. The second is a Bayesian hierarchical model (BHM) that combines this information with other predictors, such as meteorological forecasts and past observations, to obtain posterior distributions or ensembles of streamflow forecasts. Figure 2 shows a flow chart of the BHMC framework for real-time daily ensemble streamflow forecasting at different lead times. In the first component, for each forecast day, t, the calibrated hydrological model j is forced with observed meteorological data to simulate the initial hydrologic conditions for the day, t. With these initial conditions and meteorological forecasts, the z hydrological models generate streamflow forecasts for the following days. Then, in the second component, forecasted streamflows from z hydrological models and other suitable covariates are incorporated into a BHMC to generate predictive posterior distribution (i.e., ensembles) of streamflow at all the gauges on the river network.

While Figure 2 provides a general framework with multiple model inputs for combination, for this study in the NRB, we have a single hydrologic model, VIC, combined with observed and forecast meteorology. A brief description of these components of the daily ensemble streamflow forecasting system, their implementation, and the calibration approaches is provided below.

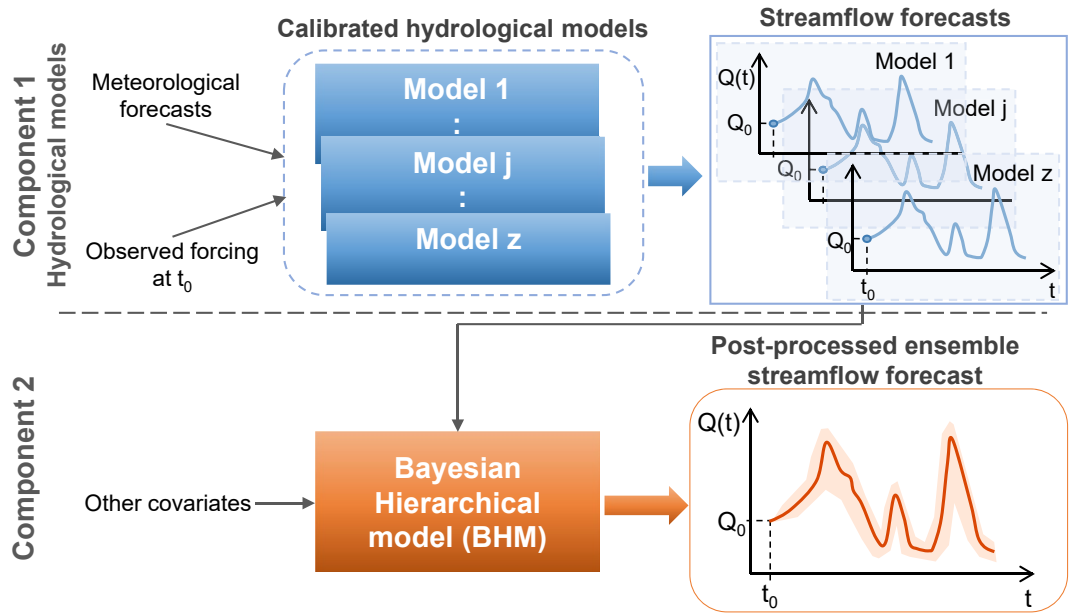

Figure 2. Flow chart of the general Bayesian hierarchical model combination (BHMC) framework for real-time daily ensemble streamflow forecasting at different lead times.

2.3.1. Hydrological Model: VIC Model

We used the VIC macroscale semidistributed hydrological model with three soil layers (Cherkauer et al., 2003; Liang & Xie, 2001; Liang et al., 1994, 1996) at a 0.25° horizontal resolution, considering the reservoir module developed by Haddeland et al. (2006) to simulate regulation effects on streamflow. The model requires gridded meteorological forcings (Section 2.2.1) and land surface characteristics for its implementation. Soil parameters were obtained from the Harmonized World Soil Database (FAO/IIASA/ISRIC/ISSCAS/JRC, 2012). The vegetation library and vegetation parameters were obtained from the Land Data Assimilation System (Rodell et al., 2004) and Advanced Very High-Resolution Radiometers (Hansen et al., 2000; Sheffield & Wood, 2007), respectively.

The VIC model was calibrated and validated manually against observed daily streamflow at four gauges (Garudeshwar, Mandleshwar, Handia, and Sandiya) in the NRB (Figure 1). Its performance was evaluated using the coefficient of determination (R^2) and the Nash-Sutcliffe efficiency (NSE; Nash & Sutcliffe, 1970). NSE is a deterministic metric that varies (∞, 1], with a perfect score of 1, and it compares how accurate and informative the hydrological model is with respect to climatology. The results showed that the VIC model simulated daily streamflow agrees well against the observed flow at all the selected locations in the Narmada basin during the calibration and evaluation periods since NSE and R^2 were higher than 0.6 (Table 1). However, the VIC model did not capture the high events well at the majority of the locations in the NRB. We developed daily meteorological forcing of 10 days forecast at the five gauges in the NRB (Figure 1) for the monsoon (June–September) season for the 2003–2018 period. We simulated initial hydrologic conditions forcing the VIC model with the observed precipitation and maximum and minimum temperatures from IMD for each forecast date during the monsoon season. Using these simulated initial hydrologic conditions and meteorological forecasts from GFS as forcings, the VIC model forecast was obtained for each forecast day during the monsoon season. Although the VIC streamflow forecasts were generated for lead times of up to 10 days for the entire monsoon season (June–September), this study only focuses on the 1–10-day forecast for the peak monsoon season (July–August). For further details about the VIC model implementation, calibration, and forecast streamflow generation, the reader is referred to Tiwari et al. (2022).

Table 1
Metrics for Calibration and Evaluation of the VIC Model Against the Observed Daily Streamflows at Four Gauge Stations in the Narmada River Basin

	Calibration			Validation		
Station	Period	NSE	R^2	Period	NSE	R^2
Garudeshwar	1973–1982	0.72	0.79	1984–1993	0.6	0.64
Mandleshwar	1973–1982	0.73	0.75	1987–1996	0.57	0.6
Handia	1979–1988	0.77	0.78	1994–2003	0.75	0.77
Sandiya	1990–1998	0.66	0.67	1999–2008	0.67	0.67

2.3.2. Bayesian Hierarchical Model Combination

The BHMC is implemented using the BHM proposed by Ossandón et al. (2022) with modifications to allow streamflow forecasts from z (in this study, $z = 1$) hydrological models at the n (in this study, $n = 5$) gauges of the NRB (Figure 1) and each lead time as predictors. Thus, in the data layer, for a k-day lead time, the daily streamflow at the gauge i and day t, $Q_t^{(i)}$, is assumed to follow a conditionally independent Gamma density function (Equation 1). In the first process layer, the density function parameters are expressed as a function of the expected value $\left(\mu_t^{(i)}\right)$ and variance $\left(\sigma_t^{(i)}\right)^2$ of $Q_t^{(i)}$ (Equation 2).

$$Gamma\left(Q_t^{(i)} \mid \alpha_t^{(i)}, \lambda_t^{(i)}, \boldsymbol{Q}_{f,k,t}^{(i)}, \boldsymbol{X}_{t-k}^{(i)}\right), \qquad i = 1,\ 2, \ldots,\ n \tag{1}$$

$$\alpha_t^{(i)} = \frac{\left(\mu_t^{(i)}\right)^2}{\left(\sigma_t^{(i)}\right)^2}, \quad \lambda_t^{(i)} = \frac{\mu_t^{(i)}}{\left(\sigma_t^{(i)}\right)^2} \tag{2}$$

Finally, the second process layer considers that $\mu_t^{(i)}$, and $\sigma_t^{(i)}$ depend on temporally varying covariates based on the following step functions to avoid fitting issues related to zero rainfall values, and to improve model fitting for high flow values:

$$\mu_t^{(i)} = \begin{cases} \beta_{0_1}^{(i)} + \left[\boldsymbol{\beta}_{Q_1}^{(i)}\right]^T g\left(\boldsymbol{Q}_{f,k,t}^{(i)}\right) & P_{t-k}^{(i)} = 0 \\ \beta_{0_2}^{(i)} + \left[\boldsymbol{\beta}_{Q_2}^{(i)}\right]^T g\left(\boldsymbol{Q}_{f,k,t}^{(i)}\right) + \left[\boldsymbol{\beta}_{X_1}^{(i)}\right]^T \boldsymbol{X}_{t-k}^{(i)} & P_{t-k}^{(i)} > 0 \ \wedge\ Q_{f,k,t,l}^{(i)} \le Q_{f,k,l,th}^{(i)} \\ \beta_{0_3}^{(i)} + \left[\boldsymbol{\beta}_{Q_3}^{(i)}\right]^T g\left(\boldsymbol{Q}_{f,k,t}^{(i)}\right) + \left[\boldsymbol{\beta}_{X_2}^{(i)}\right]^T \boldsymbol{X}_{t-k}^{(i)} & P_{t-k}^{(i)} > 0 \ \wedge\ Q_{f,k,t,l}^{(i)} > Q_{f,k,l,th}^{(i)} \end{cases} \tag{3}$$

$$\sigma_t^{(i)} = \begin{cases} \phi_{0_1}^{(i)} + \left[\boldsymbol{\phi}_{Q_1}^{(i)}\right]^T \boldsymbol{Q}_{f,k,t}^{(i)} & P_{t-k}^{(i)} = 0 \\ \phi_{0_2}^{(i)} + \left[\boldsymbol{\phi}_{Q_2}^{(i)}\right]^T \boldsymbol{Q}_{f,k,t}^{(i)} + \left[\boldsymbol{\phi}_{X_1}^{(i)}\right]^T \boldsymbol{X}_{t-k}^{(i)} & P_{t-k}^{(i)} > 0 \ \wedge\ Q_{f,k,t,l}^{(i)} \le Q_{f,k,l,th}^{(i)} \\ \phi_{0_3}^{(i)} + \left[\boldsymbol{\phi}_{Q_3}^{(i)}\right]^T \boldsymbol{Q}_{f,k,t}^{(i)} + \left[\boldsymbol{\phi}_{X_2}^{(i)}\right]^T \boldsymbol{X}_{t-k}^{(i)} & P_{t-k}^{(i)} > 0 \ \wedge\ Q_{f,k,t,l}^{(i)} > Q_{f,k,l,th}^{(i)} \end{cases} \tag{4}$$

where $\boldsymbol{\beta}^{(i)} = \left[\beta_{0_1}^{(i)}, \beta_{0_2}^{(i)}, \beta_{0_3}^{(i)}, \boldsymbol{\beta}_{Q_1}^{(i)}, \boldsymbol{\beta}_{Q_2}^{(i)}, \boldsymbol{\beta}_{Q_3}^{(i)}, \boldsymbol{\beta}_{X_1}^{(i)}, \boldsymbol{\beta}_{X_2}^{(i)}\right]$ and $\boldsymbol{\phi}^{(i)} = \left[\phi_{0_1}^{(i)}, \phi_{0_2}^{(i)}, \phi_{0_3}^{(i)}, \boldsymbol{\phi}_{Q_1}^{(i)}, \boldsymbol{\phi}_{Q_2}^{(i)}, \boldsymbol{\phi}_{Q_3}^{(i)}, \boldsymbol{\phi}_{X_1}^{(i)}, \boldsymbol{\phi}_{X_2}^{(i)}\right]$ correspond to the vectors of regression coefficients for $\mu_t^{(i)}$ and $\sigma_t^{(i)}$, respectively; $\boldsymbol{Q}_{f,k,t}^{(i)}$ is the vector of the k-day lead streamflow forecasts from z hydrological models at the gauge i for day t; $\boldsymbol{X}_{t-k}^{(i)}$ is the set of other m hydrometeorological covariates at the gauge i for day $t-k$; $\boldsymbol{P}_{t-k}^{(i)}$ is the precipitation covariate at gauge i for day $t-k$; $Q_{f,k,t,l}^{(i)}$ is the k-day lead streamflow forecasts (f) at gauge i and day t from the model l; $Q_{f,k,l,th}^{(i)}$ denotes the threshold considered for k-day lead streamflow forecasts at gauge i from the model l; and $g(\cdot)$ corresponds to a nonlinear transformation applied to the hydrological models' streamflow forecasts. Only temporal covariates are considered to help capture nonstationarity. We included the nonlinear term in Equation 3 to address the hydrological model's potential inability to capture high flow events. Equations 3 and 4 are simplified to a simple equation with $2(z + m + 1)$ regression coefficients when precipitation is not included as a covariate. We considered weakly informative normal priors centered at 0, with a standard deviation of 10 for the regression coefficients. The reader can find more details about the likelihood function in Equations S1 and S2 in Supporting Information S1.

In addition to the VIC streamflow forecast, other hydrometeorological covariates considered are as follows: the spatially averaged daily forecasted precipitation for the area between two consecutive gauges at the k-day lead time $\left(P_{f,t-k}^{(i)}\right)$, the daily VIC simulated streamflow $\left(Q_{s,t-k}^{(i+1)}\right)$ from the immediate upstream gauge, and the spatially averaged daily observed precipitation for the area between two consecutive gauges $\left(P_{t-k}^{(i)}\right)$. Also, we considered 80th quantiles from VIC streamflow forecast for different lead times as the threshold in Equations 3 and 4 (see Table S1 in Supporting Information S1).

For each lead time, we fitted various candidate BHMCs using different combinations of covariates for 16 years (2003–2018). The best BHMC was selected based on the lowest value of the leave-one-out cross-validation information criteria (LOOIC; Vehtari et al., 2017). Candidate BHMCs were fit using the program STAN (Stan

Table 2
Performance Metrics Considered to Assess the Bayesian Hierarchical Model Combination Performance

Notation	Name	Equation	Description
CE	Coefficient of efficiency	$CE = 1 - \frac{\sum_{t=1}^{T}\left(Q_t - \hat{Q}\right)^2}{\sum_{t=1}^{T}\left(Q_t - Q_{\mathrm{ref},t}\right)^2}$	Deterministic metric that varies (∞, 1], with a perfect score of 1. It compares how accurate and informative the model forecast is with respect to a reference forecast (Briffa et al., 1988).
R	Correlation coefficient	$R = \frac{\sum_{t=1}^{T}\left(\hat{Q}_t - \bar{\hat{Q}}\right)\left(Q_t - \bar{Q}\right)}{\sqrt{\sum_{t=1}^{T}\left(\hat{Q}_t - \bar{\hat{Q}}\right)^2}\sqrt{\sum_{t=1}^{T}\left(Q_t - \bar{Q}\right)^2}}$	A deterministic metric that varies [−1, 1] with a perfect score of 1. It measures the linear association between forecasts and observations independent of the mean and variance of the marginal distributions.
BIAS	Relative bias	$BIAS = \frac{\left(\bar{\hat{Q}} - \bar{Q}\right)}{\bar{Q}} \times 100$	Deterministic metric that varies (−∞, ∞), with a perfect score of 0%.
CRPSS	Continuous ranked probability skill score	$CRPSS = 1 - \frac{CRPS_{fcst}}{CRPS_{ref}}$ $CRPS = \frac{1}{T}\sum_{t=1}^{T}\int_{-\infty}^{\infty}\left[F(Q) - F_o(Q)\right]^2 dQ$ $Fo(Q) = \begin{cases} 0 & Q < Qo \\ 1 & Q < Qo \end{cases}$	A probabilistic metric that varies (− ∞, 1, with a perfect score of 1. It measures the skill of CRPS relative to a reference forecast (Hersbach, 2000). A score of 0 means the same skill as the reference forecast, while a value below 0 indicates a lower skill than the reference forecast. CRPS quantifies the difference between the cumulative distribution function (CDF) of a forecast (F) and the corresponding CDF of the observations (F_o).
PIT	Probability integral transform (PIT) value	$p(t) = F(t, Q_t)$	If PIT values are uniformly distributed, a set of forecasts at $t = 1$, 2, …, T is reliable (Wang et al., 2009).
A	α-Index	$\alpha = 1 - \frac{2}{T}\sum_{t=1}^{T}\left\|p(t) - p_U(t)\right\|$	The α-index describes the tendency of PIT values to deviate from the diagonal in PIT plots. It varies between 0 (worst reliability, with all observed p values equal to 0 or 1) and 1 (perfect reliability; Renard et al., 2010)

Note. T, number of observations; $\hat{Q}_t$, forecast ensemble mean for day t; $\bar{\hat{Q}}$, temporal average over forecast ensemble means; Q_t, observation for day t; $F(t, [])$, CDF of the forecast ensemble at time t; $p_U(t)$, theoretical value corresponding to $p(t)$.

Development Team, 2014) and its R extension (Stan Development Team, 2020). It uses a Markov chain Monte Carlo simulation algorithm (the No-U-Turn Sampler for the current example) to simulate the posterior probability distribution of the regression coefficients. We simulated three chains, ran the model for 4,000 iterations, and the first half of the simulations were discarded as burn-in to ensure convergence. We set a thinning factor of 2 to reduce the sample autocorrelation, resulting in 3,000 posterior samples (1,000 samples from each chain). To check the convergence of the posterior distribution of each regression coefficient, we used the shrink factor, $\hat{R}$, proposed by Gelman and Rubin (1992)—values lower than 1.1 for all the regression coefficients suggest model convergence. Then, the posterior distributions of the parameters and the predictive posterior distribution of daily streamflow forecasts were represented by 3,000 ensemble members for each lead time.

2.4. Evaluation of the Streamflow Forecast Performance

The out-of-sample predictability of the multimodel forecast from the best BHMC for each lead time is assessed by performing a leave-1-year-out cross-validation for 2003–2018 (16 years). For this, the daily observations in a year are selected as validation data (62 days), and the data from the remaining 15 years are selected as training data for the best BHMC. Then, the trained BHMC is used to provide estimates for the one validation year. This procedure is repeated 16 times. We model and issue forecasts for 1–10-day lead times for each day in the season.

We compute two deterministic and one probabilistic performance metrics (Table 2). The deterministic performance metrics are the Pearson correlation coefficient (R) and the relative bias (BIAS). The overall probabilistic skill is evaluated using the continuous ranked probability skill score (CRPSS; Hersbach, 2000), which quantifies the average error between a forecast cumulative distribution function (CDF) with that from the observation. CRPS corresponds to the mean absolute error for a deterministic forecast such as the VIC streamflow forecast. The associated skill score (CRPSS) is computed using the formulation provided by Wilks (2011) and shown in Table 2.

Table 3
LOOIC Values for the Best Candidate BHMC for Different Lead Times and Structure Specifications

		Garudeshwar		Mandleshwar		Handia		Hoshangabad		Sandiya	
Lead time k (days)	LOOIC	a	$X^{(1)}_{t-k}$	a	$X^{(2)}_{t-k}$	a	$X^{(3)}_{t-k}$	a	$X^{(4)}_{t-k}$	a	$X^{(5)}_{t-k}$
1	9,967	2.0	$Q^{(2)}_{s,t-1}$	2.5	$Q^{(3)}_{s,t-1}$	1.5	$P^{(3)}_{t-1}$	2.0	$Q^{(5)}_{s,t-1}$	2.0	$Q^{(6)}_{s,t-1}$
2	9,672	1.5	$Q^{(2)}_{s,t-2}$	2.0	$Q^{(3)}_{s,t-2}$	1.5	$Q^{(4)}_{s,t-2}$	1.5	$Q^{(5)}_{s,t-2}$	1.5	$Q^{(6)}_{s,t-2}$
3	11,302	3.5	$Q^{(2)}_{s,t-3}$	2.0	$Q^{(3)}_{s,t-3}$	2.5	$Q^{(4)}_{s,t-3}$	2.5	$Q^{(5)}_{s,t-3}$	1.5	$Q^{(6)}_{s,t-3}$
4	11,725	2.0	$Q^{(2)}_{s,t-4}$	1.5	$Q^{(3)}_{s,t-4}$	1.5	$Q^{(4)}_{s,t-4}$	1.5	$Q^{(5)}_{s,t-4}$	1.5	$Q^{(6)}_{s,t-4}$
5	12,491	2.0	$Q^{(2)}_{s,t-5}$	1.5	$Q^{(3)}_{s,t-5}$	1.5	$Q^{(4)}_{s,t-5}$	1.5	$Q^{(5)}_{s,t-5}$	3.5	$Q^{(6)}_{s,t-5}$
6	12,932	2.0	$Q^{(2)}_{s,t-6}$	2.0	$Q^{(3)}_{s,t-6}$	2.0	$Q^{(4)}_{s,t-6}$	2.0	$Q^{(5)}_{s,t-6}$	2.0	$Q^{(6)}_{s,t-6}$
7	13,242	2.0	$Q^{(2)}_{s,t-7}$	2.0	$Q^{(3)}_{s,t-7}$	2.0	$Q^{(4)}_{s,t-7}$	2.0	$Q^{(5)}_{s,t-7}$	2.0	$Q^{(6)}_{s,t-7}$
8	13,401	2.5	$Q^{(2)}_{s,t-8}$	2.5	$Q^{(3)}_{s,t-8}$	2.5	$Q^{(4)}_{s,t-8}$	2.5	$Q^{(5)}_{s,t-8}$	2.5	$Q^{(6)}_{s,t-8}$
9	13,453	2.0	$Q^{(2)}_{s,t-9}$	2.0	$Q^{(3)}_{s,t-9}$	3.0	$Q^{(4)}_{s,t-9}$	2.0	$Q^{(5)}_{s,t-9}$	2.0	$Q^{(6)}_{s,t-9}$
10	13,507	2.5	$Q^{(2)}_{s,t-10}$	2.5	$Q^{(3)}_{s,t-10}$	3.0	$Q^{(4)}_{s,t-10}$	2.5	$Q^{(5)}_{s,t-10}$	2.5	$Q^{(6)}_{s,t-10}$

For CRPSS, we considered the VIC streamflow forecast and climatology as reference forecasts. Bootstrapping is applied to the historical streamflows (3,000 samples) to generate uncertainty for the performance metrics.

Although CRPSS measures both accuracy and reliability, we checked the reliability of the BHM streamflow forecast by the uniformity of the probability integral transform (PIT; Gneiting et al., 2007). For a perfectly reliable forecast, PIT values should be uniformly distributed. Thus, PIT values are plotted against a standard uniform distribution in a PIT plot. Perfectly reliable forecasts follow the 1:1 line in PIT plots. Along with the PIT plot, we also computed the α-index to quantify the tendency of PIT values to deviate from the 1:1 line in PIT plots.

Reliability is a relevant feature but not enough for a good forecast—for example, a climatology forecast is perfectly reliable but does not hold information for short-term streamflow forecasting. Good forecasts need to be sharp while being reliable. If a forecast distribution is narrow, it is sharp. Here, we checked the forecast sharpness by the average width of the 95% confidence intervals (AWCI), which describes the ensemble spread (Gneiting et al., 2007; Li et al., 2017).

3. Results

3.1. Best Candidate BHMC Selection for Each Lead Time

For each lead time, the best BHMC was found to have the following structure for $\mu_t^{(i)}$ and $\sigma_t^{(i)}$ at gauge i for day t:

$$\mu_t^{(i)} = \begin{cases} \beta_{0_1}^{(i)} + \beta_{Q_1}^{(i)} \left(Q_{f,k,t}^{(i)}\right)^a + \beta_{X_1}^{(i)} X_{t-k}^{(i)} & Q_{f,k,t}^{(i)} \le Q_{f,k,80th}^{(i)} \\ \beta_{0_2}^{(i)} + \beta_{Q_2}^{(i)} \left(Q_{f,k,t}^{(i)}\right)^a + \beta_{X_2}^{(i)} X_{t-k}^{(i)} & Q_{f,k,t}^{(i)} > Q_{f,k,80th}^{(i)} \end{cases} \tag{5}$$

$$\sigma_t^{(i)} = \begin{cases} \phi_{0_1}^{(i)} + \phi_{Q_1}^{(i)} Q_{f,k,t}^{(i)} + \phi_{X_1}^{(i)} X_{t-k}^{(i)} & Q_{f,k,t}^{(i)} \le Q_{f,k,80th}^{(i)} \\ \phi_{0_2}^{(i)} + \phi_{Q_2}^{(i)} Q_{f,k,t}^{(i)} + \phi_{X_2}^{(i)} X_{t-k}^{(i)} & Q_{f,k,t}^{(i)} > Q_{f,k,80th}^{(i)} \end{cases} \tag{6}$$

The gauges were enumerated from downstream to upstream (Figure 1), that is, Garudeshwar ($i = 1$), Mandleshwar ($i = 2$), Handia ($i = 3$), Hoshangabad ($i = 4$), Sandiya ($i = 5$), and the artificial gauge ($i = 6$). Table 3 displays for each lead time, LOOIC values from the best candidate BHMC, the exponent of the VIC streamflow forecast covariate (a), and the specification of the additional covariate $\left(X_{t-k}^{(i)}\right)$. The significance of the regression coefficients (β and ϕ) was checked (posterior PDFs do not contain zero in the 95% credible interval).

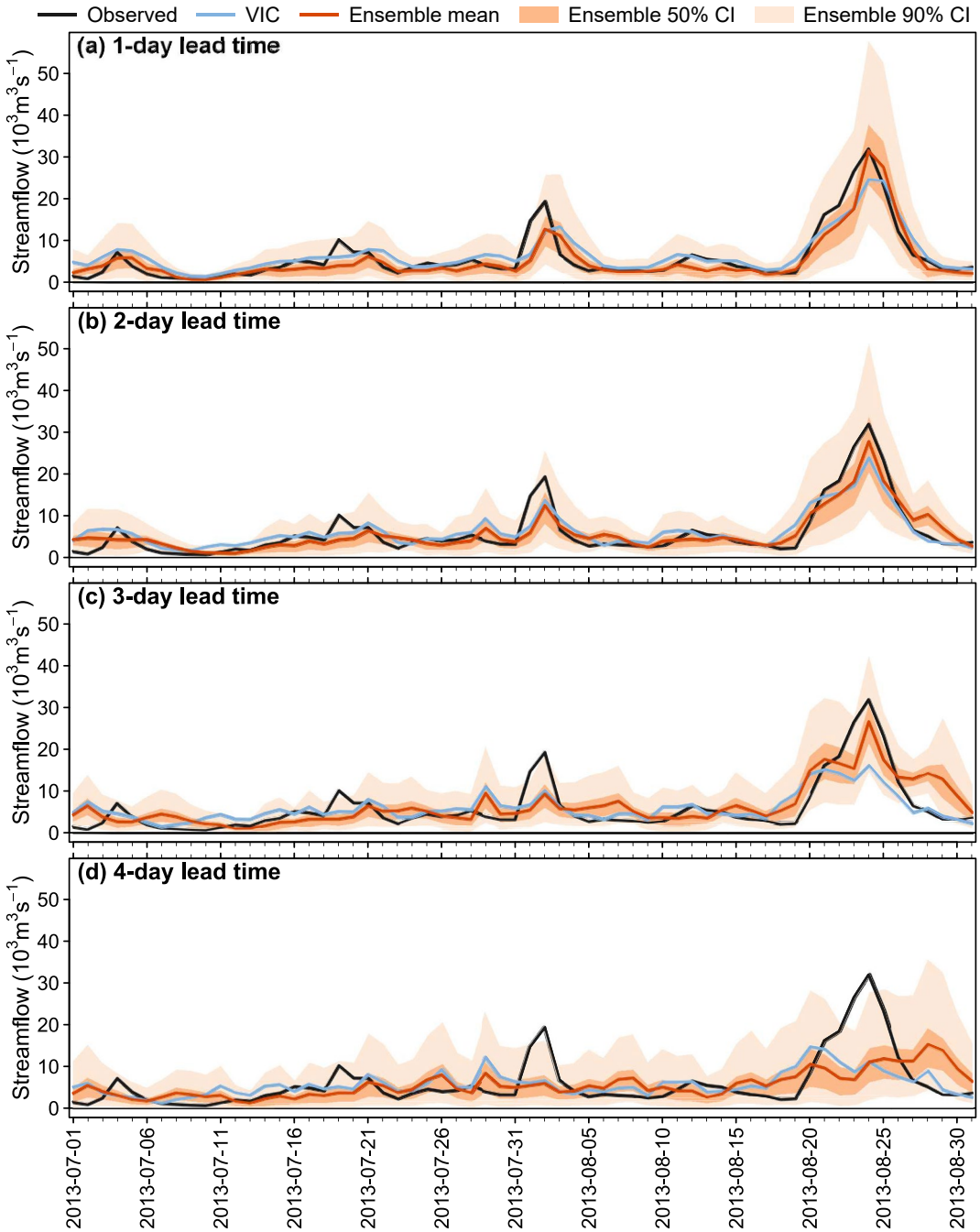

Figure 3. Times series of observations, Variable Infiltration Capacity (VIC) forecast, and cross-validated Bayesian hierarchical model combination (BHMC) ensembles forecast of daily monsoon streamflow for 2013 at Handia gauge for (a) a 1-day, (b) a 2-day, (c) a 3-day, and (d) a 4-day lead time. Black lines denote observed streamflow, light blue lines VIC forecast, orange lines BHMC ensemble mean, and orange and light orange bands 50% and 90% ensembles credible intervals.

3.2. Cross-Validation

Figure 3 displays cross-validated streamflow ensembles forecast from the best BHMC for 2013 (the highest flow year on the record) at Handia gauge for 1–4-day lead times. Most of the high flows (except 19 August) are captured by the BHMC streamflow forecast ensemble spread up to a 3-day lead time. Also, the ensemble mean is closer to the highest observed flow on the record (24 August). For 4–10-day lead times, neither VIC nor BHMC ensemble forecast can capture the timing and magnitude of the peak flows. The other gauges show simi-

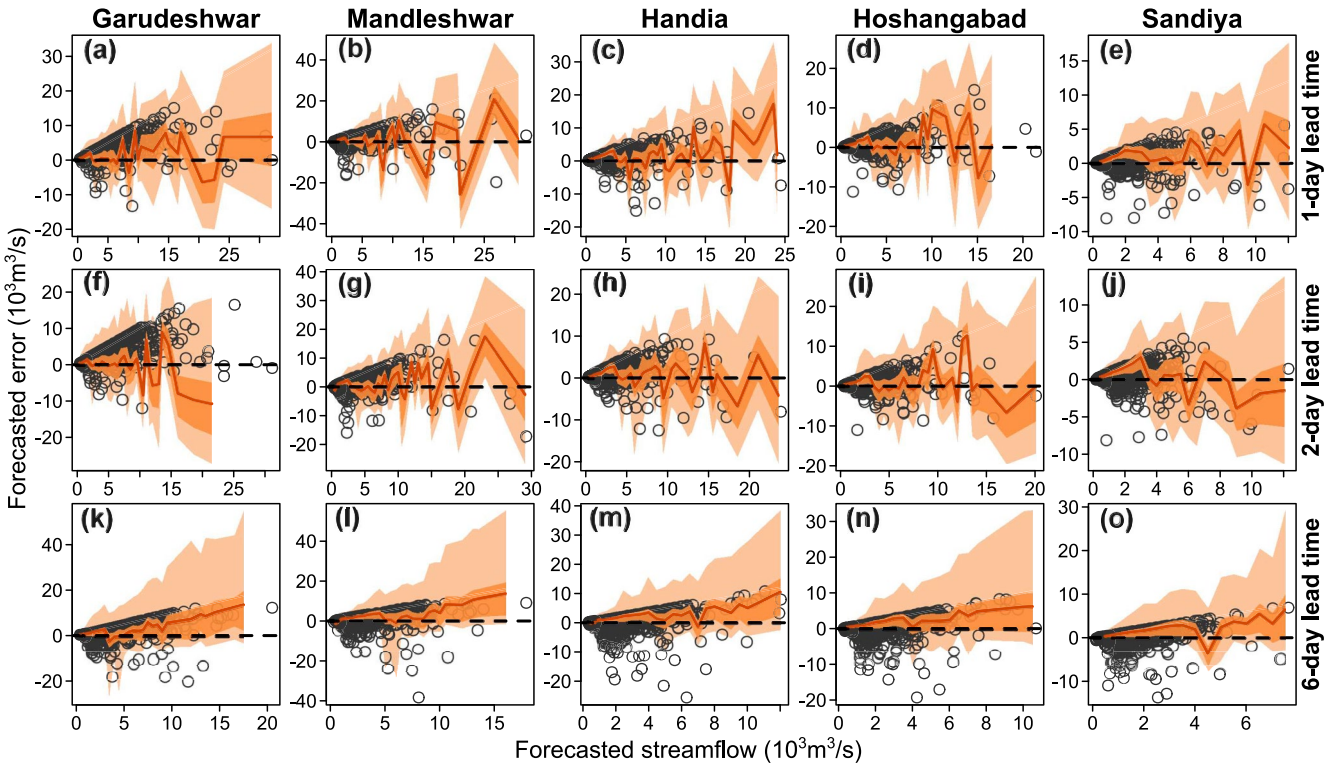

Figure 4. Plots of forecasted error versus forecasted streamflow at the five gauges of the Narmada River basin (NRB) for three lead times. Orange lines denote the ensemble mean; orange and light orange bands 50% and 95% confidence intervals; and the gray dots the errors in Variable Infiltration Capacity (VIC) forecasted streamflow. Rows and columns are associated with error uncertainty obtained from Bayesian hierarchical model combination (BHMC) cross-validation for three lead times and different gauges of the NRB. The forecasted error corresponds to the historical flow minus the forecasted flow from the ensembles.

lar performance (Figures S1–S4 in Supporting Information S1). The overall assessment of the BHMC ensemble forecasts is presented in the following sections.

3.2.1. Homoscedasticity of the Residuals

Homoscedasticity of the residuals (i.e., constant error variance for all magnitude of forecasted flows) is a desired feature of a forecast model. Still, for higher flow values, heteroscedasticity is not uncommon. To examine the degree of heteroscedasticity at the five gauges in the NRB, Figure 4 shows the relationships between forecasted streamflow and the corresponding forecasted error from the BHMC cross-validation for three lead times. The BHMC forecasted flows show low heteroscedasticity in the ensemble mean (orange lines oscillating around the null error line) compared to VIC forecasted streamflow (gray points) for 1- and 2-day lead times (Figures 4a–4j) at all gauges except for Garudeshwar. At Garudeshwar, the BHMC ensembles overestimate (Figure 4a) and underestimate (Figure 4f) high flows for 1- and 2-day lead. Low heteroscedasticity is also checked until the 5-day lead time (Figure S5 in Supporting Information S1). For the 6-day lead time (Figures 4k–4o), although there is a reduction of the error in the ensemble means, they exhibit a clear positive trend at all gauges. This trend is also observed from 7–10-day lead (Figures S5 and S6 in Supporting Information S1), indicating that homoscedasticity is not achieved at lead times longer than 5-day lead.

3.2.2. Accuracy Metrics

Figures 5 and 6 display error bars of the BIAS and Pearson correlation coefficient (R), respectively, for the entire period (2003–2018) and the three flow event strata (low, midrange, and high forecasted events) as a function of lead time at the five gauges of the NRB. For the entire period and high flow event strata, the BHMC forecast outperforms the VIC forecast significantly (95% confidence levels [CLs] do not overlap) based on the BIAS for most of the lead times with the median close to zero, except at Garudeshwar for lead times longer than 6-day (Figures 5a and 5d). These enhancements translate into absolute bias reductions ranging between 28% and 140%, highlighting the high bias reduction for Garudeshwar. In the case of the correlation (Figures 6a and 6d), the

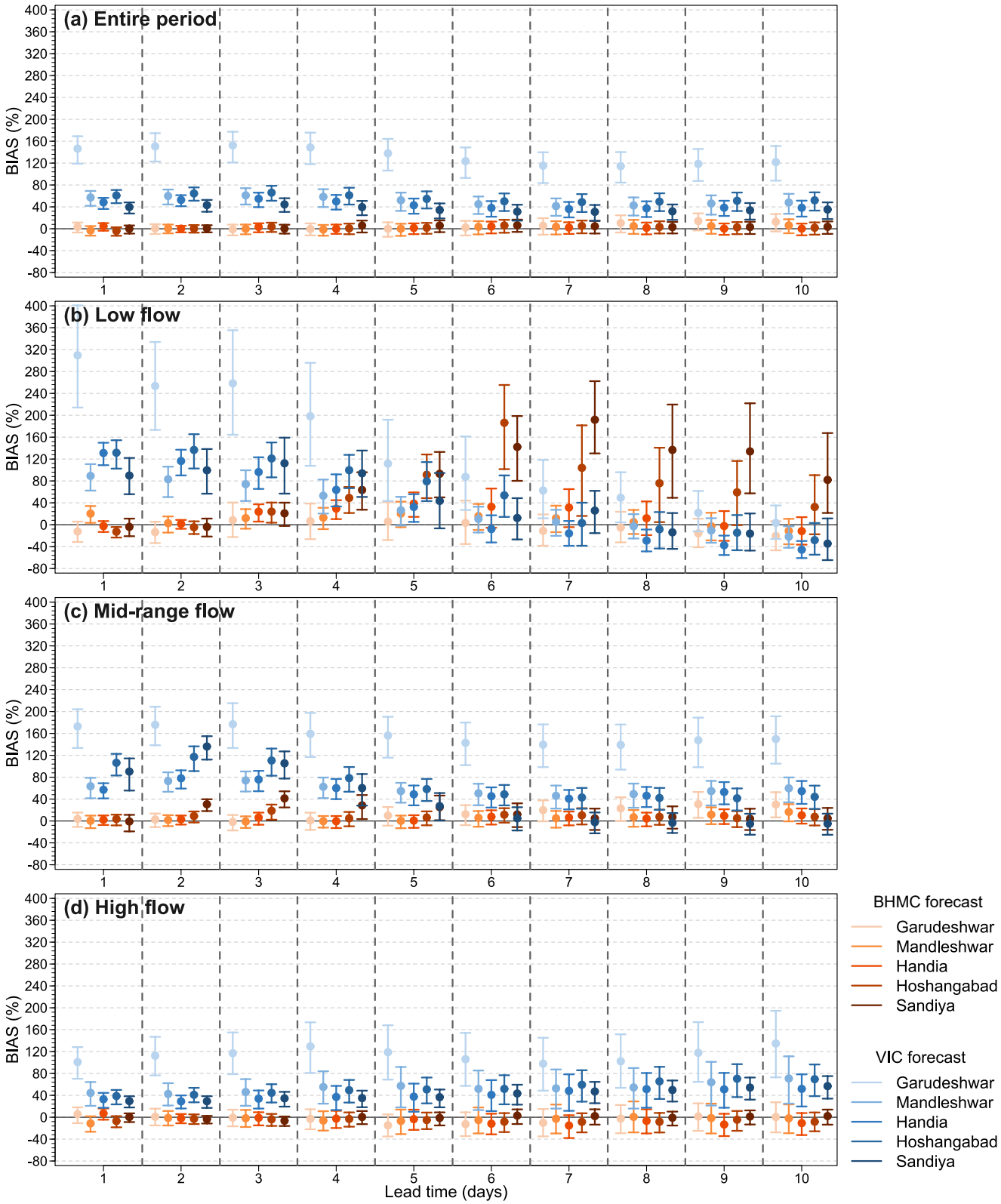

Figure 5. Error bars of BIAS as a function of the lead time at the five gauges of the Narmada River basin (NRB) for (a) the entire period, (b) low, (c) midrange, and (d) high flow forecasted flow events strata. The points and error bars denote the medians, and the 95% confidence levels (CLs) obtained through bootstrapping.

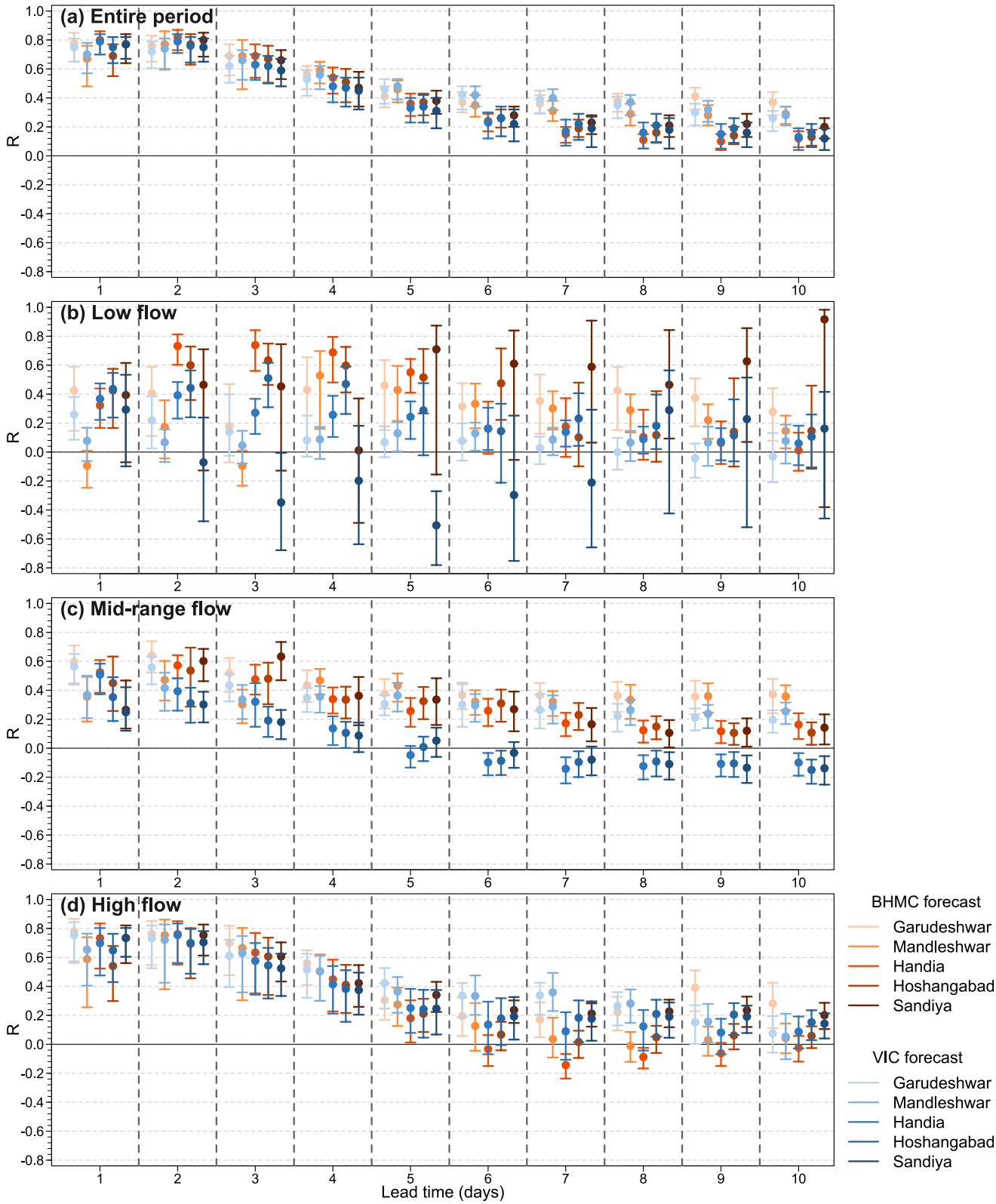

Figure 6. As in Figure 5 but for the Pearson correlation coefficient (*R*).

BHMC forecast at least preserves (95% CLs almost overlap) the correlation of the VIC forecast, and *R* median values are above 0.6 for 1–3-day lead times. For the low flow strata, there are considerable reductions of the BIAS relative to VIC (medians close to 0) at all gauges up to a 3-day lead time (Figure 5b). Besides, there is an increase in BIAS (positive BIAS) relative to VIC at Hoshangabad and Sandiya gauges for lead times above 5-day. For the correlation coefficient (Figure 6b), for a 1-day lead, there is no significant difference between HBMC and VIC for all gauges but Mandleshwar. From the 2-day lead time on, a considerable improvement on *R* is seen at all gauges except for a few cases (e.g., Garudeshwar for 3-day lead time). Lastly, in the case of the midrange flow strata, BHMC outperforms VIC based on the BIAS for all lead times except for the Sandiya gauge from 5- to 10-day lead time (Figure 5c). For *R* (Figure 6c), *R* values' improvements are observed at most gauges for 2–10-day lead times. For a 1-day lead time, there are no significant differences between BHMC and VIC forecast.

Figure 7 presents the error bars of CRPSS from BHMC with respect to lead time at all gauges for the entire period and the three flow strata. Compared to the VIC forecast (orange colors scheme), the BHMC forecast yields a substantial improvement in skill (95% CL and median above 0.2 and 0.4, respectively) for the entire period and all the flow strata at most of the gauges, except for the low flow forecasted after the 5-day lead time (Figure 7b). On the other hand, BHMC yields a significant increase in skill at all gauges compared to observed climatology (purple colors scheme) for the entire period and the three strata up to 5-day lead times. For 6–10-day lead times, differences in skill between BHMC and climatology are negligible for the entire period and the three strata.

3.2.3. Reliability and Sharpness

The reliability metric, PIT plots, of the BHMC forecast for different lead times and the α-index versus the lead time for the five gauges of the NRB are shown in Figure 8. From 1- to 3-day lead time (Figures 8a and 8b), the PIT plots are very close to the 1:1 line for all gauges, except for Garudeshwar at the 3-day lead (slight under forecasting), indicating good forecast reliability. From the 4-day lead time on (Figures 8c–8e), PIT plots are also reasonably close to the 1:1 line but tend to fall above it, indicating under forecasting. However, PIT plots fall inside the 95% Kolmogorov-Smirnoff confidence intervals for all the gauges. The α-index plot confirms the excellent reliability for 1–3-day lead times (above 0.94 except for Garudeshwar at a 3-day lead). It also exhibits an abrupt reduction for the 4-day lead time at all gauges except at Handia (drops up to 0.9). Also, it shows a sustained decline until the 9-day lead time, but its values remained high (above 0.87).

To assess the reliability of the forecasted flow events strata, Figures 9 and 10 display PIT plots for the three strata (low, midrange, and high flow forecasted events) for 1- and 4-day lead times, respectively. For a 1-day lead time (Figure 9), the low flow strata exhibit the lowest reliability of the three strata, with PIT plot shapes varying between under forecasting, over forecasting, and overconfident (s-shape) at different gauges. Also, PIT plots fall outside of the 95% Kolmogorov-Smirnoff confidence intervals at some gauges (Sandiya mostly). The PIT plots for midrange (Figure 9b) and high (Figure 9c) flow events strata are reasonably close to the identity line, and they fall inside the 95% confidence intervals. Still, their shapes are opposite—under forecasting for midrange flows and over forecasting for high flows. The α-index for low flows strata shows values above 0.79 and above 0.9 for midrange and high flows, indicating good reliability overall. Similar reliability performance is achieved for 2- and 3-day lead times (Figures S7 and S8 in Supporting Information S1).

For a 4-day lead time (Figure 10), PIT plots for midrange and high flow strata (Figures 10b and 10c) exhibit shapes similar to those obtained in Figures 9b and 9c but accentuated for the midrange flow strata (PIT plots close to or falling outside of the 95% confidence intervals). PIT plots for low flow strata show good reliability at Handia, underconfident at Garudeshwar, and clear under forecasting at the rest of the gauges. These findings are reflected by the α-index (Figure 10d). The α-index values are below 0.9 for low flow strata at all gauges except Handia and above 0.9 for the high flow strata. Similar reliability is observed for longer lead times (Figures S9–S11 in Supporting Information S1).

Finally, Figure 11 shows AWCI for various lead times for the BHMC and reference (observed climatology) forecasts at the five gauges to evaluate the forecast sharpness. BHMC forecasts (orange lines) have much narrower 95% confidence intervals than the reference forecast (purple lines). For most of the gauges, the lowest AWCI occurs at 1- or 2-day lead time, and then it increases with lead time until a 6-day lead and remains relatively constant until a 10-day lead time. The exception is Garudeshwar, where the AWCI varies almost linearly from 1- to 9-day lead times. Overall, this reflects the increasing forecast uncertainty at longer lead times.

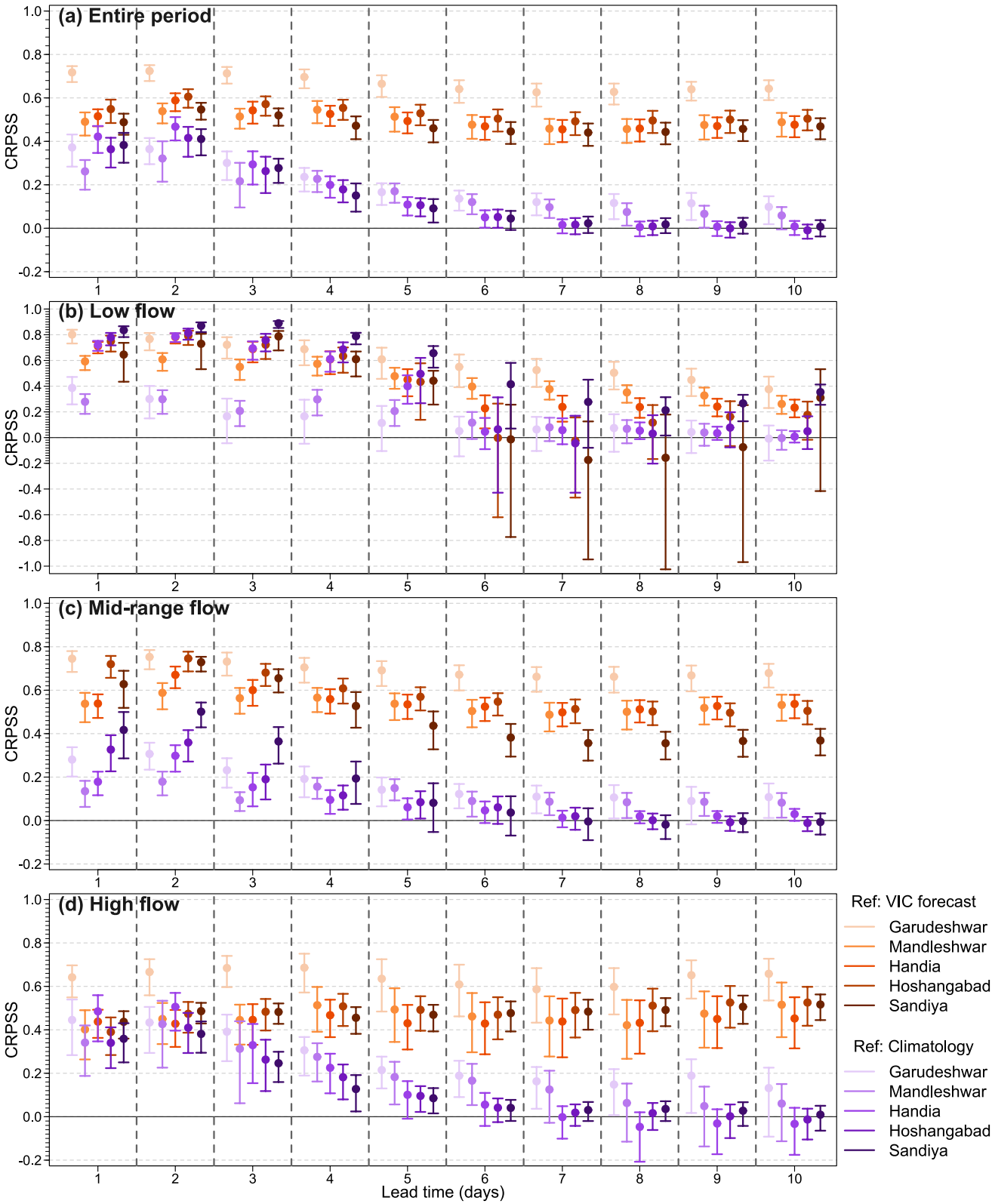

Figure 7. Error bars of the continuous ranked probability skill score (CRPSS) as a function of the lead time at the five gauges of the Narmada River basin (NRB) for (a) the entire period, (b) low, (c) midrange, and (d) high flow forecasted flow events strata. The points and error bars denote the medians, and the 95% confidence levels (CLs) obtained through bootstrapping.

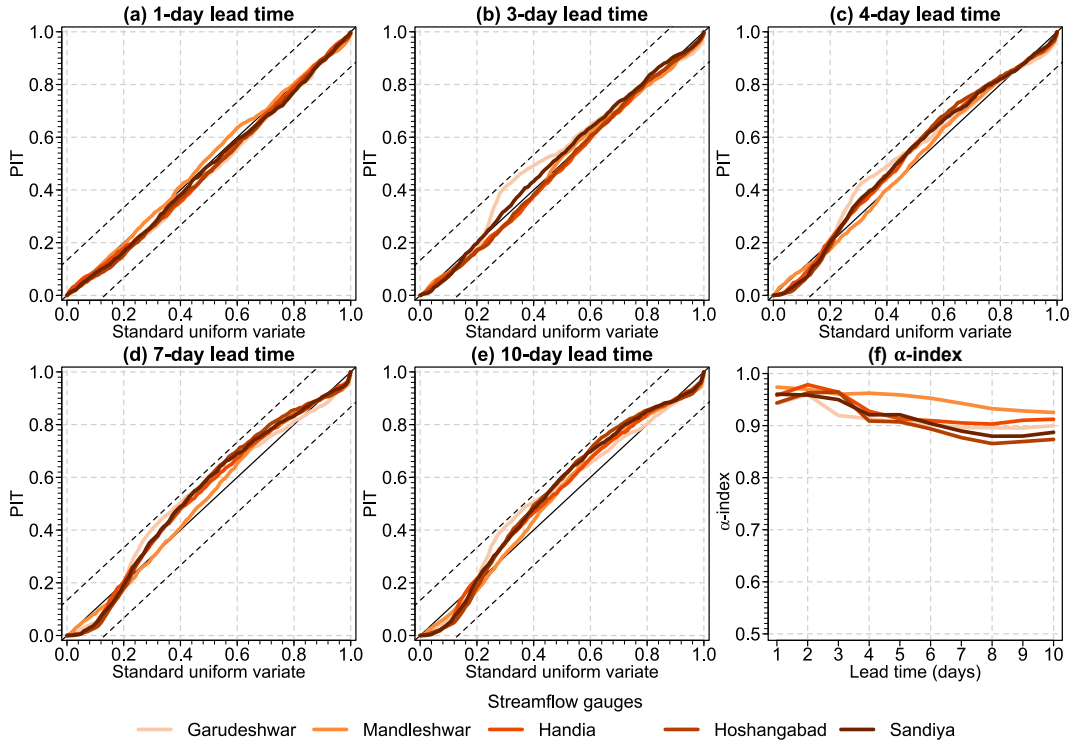

Figure 8. Probability integral transform (PIT) plots for forecasts Bayesian hierarchical model combination (BHMC) forecast at the five stations of the Narmada River basin (NRB) for (a) 1-day, (b) 3-day, (c) 4-day, (d) 7-day, and (e) 10-day lead time. (f) The α-index as a function of the lead time at the five gauges. Dashed lines in panels (a)–(e) denote the 95% Kolmogorov-Smirnoff confidence intervals.

3.3. Real-Time Forecast Performance for the 2021 Monsoon Season

As a short demonstration of the real-time application of the BHMC framework proposed, we obtained streamflow forecasts for July–August 2021 at the Handia gauge. We obtained stage data and corresponding streamflow for the 2016 season. We fitted the rating curve based on the simultaneous record of streamflow and stage for 2016 (Figure S12 in Supporting Information S1). Observed streamflow at Handia for 2021 was obtained as estimates from the fitted stage–discharge relationship, as there is a delay in receiving the "official" flows from the water agencies. Figure 12 shows the times series of observed streamflow, VIC, and BHMC forecasted streamflow for the 2021 monsoon (July–August) season. Although both VIC and BHMC forecasts overestimate the observed streamflow between 27 July and 11 August at the three lead times, there is a clear improvement in the BHMC performance, as it captures most of the observations inside of the 50% credible interval (orange bands), excepts for a few days (e.g., 20 August). Also, the BHMC forecasts show high correlation and low BIAS, up to three lead times. Note that 2021 was not a wet year since all the observed streamflow values are below the historical median, 1,210 (m^3/s), from the calibration period (2003–2018). This demonstration of a real-time forecast for the 2021 season offers bright prospects for skillful streamflow and river stage forecasts at long lead times, which are crucial for flood mitigation.

4. Discussion

The best candidate BHMCs selection showed that in addition to VIC forecasted streamflow, VIC simulated streamflow was selected as the second covariate at different lead times except for a 1-day lead time at Handia, where the observed precipitation was chosen as the second covariate. The VIC simulated streamflow has a strong streamflow persistence. Various authors have reported it as a significant skill contribution to streamflow forecasting (Bennett et al., 2021; Li et al., 2015, 2016, 2017); hence it is selected in the model. We did not lag observed streamflow as a potential covariate since its inclusion does not allow the implementation of the framework for real-time forecasting, as there is a delay between when the streamflow is recorded and when it becomes available

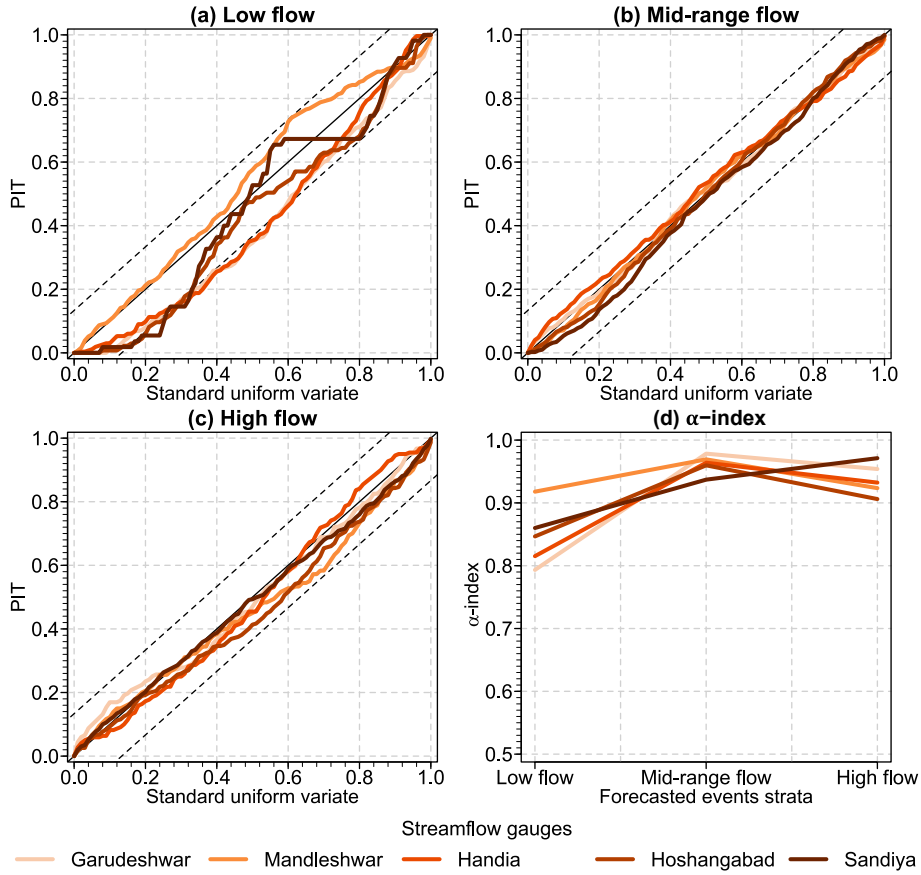

Figure 9. Probability integral transform (PIT) plots for forecasts Bayesian hierarchical model combination (BHMC) forecast at the five stations of the Narmada River basin (NRB) for a 1-day lead time for (a) low, (b) midrange, and (c) high flow forecasted events strata. (d) The α-index as a function of lead times at the five gauges. Dashed lines in panels (a)–(e) denote the 95% Kolmogorov-Smirnoff confidence intervals.

at each gauge in the study basin. As noted in the previous section, even for the 2021 season, the official streamflows are yet to be released. In addition, nonlinear transformations of the VIC streamflow forecast help reduce the bias related to high flow events, which is of interest.

Cross-validation results revealed that the BHMC approach enables handling the heteroscedasticity of the residuals in terms of the ensemble mean until 5-day lead times. In addition, deterministic accuracy metrics showed enhanced performance of BHMC ensembles mean compared to the raw VIC forecasted streamflow—i.e., absolute bias reductions of at least 28% for the entire period and the high flow forecasted events strata across the basin for all lead times. The highest bias reduction occurs at the terminal gauge (Garudeshwar), with values above 100% for the entire period and midrange and high flow forecasted events strata. Significant bias reductions for the low flow forecasted strata are obtained only until 5-day lead times. Regarding the Pearson correlation coefficients (R), results showed that BHMC enhanced or preserved the correlation between observed and forecasted streamflow for the entire period and all flow events strata at most lead times (high flow events at lead times higher than 5-day are the exception). The results also revealed a high correlation reduction after the 3-day lead time, which is caused by the poor precipitation forecast skill for lead times longer than 3-day. Furthermore, the ability of BHMC to capture the temporal coherence is not ensured by QM (Zhao et al., 2017), an approach widely used in hydroclimate applications and employed by Tiwari et al. (2022) to postprocess the VIC streamflow forecast considered for this study.

Regarding the probabilistic accuracy metrics, CRPSS error bars showed that BHMC ensemble forecast streamflow outperforms raw VIC streamflow forecast for the entire period—i.e., skill improvements of at least 40%. The BHMC forecast skill results were consistently better than the VIC streamflow forecast for the three strata (low, midrange, and high flows forecasted events strata) at all gauges but for the two upper gauges (Hoshangabad and

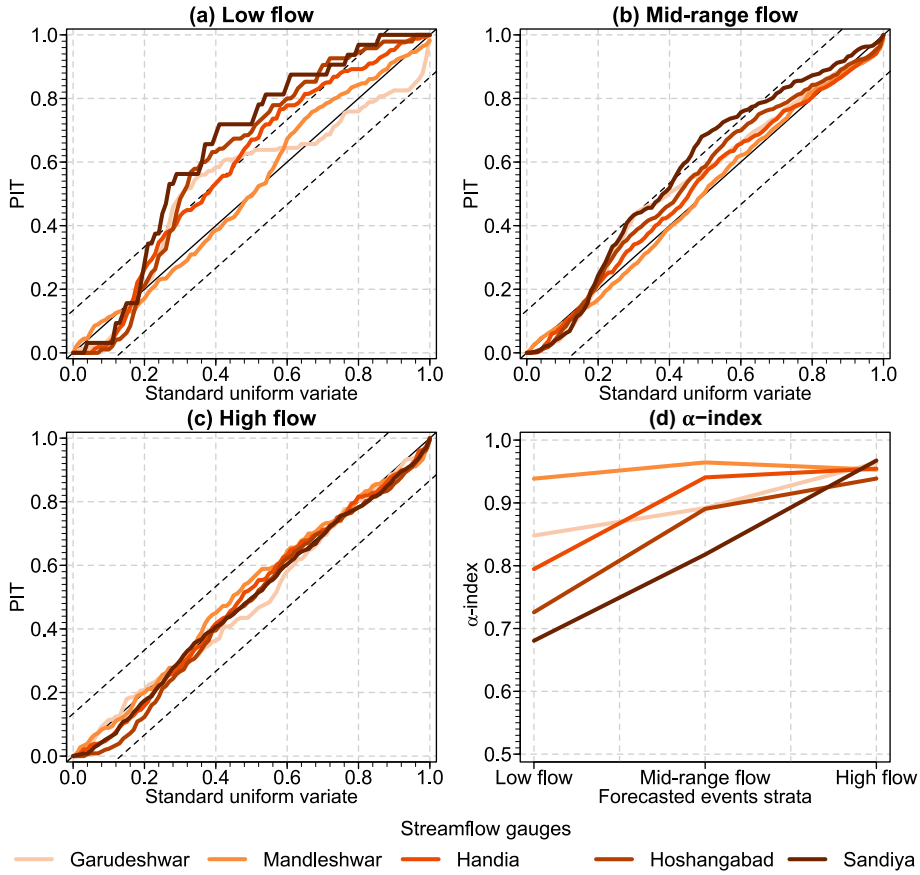

Figure 10. As in Figure 9 but for the 4-day lead time.

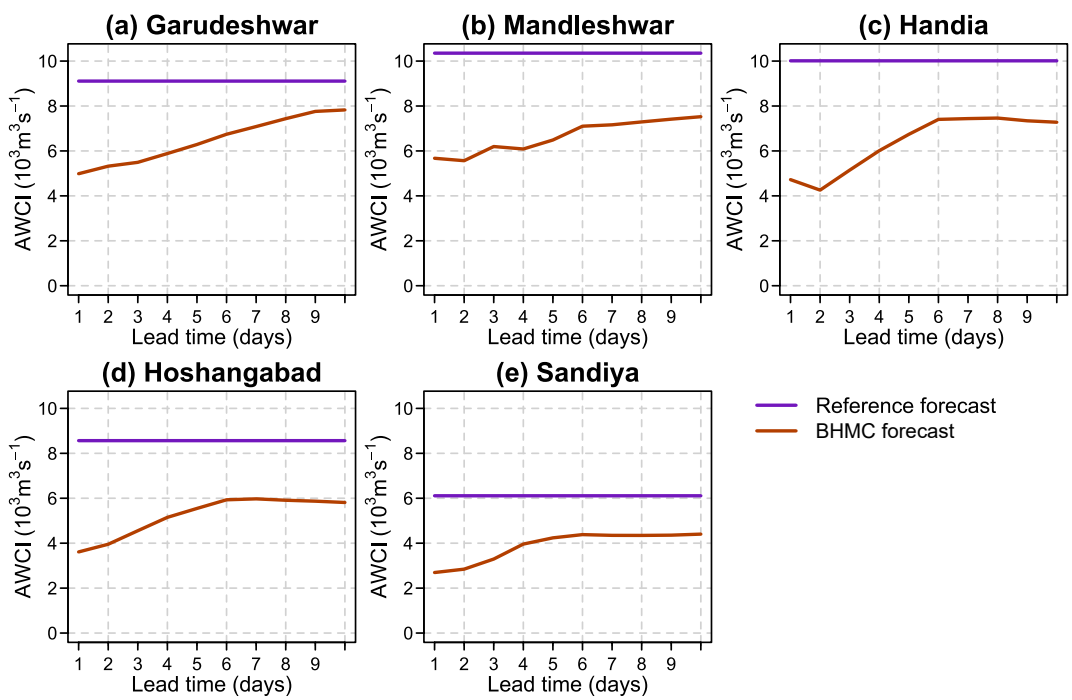

Figure 11. Average width of the 95% confidence intervals (AWCI) of the Bayesian hierarchical model combination (BHMC) ensemble streamflow forecasts as a function of the lead time at the five gauges of the Narmada River basin. The reference forecast corresponds to the observed climatology.

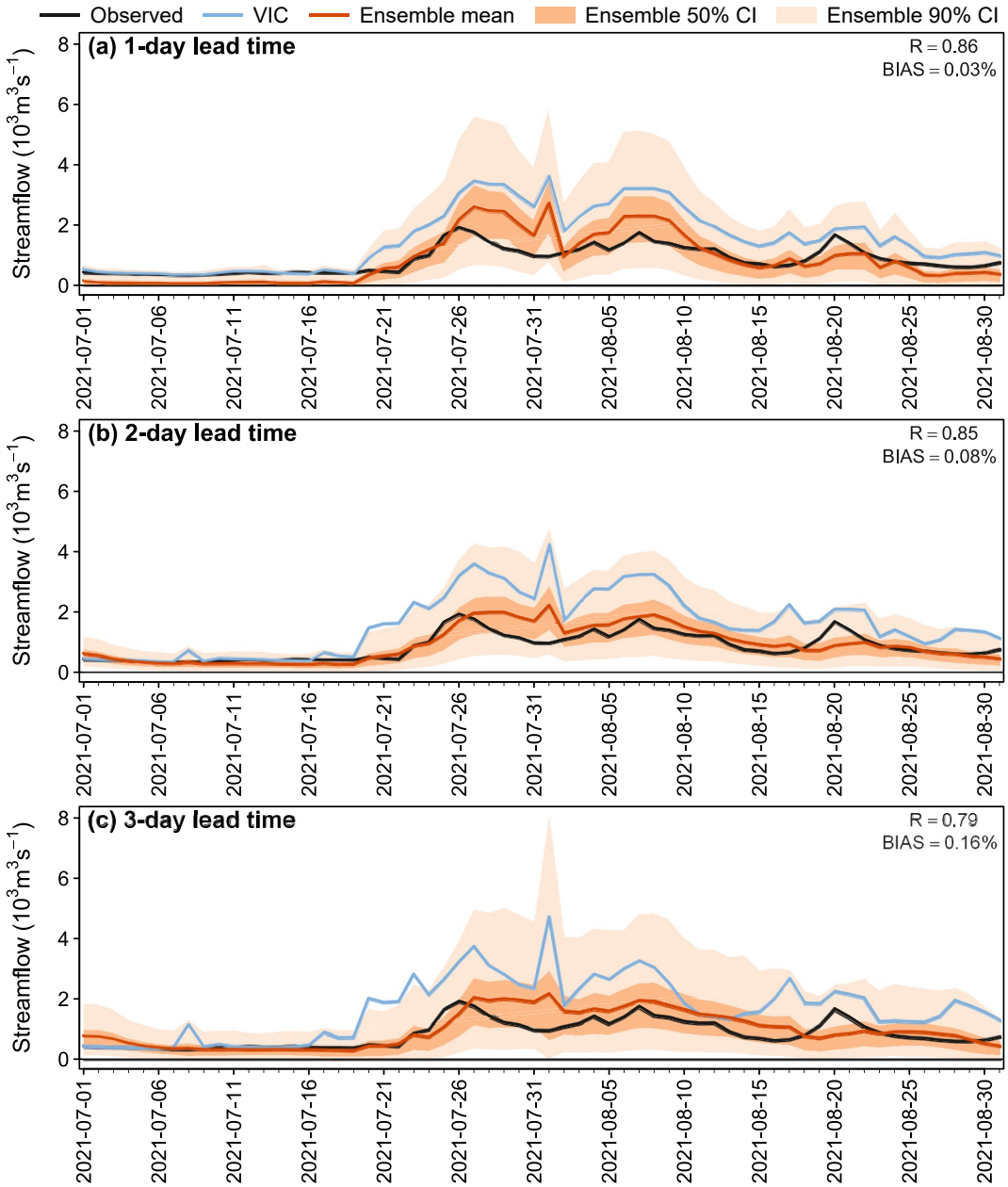

Figure 12. Times series of observations, Variable Infiltration Capacity (VIC) forecast, and Bayesian hierarchical model combination (BHMC) ensembles forecast of daily monsoon streamflow for 2021 at Handia gauge for (a) 1-day, (b) 2-day, and (c) 3-day lead time. Black lines denote observed streamflow, light blue lines VIC forecast, orange lines BHMC ensemble mean, and orange and light orange bands 50% and 90% ensembles credible intervals.

Sandiya) at lead times longer than 5-day. Compared to climatology, BHMC performs better for the entire period and flow events strata until 5-day lead time and with similar skill at longer lead times. This feature is consistent with the findings of other studies (Bennett et al., 2017, 2021). Further, PIT and AWCI plots showed that the BHMC frameworks provide sharp and reliable streamflow forecast ensembles until the 3-day lead times for the entire period and flow events strata, except at some gauges for the low flow events strata.

The results of this study indicate that the proposed BHMC framework can improve the forecast skill by reducing the nonsystematic bias in VIC model forecasts. This improved the temporal coherence and forecast reliability until 3-day lead time, until when the precipitation forecasts are skillful. For lead times longer than 3-day, accurate meteorological forcings are unavailable, and physically based models cannot predict streamflow processes

well. In that case, the BHMC proposed cannot fully correct the streamflow forecast errors as the covariates in the model are not skillful. Consequently, it can only provide a significant bias reduction but not a reliable forecast. This drawback could be overcome by including outputs from additional hydrological models, which could capture other features of streamflow processes, as potential covariates (e.g., Georgakakos et al., 2004; Krishnamurti et al., 2000; Mendoza et al., 2014; Rajagopalan et al., 2002).

In order to account for the impact of reservoir operations in the NRB, the VIC model implementation included a reservoir module and observations of water level storage at reservoirs in the basin for its calibration. Nevertheless, the observed water storage level was not included as a covariate in the Bayesian model as it is not available in real-time at present. However, if this information were available in future or in another basin where the BHMC is considered for implementation, it can be incorporated as a covariate to capture the reservoir effects.

5. Summary and Conclusions

The high occurrence of floods during the monsoon season due to most of the rainfall in India falling during this season and the unavailability of an accurate and robust streamflow forecast system have posed the need for skillful daily streamflow forecast over this region. To address this need, we adapted the Bayesian hierarchical framework developed by Ossandón et al. (2021, 2022) to incorporate multiple sources of information and provide robust real-time daily ensemble streamflow forecasting across a rainfed river basin. This BHMC considers a Gamma marginal distribution and includes information from different sources as covariates in the second layer of hierarchy. The study case presented here aims to generate real-time daily ensemble streamflow forecasts for the monsoon peak season (July–August) in the NRB in Central India.

Overall, the results obtained in this study demonstrate the enormous potential of this framework for short-term (1–3-day lead time) streamflow forecasting in the NRB, complementing and enhancing the current operational flood forecasting system (CWC, 2015). In a future extension of this work, additional hydrological and meteorological model forecasts can be incorporated into the framework as potential covariates in the BHMC. Further, the proposed framework can be easily implemented in other river basins worldwide, with varied network topologies and climates. Finally, the proposed framework has the potential to provide a skillful stage (water levels) forecast in the NRB, which will be of immense help to emergency managers, as it is the key decision variable for public officials in evacuations and mitigation.

Data Availability Statement

The data set used in this study, which consists of time series of potential covariates and daily streamflow for the peak monsoon season (July–August) at five station gauges in the Narmada River basin, and the scripts for the implementation of the BHMC proposed here, can be downloaded from https://zenodo.org/badge/latestdoi/502150441.

Acknowledgments
The publication of this article was funded by the University of Colorado Boulder Libraries Open Access Fund. This project was funded by the Monsoon Mission project of the Ministry of Earth Sciences, India. We also acknowledge the support from the Fulbright Foreign Student Program and the National Agency for Research and Development (ANID) Scholarship Program/DOCTORADO BECAS CHILE/2015-56150013 to the first author.

References

Ali, H., & Mishra, V. (2018). Increase in subdaily precipitation extremes in India under 1.5 and 2.0°C warming worlds. *Geophysical Research Letters*, *45*, 6972–6982. https://doi.org/10.1029/2018GL078689

Bennett, J. C., Wang, Q. J., Robertson, D. E., Bridgart, R., Lerat, J., Li, M., & Michael, K. (2021). An error model for long-range ensemble forecasts of ephemeral rivers. *Advances in Water Resources*, *151*, 103891. https://doi.org/10.1016/J.ADVWATRES.2021.103891

Bennett, J. C., Wang, Q. J., Robertson, D. E., Schepen, A., Li, M., & Michael, K. (2017). Assessment of an ensemble seasonal streamflow forecasting system for Australia. *Hydrology and Earth System Sciences*, *21*(12), 6007–6030. https://doi.org/10.5194/HESS-21-6007-2017

Briffa, K. R., Jones, P. D., Pilcher, J. R., & Hughes, M. K. (1988). Reconstructing summer temperatures in northern Fennoscandinavia back to AD 1700 using tree-ring data from Scots pine. *Arctic and Alpine Research*, *20*(4), 385–394. https://doi.org/10.2307/1551336

Cherkauer, K. A., Bowling, L. C., & Lettenmaier, D. P. (2003). Variable infiltration capacity cold land process model updates. *Global and Planetary Change*, *38*(1–2), 151–159. https://doi.org/10.1016/S0921-8181(03)00025-0

CWC. (2015). Flood forecasting and warning system in India. In *Regional flood early warning system workshop* (pp. 6–7).

FAO/IIASA/ISRIC/ISSCAS/JRC. (2012). *Harmonized World Soil Database (version 1.2)* (Tech. Rep.). IIASA/FAO.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Georgakakos, K. P., Seo, D. J., Gupta, H., Schaake, J., & Butts, M. B. (2004). Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. *Journal of Hydrology*, *298*(1–4), 222–241. https://doi.org/10.1016/J.JHYDROL.2004.03.037

Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *69*(2), 243–268. https://doi.org/10.1111/J.1467-9868.2007.00587.X

Haddeland, I., Skaugen, T., & Lettenmaier, D. P. (2006). Anthropogenic impacts on continental surface water fluxes. *Geophysical Research Letters*, *33*, L08406. https://doi.org/10.1029/2006GL026047

Hansen, M. C., Sohlberg, R., Defries, R. S., & Townshend, J. R. (2000). Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, *21*(6–7), 1331–1364. https://doi.org/10.1080/014311600210209
Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, *15*(5), 559–570. https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2
Hunt, K. M. R., & Fletcher, J. K. (2019). The relationship between Indian monsoon rainfall and low-pressure systems. *Climate Dynamics*, *53*(3–4), 1859–1871. https://doi.org/10.1007/s00382-019-04744-x
Hunt, K. M. R., Turner, A., & Parker, D. E. (2016). The spatiotemporal structure of precipitation in Indian monsoon depressions. *Quarterly Journal of the Royal Meteorological Society*, *142*(701), 3195–3210. https://doi.org/10.1002/qj.2901
Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–472. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2
Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., et al. (2001). The NCEP–NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bulletin of the American Meteorological Society*, *82*(2), 247–267. https://doi.org/10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2
Krishnamurti, T. N., Kishtawal, C. M., Zhang, Z., LaRow, T., Bachiochi, D., Williford, E., et al. (2000). Multimodel ensemble forecasts for weather and seasonal climate. *Journal of Climate*, *13*(23), 4196–4216. https://doi.org/10.1175/1520-0442(2000)013<4196:MEFFWA>2.0.CO;2
Kumar, A., Sridevi, C., Durai, V. R., Singh, K. K., Mukhopadhyay, P., & Chattopadhyay, N. (2019). MOS guidance using a neural network for the rainfall forecast over India. *Journal of Earth System Science*, *128*(5), 1–12. https://doi.org/10.1007/S12040-019-1149-Y
Li, M., Wang, Q. J., Bennett, J. C., & Robertson, D. E. (2015). A strategy to overcome adverse effects of autoregressive updating of streamflow forecasts. *Hydrology and Earth System Sciences*, *19*(1), 1–15. https://doi.org/10.5194/HESS-19-1-2015
Li, M., Wang, Q. J., Bennett, J. C., & Robertson, D. E. (2016). Error reduction and representation in stages (ERRIS) in hydrological modelling for ensemble streamflow forecasting. *Hydrology and Earth System Sciences*, *20*(9), 3561–3579. https://doi.org/10.5194/HESS-20-3561-2016
Li, M., Wang, Q. J., Robertson, D. E., & Bennett, J. C. (2017). Improved error modelling for streamflow forecasting at hourly time steps by splitting hydrographs into rising and falling limbs. *Journal of Hydrology*, *555*, 586–599. https://doi.org/10.1016/J.JHYDROL.2017.10.057
Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, *99*(D7), 415–429. https://doi.org/10.1029/94JD00483
Liang, X., Wood, E. F., & Lettenmaier, D. P. (1996). Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification. *Global and Planetary Change*, *13*(1–4), 195–206. https://doi.org/10.1016/0921-8181(95)00046-1
Liang, X., & Xie, Z. (2001). A new surface runoff parameterization with subgrid-scale soil heterogeneity for land surface models. *Advances in Water Resources*, *24*(9–10), 1173–1193. https://doi.org/10.1016/s0309-1708(01)00032-x
Marcolongo, A., Vladymyrov, M., Lienert, S., Peleg, N., Haug, S., & Zscheischler, J. (2022). Predicting years with extremely low gross primary production from daily weather data using Convolutional Neural Networks. *Environmental Data Science*, *1*, e2. https://doi.org/10.1017/EDS.2022.1
Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P., & Nijssen, B. (2002). A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *Journal of Climate*, *15*(22), 3237–3251. https://doi.org/10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2
Mendoza, P. A., Rajagopalan, B., Clark, M. P., Cortés, G., & McPhee, J. (2014). A robust multimodel framework for ensemble seasonal hydroclimatic forecasts. *Water Resources Research*, *50*, 6030–6052. https://doi.org/10.1002/2014WR015426
Mukhopadhyay, P., Prasad, V. S., Krishna, R. P. M., Deshpande, M., Ganai, M., Tirkey, S., et al. (2019). Performance of a very high-resolution Global Forecast System model (GFS T1534) at 12.5 km over the Indian region during the 2016–2017 monsoon seasons. *Journal of Earth System Science*, *128*(6), 1–18. https://doi.org/10.1007/S12040-019-1186-6
Nanditha, J. S., Rajagopalan, B., & Mishra, V. (2022). Combined signatures of atmospheric drivers, soil moisture, and moisture source on floods in Narmada River basin, India. *Climate Dynamics*, *59*, 2831–2851. https://doi.org/10.1007/S00382-022-06244-X
Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6
Ossandón, Á., Nanditha, J. S., Mendoza, P. A., Rajagopalan, B., & Mishra, V. (2022). A Bayesian hierarchical framework for post-processing daily streamflow simulations across a river network. *Journal of Hydrometeorology*, *23*(6), 947–963. https://doi.org/10.1175/jhm-d-21-0167.1
Ossandón, Á., Rajagopalan, B., Lall, U., Nanditha, J. S., & Mishra, V. (2021). A Bayesian hierarchical network model for daily streamflow ensemble forecasting. *Water Resources Research*, *57*, e2021WR029920. https://doi.org/10.1029/2021WR029920
Pai, D., Sridhar, L., Rajeevan, M., Sreejith, O. P., Satbhai, N. S., & Mukhopadhyay, B. (2014). Development of a new high spatial resolution (0.25° × 0.25°) long period (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. *MAUSAM*, *65*(1), 18. https://doi.org/10.54302/mausam.v65i1.851
Papalexiou, S. M., & Montanari, A. (2019). Global and regional increase of precipitation extremes under global warming. *Water Resources Research*, *55*, 4901–4914. https://doi.org/10.1029/2018WR024067
Rajagopalan, B., Lall, U., & Zebiak, S. (2002). Categorical climate forecasts through optimal combination of multiple GCM ensembles. *Monthly Weather Review*, *130*(7), 1792–1811. https://doi.org/10.1061/40685(2003)184
Renard, B., Kavetski, D., Kuczera, G., Thyer, M., & Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, *46*, W05521. https://doi.org/10.1029/2009WR008328
Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., et al. (2004). The global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, *85*(3), 381–394. https://doi.org/10.1175/BAMS-85-3-381
Sheffield, J., & Wood, E. F. (2007). Characteristics of global and regional drought, 1950–2000: Analysis of soil moisture data from off-line simulation of the terrestrial hydrologic cycle. *Journal of Geophysical Research*, *112*, D17115. https://doi.org/10.1029/2006JD008288
Sridevi, C., Singh, K. K., Suneetha, P., Durai, V. R., & Kumar, A. (2020). Rainfall forecasting skill of GFS model at T1534 and T574 resolution over India during the monsoon season. *Meteorology and Atmospheric Physics*, *132*(1), 35–52. https://doi.org/10.1007/s00703-019-00672-x
Stan Development Team. (2014). *Stan modeling language user's guide and reference manual*. Stan Development Team.
Stan Development Team. (2020). *RStan: The R interface to stan*. Stan Development Team. Retrieved from http://mc-stan.org/
Tiwari, A. D., Mukhopadhyay, P., & Mishra, V. (2022). Influence of bias correction of meteorological and streamflow forecast on hydrological prediction in India. *Journal of Hydrometeorology*, *23*(7), 1171–1192. https://doi.org/10.1175/JHM-D-20-0235.1
Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4
Wallemacq, P., & House, R. (2018). *Economic losses, poverty and disasters: 1998–2017* (Tech. Rep.). UNISDR and CRED.
Wang, Q. J., Robertson, D. E., & Chiew, F. H. (2009). A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. *Water Resources Research*, *45*, W05407. https://doi.org/10.1029/2008WR007355

Wasko, C., & Sharma, A. (2017). Continuous rainfall generation for a warmer climate using observed temperature sensitivities. *Journal of Hydrology*, *544*, 575–590. https://doi.org/10.1016/j.jhydrol.2016.12.002

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (3rd ed., Vol. 100). Academic Press Inc.

Zhao, T., Bennett, J. C., Wang, Q. J., Schepen, A., Wood, A. W., Robertson, D. E., & Ramos, M. H. (2017). How suitable is Quantile Mapping for postprocessing GCM precipitation forecasts? *Journal of Climate*, *30*(9), 3185–3196. https://doi.org/10.1175/JCLI-D-16-0652.1