Forecasting Magnitude and Frequency of Seasonal Streamflow Extremes Using a Bayesian Hierarchical Framework

Álvaro Ossandón[1,2] , **Balaji Rajagopalan**[1,3] , **and William Kleiber**[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]Department of Applied Mathematics, University of Colorado, Boulder, CO, USA

Abstract We develop a space-time Bayesian hierarchical modeling (BHM) framework for two flood risk attributes—seasonal daily maximum flow and the number of events that exceed a threshold during a season (NEETM)—at a suite of gauge locations on a river network. The model uses generalized extreme value (GEV) and Poisson distributions as marginals for these flood attributes with non-stationary parameters. The rate parameters of the Poisson distribution and location, scale, and shape parameters of the GEV are modeled as linear functions of suitable covariates. Gaussian copulas are applied to capture the spatial dependence. The best covariates are selected using the Watanabe-Akaike information criterion (WAIC). The modeling framework results in the posterior distribution of the flood attributes at all the gauges and various lead times. We demonstrate the utility of this modeling framework to forecast the flood risk attributes during the summer peak monsoon season (July-August) at five gauges in the Narmada River basin (NRB) of West-Central India for several lead times (0–3 months). As potential covariates, we consider climate indices such as El Niño–Southern Oscillation (ENSO), the Indian Ocean Dipole (IOD), and the Pacific Warm Pool Region (PWPR) from antecedent seasons, which have shown strong teleconnections with the Indian monsoon. We also include new indices related to the East Pacific and West Indian Ocean regions depending on the lead times. We show useful long lead skill from this modeling approach which has a strong potential to enable robust risk-based flood mitigation and adaptation strategies 3 months before flood occurrences.

1. Introduction

Climate extremes such as floods and heavy precipitation every year lead to severe infrastructural damage and human lives lost across the world (Field et al., 2012; Marcolongo et al., 2022; Tanoue et al., 2016; Wallemacq & House, 2018). Besides, they are expected to become more frequent 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. For example, India receives more than 80% of the total annual rainfall during summer monsoon (June–September) rainfall events caused by synoptic-scale cyclonic depressions (Hunt et al., 2016; Hunt & Fletcher, 2019), which lead to the occurrence of several floods during this season in different basins across the country. To effectively mitigate the negative impacts of these extreme events, we need to provide decision-makers with robust tools that they can use in the mid-and long-term planning of flood risk adaptation strategies. Most available tools consist of hydrological models to provide daily operational forecasts for short-range (1 day to a couple of weeks) or statistical models considering hydroclimatic variables from the previous season to generate seasonal mean streamflow forecasts.

Operational streamflow forecasts are implemented using physically based hydrologic or conceptual models that consider forecasts of hydrometeorological variables, such as rainfall and temperature, as their forcing (Clark et al., 2004; Tiwari et al., 2021; Werner & Yeager, 2013; Wijayarathne & Coulibaly, 2020). Other alternatives to such hydrological models are statistical-physical hybrid models (Kurian et al., 2020; Li et al., 2015; McInerney et al., 2017; Ossandón, Rajagopalan, et al., 2022), and purely statistical models (Govindaraju, 2000; Hadi & Tombul, 2018; Papacharalampous & Tyralis, 2018; Sivakumar, 2016). These modeling approaches only provide daily streamflow forecasts for short lead times (no longer than 1 or 2 weeks), and the skills decline with increasing lead times, more so for peak flows. While they are useful for short-term actions for flood impact mitigation, they do not allow for long-term planning of mitigation actions, such as months ahead of the flood season.

Models for forecasting seasonal and sub-seasonal streamflow rely on the skill of hydroclimatic variables from the previous season, such as snow cover (e.g., Koster et al., 2010; Livneh & Badger, 2020; Pagano et al., 2009; Wood et al., 2016), changes in land cover conditions (De Perez et al., 2017; Penn et al., 2020), or large-scale climate indices (Robertson & Wang, 2012; Steirou et al., 2019; Wang & Robertson, 2011; Zhao et al., 2016), among others. Among the modeling approaches are: statistical techniques based on multiple linear regression (e.g., Penn et al., 2020; Ruiz et al., 2007; Sankarasubramanian & Lall, 2003; Steinschneider & Lall, 2016) or Bayesian approaches that account for parameter uncertainty (e.g., Kwon et al., 2008, 2009; Lima & Lall, 2010; Zhao et al., 2016), physicallybased models that consider the uncertainty of initial conditions or inputs by perturbing them (Anghileri et al., 2016; Wood et al., 2016), and hybrid models, which combine hydrological models with statistical models to post-process their output and to enhance forecast skill (Bennett et al., 2016, 2017, 2021; Mendoza et al., 2014). Although this type of forecast help inform reservoir operations during the dry season at a local scale, they do not consider crucial features for estimating regional flood hazards, such as the spatial dependencies in high-flow occurrences across different catchments on a river network (Brunner et al., 2020; Ossandón, Brunner, et al., 2022).

Among the statistical approaches, Bayesian models emerge as an attractive alternative as they provide a robust estimation of streamflow extremes and parameters uncertainty via posterior distributions (Reis & Stedinger, 2005; Reza Najafi & Moradkhani, 2013; Yan & Moradkhani, 2015) and allow considering the spatial dependence structure in a process layer (Bracken et al., 2018; Renard, 2011; Renard & Lang, 2007). However, most of the applications of Bayesian frameworks are mainly for obtaining return levels from frequency analysis of extreme precipitation (e.g., S. G. Coles & Tawn, 1996; Cooley et al., 2007; Cooley & Sain, 2010; Ossandón et al., 2021; Renard, 2011) and floods (e.g., López & Francés, 2013; Najafi & Moradkhani, 2014; Steirou et al., 2019; Villarini et al., 2012; Yan & Moradkhani, 2016). Fewer studies implemented Bayesian methods for seasonal peak flow forecasts (Kwon et al., 2008, 2009; Ossandón, Brunner, et al., 2022; Steirou et al., 2022)- mostly for return-level forecasts for single sites without accounting for spatial dependencies. Only Ossandón, Brunner, et al. (2022) showed a skillful spring seasonal maximum streamflow forecast at multiple gauges in the Upper Colorado River basin (UCRB) at 0- to 2- months lead time. They used a Bayesian hierarchical model that accounts for the spatial dependence structure of the data and nonstationarity through suitable covariates from previous seasons. Most forecast skill in the UCRB came from snow accumulated during winter.

The seasonal peak forecast in rainfed basins, where we cannot rely on the skill provided by the snow, is incredibly challenging. Particularly in India, there is no record of any attempt to provide summer monsoon seasonal (i.e., June–September) peak flow forecasts and their attributes. Strong correlations between ISMR and large-scale climate variables such as El Niño–Southern Oscillation, Atlantic Multidecadal Oscillation (ENSO), and the Indian Ocean Dipole (IOD) have been documented (e.g., Curtis et al., 2001; Kumar et al., 2006; Saji et al., 1999). However the influence of ENSO (IOD) on the ISMR was found to be weakened (strengthened) in the recent decades (Hrudya et al., 2020; Krishnaswamy et al., 2015; Rajagopalan & Molnar, 2012). While continuous efforts have been made to predict the ISMR one season ahead using statistical and physical models based on these teleconnections, their skills are limited ($r \sim 0.3$–0.46; Delsole & Shukla, 2012; Jain et al., 2019; Köhn-Reich & Bürger, 2019; Rajeevan et al., 2012). Thus, the interesting questions are—do these teleconnections translate to attributes of seasonal streamflow extremes? Can they be exploited to provide long-lead forecasts in a river basin if they do?

With this motivation and need, we develop a Bayesian Hierarchical Modeling (BHM) framework, based on that proposed by Ossandón, Brunner, et al. (2022), to model and forecast two flood risk attributes—seasonal daily maximum flow and the number of events that exceed a threshold during a season (NEETM)—at a suite of gauge locations on a river network. These two attributes will help policymakers devise mitigation and adaptation strategies before the start of the season by considering early alerts that complement existing forecasts at shorter time scales to increase flood preparedness. These could include the early release of water from reservoirs to increase their flood retention capacity, devote financial resources to acquire disaster response supplies, and implementing communication strategies to enhance community disaster preparedness (Steirou et al., 2022; Zanardo et al., 2019). We apply the BHM forecasting framework to the summer peak monsoon season (July–August) at five Narmada River basin (NRB) gauges and three lead times. For the remainder of this paper, we use the term "daily monsoon maximum streamflow" to refer to the maximum daily streamflow during the summer peak monsoon season (July–August). The novelty here is that this study is one of the first efforts at a long-lead forecast of seasonal extremes in a river basin in India and using large-scale climate indices as covariates while

also accounting for spatial dependencies in streamflow on the river network and robust quantification of uncertainties in the forecasts.

2. Proposed Framework

2.1. General Model Structure

Consider that $q_i(t)$ for $t = 1, \ldots, k$ represents streamflow extremes at time point t for location i. Then, $\mathbf{q} = [\mathbf{q_1}, \ldots, \mathbf{q_m}]$ corresponds to the matrix of seasonal streamflow extremes at m gauges in a river basin over k years. To model them, we use a Bayesian hierarchical model (BHM; Cooley et al., 2007; Wikle et al., 1998) that allows for spatial dependence and nonstationarity. In addition to a data layer, which parameters are directly related to the observed data, BHM includes additional layers that allow accounting for the latent process, typically called process or latent layers. Here, in the data layer, we use a Gaussian copula with generalized extreme value (GEV) margins (S. Coles, 2001; He et al., 2015; Katz, 2013) to model the joint distribution of streamflow at m gauges in each year as follows:

$$F_1(q_1(t)), \ldots, F_m(q_m(t)) \sim C_g(\mathbf{\Sigma}) \tag{1}$$

$$q_i(t) \sim GEV(\mu_i(t), \sigma_i(t), \xi_i), \qquad i = 1, \ldots, m \tag{2}$$

where F_i is the CDF of the GEV distribution for the gauge i, C_g is an m-dimensional Gaussian copula with dependence matrix $\mathbf{\Sigma}$, and $\mu_i(t) \in (-\infty, \infty)$, $\sigma_i(t) > 0$ and $\xi_i(t) \in (-\infty, \infty)$ are the GEV parameters (location, scale, and shape). To account for the nonstationarity in the second layer (first process layer), we consider that the GEV parameters can vary over time. We model the first two GEV parameters (location and scale) as linear functions of time-dependent large-scale climate variables and regional mean variables from the previous season while the shape parameter is considered stationary:

$$\mu_i(t) = \alpha_{\mu 0_i} + \sum_{j=1}^{n} \alpha_{\mu j_i} z_j(t), \qquad i = 1, \ldots, m \tag{3}$$

$$\log(\sigma_i(t)) = \alpha_{\sigma 0_i} + \sum_{j=1}^{n} \alpha_{\sigma j_i} z_j(t), \qquad i = 1, \ldots, m \tag{4}$$

$$\xi_i(t) = \alpha_{\xi_i}, \qquad i = 1, \ldots, m \tag{5}$$

where $\alpha_{\mu j_i}$, $\alpha_{\sigma j_i}$, and α_{ξ_i} are the regression coefficients for the predictor j and gauge i, and $z_j(t)$ is the covariate j at the time t. $\log(\sigma))$ is used to ensure positive scale parameters. The validity of the nonstationary location and scale parameters can be checked through the significance of their slope coefficients' posterior PDFs ($\alpha_{\mu j_i} \neq 0$, and $\alpha_{\sigma j_i} \neq 0$ for $j > 0$). Here, we model the shape parameter for each catchment individually but consider it stationary in time and restricted to a specific range of potential values (-0.5 to 0.5; Bracken et al., 2016; Cooley et al., 2007). Covariates will be discussed in Section 3.2.

Although several copulas are available (Hochrainer-Stigler, 2020), this study focuses on Gaussian copulas because of their easy implementation in a Bayesian and high-dimensional framework. The copula dependence matrix, $\mathbf{\Sigma}$, is a symmetric positive definite matrix that captures the strength of dependence between all gauge pairs by the correlation coefficient. The element ρ_{ij} of $\mathbf{\Sigma}$ quantifies the dependence between gauges i and j, and its values can vary between -1 and 1, as follows:

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1(m-1)} & \rho_{1m} \\ \rho_{21} & 1 & \ddots & \vdots & \rho_{2m} \\ \rho_{31} & \rho_{32} & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 1 & \rho_{(m-1)m} \\ \rho_{m1} & \rho_{m2} & \cdots & \rho_{m(m-1)} & 1 \end{bmatrix} \tag{6}$$
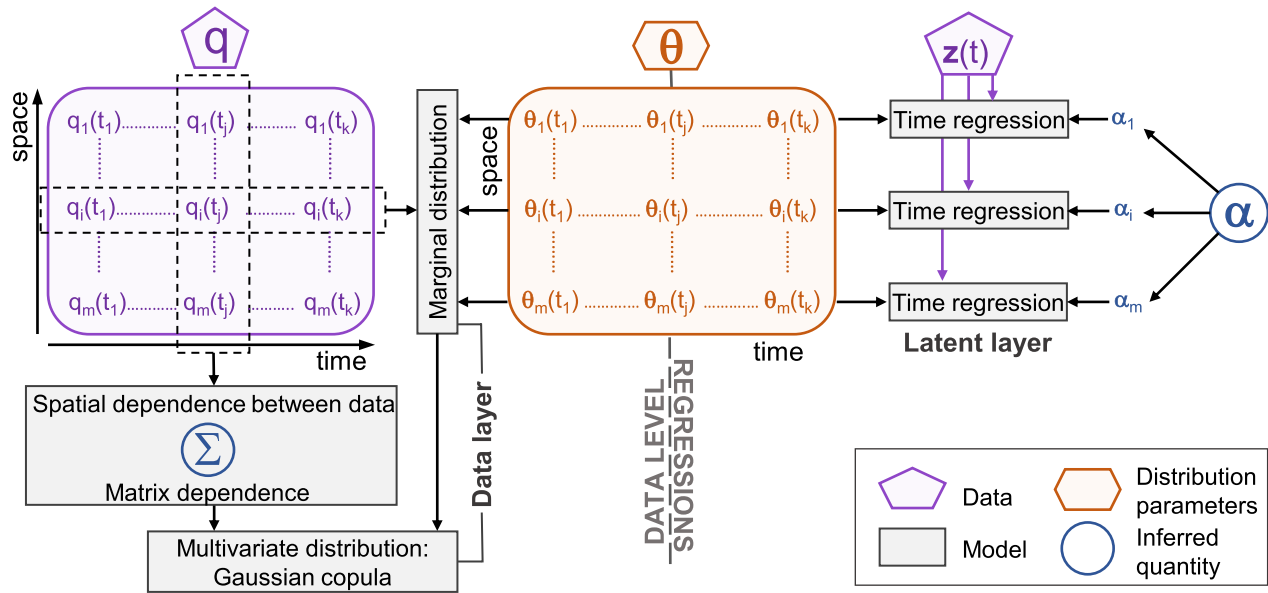
Figure 1. Conceptual schematic of Bayesian Hierarchical Model. Purple boxes denote the data, the orange box the marginal distribution parameters, gray boxes the models, and blue circles the inferred quantities. Model boxes correspond to the data layer (Gaussian copula and marginal distributions) and the latent layer (time regressions of marginal distribution parameters). $\theta_i(t_j)$ denotes the vector of the marginal distribution parameters.

Note that since Σ is symmetric, only $m(m-1)/2$ dependence parameters must be fitted (values in the lower or upper triangle of Σ). The Gaussian copula only assumes linear correlation after quantile transformation of the marginals with the inverse normal CDF. This assumption does not impose a linear correlation structure on the marginal distributions, meaning that non-linear dependence between variables can be captured at the data level (Bracken et al., 2018). The Gaussian copula constructs the joint cumulative distribution function (CDF) of $\mathbf{q}(t) = [q_i(t)]_{i=1}^m$ as

$$F(\mathbf{q}(t)) = \Phi_\Sigma(\mathbf{u}(t)) \tag{7}$$

where $\Phi_\Sigma(\cdot)$ is the m-dimensional multivariate normal distribution joint CDF with dependence matrix Σ, $\mathbf{u}(t) = [u_i(t)]_{i=1}^m$, $u_i(t) = \Phi^{-1}(F_{it}[q_i(t)])$ with Φ being the CDF of the standard normal distribution, and $F_{it}(\cdot)$ is the marginal GEV CDF or empirical CDF for the streamflow gauge i at time t. The corresponding joint PDF at the time t is

$$f(\mathbf{q}(t)) = \frac{\prod_{i=1}^m f_i(q_i(t))}{\prod_{i=1}^m \psi(u_i(t))}\Psi_\Sigma(\mathbf{u}(t)) \tag{8}$$

where ψ corresponds to the standard normal PDF, f_i to the marginal GEV PDF at site i, and Ψ_Σ to the joint PDF of an m-dimensional multivariate normal distribution (Bracken et al., 2018; Ghosh & Mallick, 2011).

There are two main approaches for estimating the unknown parameters of the conditional copula (Hochrainer-Stigler, 2020). The first one, known as pseudo-observations fitting, estimates the copula parameters without assuming specific parametric distribution functions of the marginals, and pseudo-observations are used instead. The second approach is called inference functions for margins (IFMs). This approach estimates the marginal distribution parameters in the first step, and the copula parameters are estimated in the second step (third process layer). Here, we considered the IFMs approach since it allows for joint parameter estimation in a Bayesian framework.

Figure 1 displays a conceptual sketch of the BHM, which shows the data layer (Gaussian copula and marginal distributions) and the process layer (time dependence of marginal GEV distribution parameters).

Note that this framework can be easily adapted to model another variable of interest by changing the marginal distribution in Equation 2 and modifying Equations 3–5 accordingly. For example, to model the second flood risk

attribute, NEETM, we replace the GEV distribution in Equation 2 with a Poisson distribution ($Poisson(\lambda_i(t))$), which is appropriate to model a discrete number of events, such as the case here. The rate parameter is non-stationary ($\lambda_i(t)$) and is therefore modeled as a function of temporal covariates as follows:

$$\lambda_i(t) = \alpha_{\lambda 0_i} + \sum_{j=1}^{n} \alpha_{\lambda j_i} z_j(t), \qquad i = 1, \dots, m \tag{9}$$

where $\alpha_{\lambda j_i}$ is the regression coefficient for the predictor j and gauge i, and $z_j(t)$ is the covariate j at the time t. We omit a copula in the data layer since it is unsuitable for modeling discrete data like NEETM.

2.2. Likelihood and Priors

Let $\mathbf{z} = [\mathbf{z}_1, \cdots, \mathbf{z}_n]$ be a matrix of all covariates and $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_\mu, \boldsymbol{\alpha}_\sigma, \boldsymbol{\alpha}_\xi]$ a matrix of the GEV regression coefficients. By Bayes' rule, the posterior is

$$p(\boldsymbol{\alpha}, \boldsymbol{\Sigma}|\mathbf{q}, \mathbf{z}) \propto p(\mathbf{q}|\mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\Sigma})p(\boldsymbol{\Sigma})p(\boldsymbol{\alpha}) \tag{10}$$

where the first term on the right-hand side is the copula likelihood which, assuming temporal independence of the observations, can be written as

$$p(\mathbf{q}|\mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\Sigma}) = \prod_{t=1}^{k} p(\mathbf{q}(t)|\mathbf{z}(t), \boldsymbol{\alpha}, \boldsymbol{\Sigma}) \tag{11}$$

where $p(\mathbf{q}(t)|\mathbf{z}(t), \boldsymbol{\alpha}, \boldsymbol{\Sigma})$ corresponds to the likelihood displayed in Equation 8. The terms $p(\boldsymbol{\alpha})$ and $p(\boldsymbol{\Sigma})$ from Equation 10 represents the priors of the GEV regression coefficients and Gaussian copula dependence matrix, which are defined as

$$p(\boldsymbol{\alpha}) = \prod_{i=1}^{m} \left(\prod_{j=0}^{n} p(\alpha_{\mu j_i}) p(\alpha_{\sigma j_i}) \right) p(\alpha_{\xi_i}) \tag{12}$$

$$p(\alpha_{\mu j_i}) = N(0, A_{j_i}) \tag{13}$$

$$p(\alpha_{\sigma j_i}) = N(0, B_{j_i}) \tag{14}$$

$$p(\alpha_{\xi_i}) = N(0, C_i) \tag{15}$$

$$p(\boldsymbol{\Sigma}) = LKJcorr(\eta) \tag{16}$$

Where A_{j_i}, B_{j_i}, and C_i correspond to the prior variances for the location $(\alpha_{\mu j_i})$, scale $(\alpha_{\sigma j_i})$, and shape (α_{ξ_i}) GEV regression coefficients, and $LKJcorr(\eta)$ is the LKJ correlation matrix density with η being the shape parameter (Lewandowski et al., 2009). The specific values of the priors A_{j_i}, B_{j_i}, C_i, and η will be described in Section 3.3.

2.3. Estimation of Ensembles of Seasonal Maximum Streamflow Forecasting

The posterior predictive distributions can be approximated using ensembles from Monte Carlo. Thus, the predictive posterior distribution of monsoon daily maximum streamflow (ensembles) for the m streamflow gauges are obtained via Monte Carlo from GEV distribution using the posterior distribution (N samples) of GEV regression coefficients, $\boldsymbol{\alpha}_\mu$, $\boldsymbol{\alpha}_\sigma$, and $\boldsymbol{\alpha}_\xi$, and Gaussian copula dependence matrix, $\boldsymbol{\Sigma}$. The steps for this procedure are as follows.

1. If the posterior PDFs of the slope coefficients of the GEV parameters are found to be significant, compute the GEV parameters ($\mu_i(t)$, $\log(\sigma_i(t))$) for each year and gauge using $\boldsymbol{\alpha}_{\mu_i}$, $\boldsymbol{\alpha}_{\sigma_i}$, and covariates, $\mathbf{z}(t)$, using Equations 3 and 4.
2. Then, simulate non-exceedance probabilities, ϕ_{it}, for the m streamflow gauges from a Gaussian copula with a dependence matrix, $\boldsymbol{\Sigma}$.
3. After that, compute the daily monsoon maximum streamflow for each streamflow gauge i at time t using the expression for the quantile function of the GEV.
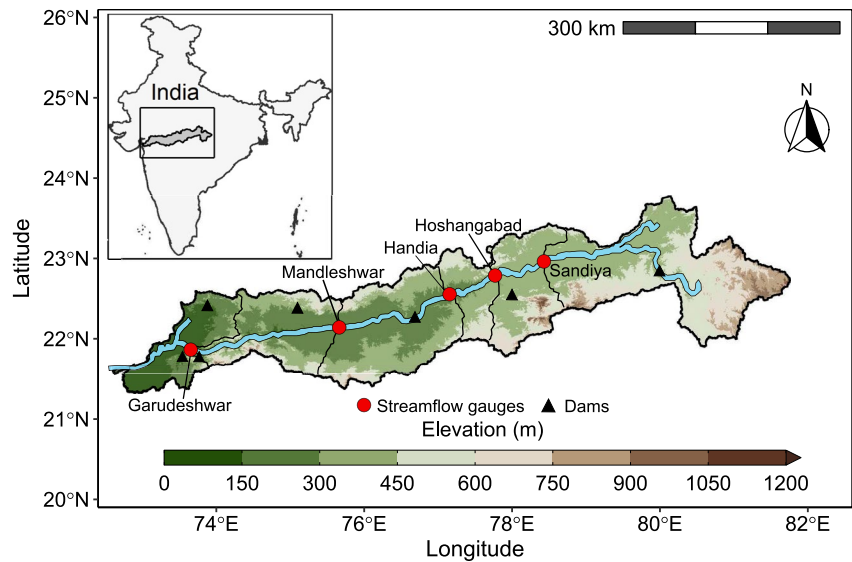
Figure 2. Map of the Narmada River basin, including five sub-basin outlets (red circles), the 150-m elevation bands, and some of the major dams in the basin: Bargi, Tawa, Indirasagar, Jobat, and Sardar Sarovar (from upstream to downstream direction).

$$q_i(t) = \mu_i(t) + \frac{\sigma_i(t)}{\xi_i}\left[(-\log(\phi_{it}))^{-\xi_i} - 1\right], \qquad i = 1, \dots, m \tag{17}$$

4. Repeat steps 2–3 for each year of the record.

This procedure is repeated N times, and the N samples form an approximation of the posterior predictive distribution. A similar approach is followed for simulating NEETM from Poisson distribution without considering a Gaussian copula.

2.4. Estimation of Ensembles of NEETM

We adapt the same procedure for simulating ensembles of NEETM at streamflow gauge i using the posterior distribution (N samples) of Poisson regression coefficients, $\boldsymbol{\alpha}_{\lambda_i}$, by replacing the GEV distribution in Equation 2 with a Poisson distribution and excluding a copula in the data layer for NEETM modeling. The steps are as follows.

1. If the posterior PDFs of the slope coefficients $\left(\boldsymbol{\alpha}_{\lambda_i}\right)$ for the Poisson parameter ($\lambda_i(t)$) are found to be significant, calculate $\lambda_i(t)$ for each year using $\boldsymbol{\alpha}_{\lambda_i}$, and covariates, $\boldsymbol{z}(t)$ according to Equation 9. If not, set $\lambda_i(t)$ equal to $\alpha_{\lambda 0_i}$.
2. Simulate the non-exceedance probability, F_{it}, for each year using the Poisson margin.
3. Compute the NEETM for each year by employing the Poisson quantile function.

Repeat this procedure N times for every gauge.

3. Application to the Narmada River Basin

We demonstrate the framework proposed in the previous section by its application to forecasting monsoon daily maximum streamflow at five gauges in the Narmada River basin (NRB) from 0- to 3-month lead time (Figure 2). The NRB is a narrow and elongated basin that stretches in the East-West direction, and it rises in the Amarkantak hills of central India. The NRB, with a catchment area of approximately 97,882 km², is the largest river, draining into the Arabian Sea in west India. The mean annual rainfall over the basin is 1,120 mm (1951–2018), with most of it occurring during the summer monsoon season (June–September), which causes flood events mainly during July–August in the basin. Therefore, providing skillful forecasting of peak monsoon season maximum streamflow several months in advance would be helpful for the implementation of risk mitigation strategies.

Table 1
Basic Data Corresponding to the Streamflow Gauges in the Narmada River Basin Considered in This Study

Gauge	Area (km²)	Elevation (m)	Mean streamflow (m³ s⁻¹)	Mean seasonal streamflow (m³ s⁻¹)	Max seasonal streamflow (m³ s⁻¹)
Garudeshwar	86,494	12	886	2698	52,000
Mandleshwar	71,739	141	997	2938	46,398
Handia	51,115	260	785	2487	31,880
Hoshangabad	44,487	292	676	2092	28,600
Sandiya	32,495	301	481	1410	19,700

3.1. Data

Daily peak monsoon season(July and August) streamflow data at five gauges in the Narmada River basin—Sandiya, Hoshangabad, Handia, Mandleshwar, and Garudeshwar—were obtained from the India Water Resource Information System (IWRIS) (Figure 2 and Table 1) for the period 1978–2018. The five gauges encompass drainage areas ranging from 32,495 to 86,494 km² and elevations varying between 12 and 301 m above sea level (m.a.s.l.). The mean streamflow, mean seasonal streamflow (July–August), and maximum seasonal streamflow (July-August) span from 301 to 997 m³ s⁻¹, 1,410 to 2,938 m³ s⁻¹, and 19,700 to 52,000 m³ s⁻¹, respectively. Reservoirs inflows and releases were not considered since they are unavailable or have long periods of missing values. We computed the daily monsoon maximum streamflow for each year at each streamflow. Garudeshwar had no monsoon daily streamflow record in 1995, so the maximum flow was set as a missing value.

To justify the inclusion of a copula in the BHM to capture the spatial dependence structure of the data, we assessed the spatial dependence of daily monsoon maximum streamflow for the five gauges using the Spearman correlation coefficient for each pair of stations (Figure S1 in Supporting Information S1). The correlation between gauges is positive, higher than 0.4, and significant at a 95% confidence level for all station pairs.

Regarding meteorological variables, we used 0.25° gridded daily precipitation data from the India Meteorology Department (IMD) from 1978 to 2018. The daily gridded precipitation data set was obtained by interpolating observations from meteorological stations across India using an inverse distance weighting scheme (Pai et al., 2014). Gridded global SST anomalies from 1856 to 2021 with 5° spatial resolution were obtained from the International Research Institute (IRI) for Climate and Society at Columbia University (Kaplan et al., 1998; Parker et al., 1994; Reynolds & Smith, 1994). Atmospheric pressure, humidity, and horizontal and vertical wind fields from 1949 to 2021 were obtained from the National Centers for Environmental Prediction (NCEP)-National Centers for Atmospheric Research (NCAR) reanalysis data set (Kalnay et al., 1996; Kistler et al., 2001) with 5° spatial resolution.

In the case of large-scale climate indices, we obtained time series of Extreme Eastern Tropical Pacific SST (Niño 1+2), East Central Tropical Pacific SST (Niño 3.4), Pacific Warm Pool Region (PWPR), and Indian Ocean Dipole (IOD) anomalies from the National Oceanic and Atmospheric Administration (NOAA; https://psl.noaa.gov/data/climateindices/list/).

3.2. Covariates Selection

First, to perform the climate diagnostics, we conducted a principal component analysis across the detrended monsoon daily maximum streamflow of the five gauges considered in this study. The first two leading principal components (PC1 and PC2) explained most data variance (72% and 20%, respectively). We considered the detrended streamflow for the climate analysis since streamflows are affected by the reservoir operations, which weakens the climate signature (a negative temporal trend for most gauges, Kendall's τ coefficient ranging between 0.13 and 0.44). To understand the physical mechanism behind these extremes, we obtained the composite maps of integrated water vapor transport (IVT; Gao et al., 2021; Neiman et al., 2008) and vector winds at 850 mb anomalies for the top five wettest years and driest years (Figure S2 in Supporting Information S1). A positive anomalous IVT in and around the basin with winds from the Arabian sea is observed for wet years. At the same time, the IVT is anomalously low around the basin, and the winds are from the Bay of Bengal and to the south of the basin for
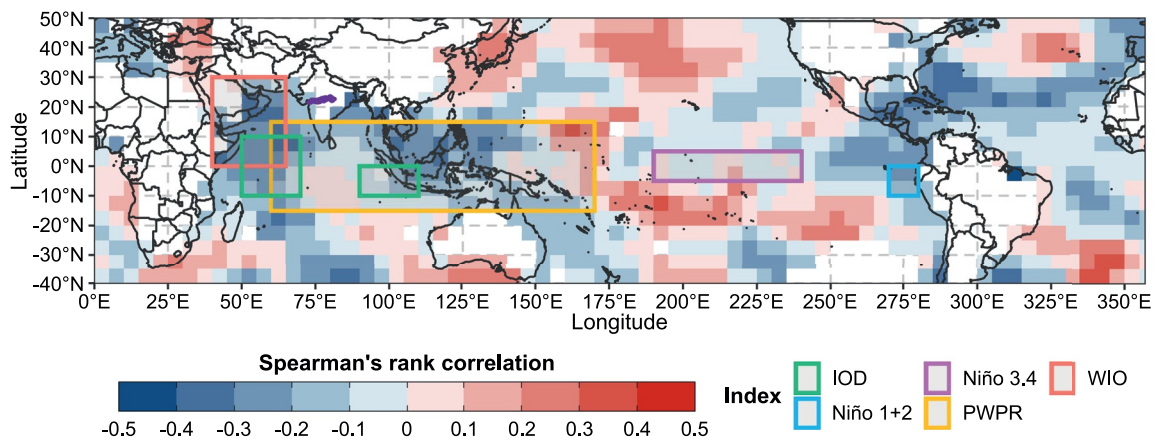
Figure 3. Correlation between PC1 of detrended monsoon (July–August) daily maximum streamflow at the Narmada River basin and gridded SST anomalies for a 0-month lead time (June) for 1978–2018.

the dry years. These moisture transport and rainfall patterns are consistent with the large-scale climate signatures (e.g., Chevuturi et al., 2021; Nanditha et al., 2022) and, consequently, support the ENSO teleconnection.

As mentioned, strong teleconnections between the summer monsoon rainfall in India and large-scale climate variables such as ENSO or IOD have been reported by other authors (e.g., Curtis et al., 2001; Kumar et al., 2006; Saji et al., 1999). These teleconnections support the initial inclusion of these large-scale climate indices as covariates. However, the influence of ENSO (IOD) on the ISMR was found to be weakened (strengthened) in recent decades (Hrudya et al., 2020; Krishnaswamy et al., 2015; Rajagopalan & Molnar, 2012). Figure 3 shows the correlation between PC1 of detrended monsoon (July–August) daily maximum streamflow at the NRB and gridded SST anomalies for a 0-month lead time (June) for the 1978–2018 period and the regions corresponding to the large-scale climate indices considered in this study. Areas of the large-scale climate indices do not show a high correlation for the July–August period. Therefore, we consider the West Indian Ocean (WIO, 42.5–62.5E and 2.5–27.5N) index that corresponds to the mean SST anomaly of the highly correlated region (the pink box in Figure 3).

We assessed the strength of the relationship between the monsoon daily maximum streamflow (July–August) and potential covariates for 0-month lead time (June) by computing Spearman's rank correlation coefficient, shown in Figure 4. WIO and PWPR (Figures 4a and 4b) show a strong negative correlation with the daily monsoon maximum streamflow. The Niño 3.4 index (Figure 4c) shows a weak and not significant correlation with the daily monsoon maximum streamflow at all the gauges. The Niño 1+2 and IOD climate indices (Figures 4d and 4e) show a significant correlation at some gauges, but it is weaker than that for WIO and PWPR.

We tested the following covariates for modeling the temporal nonstationarity of the GEV parameters (see Equations 3 and 4) for the period 1978–2018: PWPR, Niño 3.4, Niño 1+2, IOD climate indices, and the new indices defined as the mean SST anomaly of the highly correlated region from March, April, May, or June depending on the lead time (0-, 1-, 2- or 3-month lead time). The covariates use all information from the month before forecast issuance. As indicated earlier, we focus on forecasting monsoon daily maximum streamflow at lead times of 0-, 1-, 2-, and 3 months, corresponding the release of the forecast on July 1, June 1, May 1, and April 1, respectively (Figure 5).

Table 2 shows the domain of the new indices considered for each lead time, which are called WIO, the Extended West Indian Ocean (EWIO), the East Pacific Cold Tongue (EPCT), and the Extended East Pacific Cold Tongue (EEPCT) for 0-, 1-, 2-, and 3-month lead time. Note that the new indices domain varies with the lead time, moving eastward as the lead time increases. To define the new index domain for the other lead times, we followed the same procedure as a 0-month lead—that is, we identified the region with high correlations and computed the index by averaging the SST anomaly over the domain (Figures S3 and S5 in Supporting Information S1). We use the changing domain since we are interested in a predictor that gives skillful predictions. However, we note that the regions at all the lead times correspond to the ENSO phenomena and that it influences the flow variability. If we picked the standard ENSO index for longer lead times, the correlations are weak with the flow; hence, we
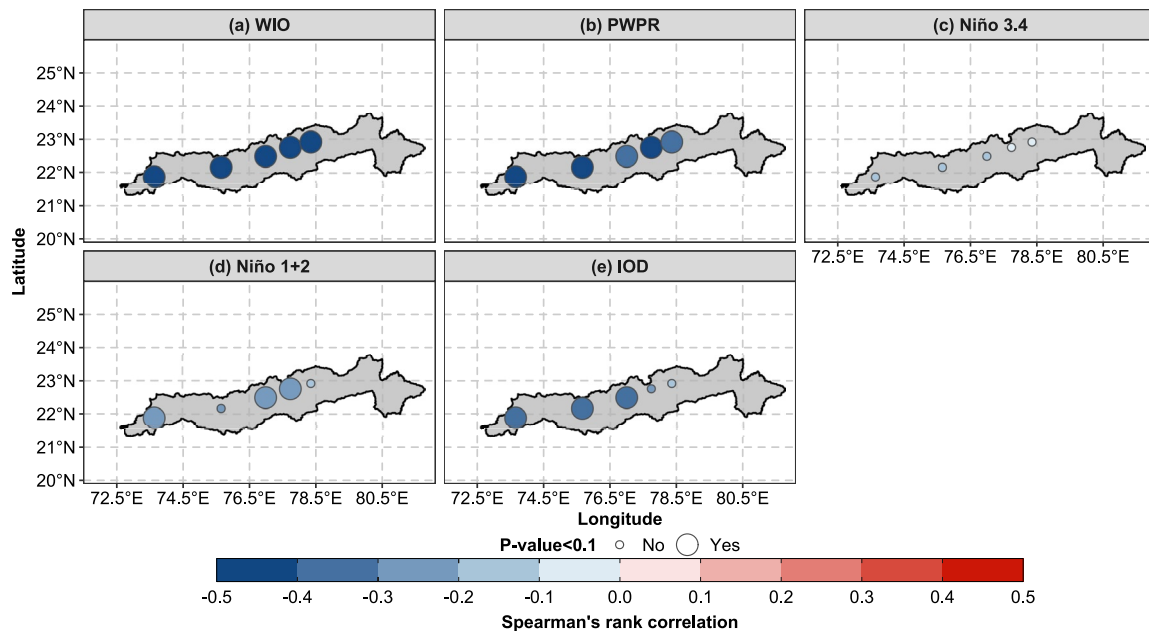
Figure 4. Spearman's rank correlation coefficient between daily monsoon (July–August) maximum streamflow and covariates for a 0-month lead time (June). (a) The West Indian Ocean (WIO) index, (b) Pacific Warmpool Region (PWPR). (c) Niño 3.4, (d) Niño 1+2, and (e) Indian Ocean Dipole (IOD). Big circles indicate that Spearman's rank correlation is significant (p value < 0.05).

shifted the SST region (new indices domain) lead time. Other authors have considered this approach for finding skillful covariates for forecasting streamflow in the western US (e.g., Grantz et al., 2005; Regonda et al., 2006, being the early proponents).

The reader can find Spearman's rank correlation coefficient plots between monsoon daily maximum streamflow and covariates for 1- to 3-month lead time in Supporting Information S1 (Figures S6–S8).

3.3. Model Structure for the NRB

The specific structure of the BHM for the NRB incorporated the covariates described in Section 3.2 for the first process layer. We modeled the location parameter of the GEV at each gauge as non-stationary, and the scale parameter of the GEV was kept stationary for all gauges since an initial run of the BHM considering it as non-stationary showed that their regression coefficients ($\alpha_{\sigma_j} = 0$ for $j > 0$ in Equation 4) were not significant (posterior PDFs contain zero in their domain at a 90% confidence level). For the normal priors of α_{μ_j}, α_{σ_0}, and α_{ξ_0}, we considered a variance of $A_{j_i} = 10$, $B_{0_i} = 10$, and $C_i = 1$, respectively. We considered $\eta = 2$ for the LKJ correlation matrix density prior. They correspond to weakly informative priors.
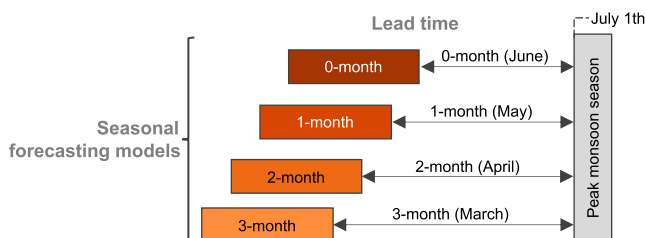

Figure 5. Schematic of non-stationary models considered for four different lead times. The boxes denote, from the darker to lightest orange color: the model for a 0-month lead time (forecasts released on July 1), a 1-month lead time (forecasts released on June 1), a 2-month lead time (forecasts released on May 1), and a 3-month lead time (forecasts released on April 1). The same color scheme will be considered for Section 4.

3.4. Implementation and Model Fitting

Candidate BHMs were fit using the program STAN (Stan Development Team, 2014), its R extension (Stan Development Team, 2020), and the priors assigned in Section 3.3. STAN uses a Markov chain Monte Carlo (MCMC) simulation method that considers a No-U-Turn Sampler (NUTS; Hoffman & Gelman, 2014) to simulate the posterior probability distribution of the GEV regression coefficients and the Gaussian copula dependence matrix. We ran three parallel chains with different initial values, and each chain has a length of 12,000 simulations (iterations) with a burn-in size of 6,000 to ensure convergence. To reduce the sample dependence (autocorrelation), we chose a thinning factor of 6. The convergence of the posterior distribution of each regression coefficient was checked using the scale reduction factor, $\hat{R}$, proposed by (Gelman & Rubin, 1992)—$\hat{R}$ values lower than 1.1 for all the

Table 2
Domain of the New Indices Considered for Different Lead Times

Lead time	Index	Domain	
		Longitude (E)	Latitude (N)
0-month	WIO	42.5, 62.5	2.5, 27.5
1-month	EWIO	42.5, 82.5	−32.5, 27.5
2-month	EPCT	247.5, 277.5	−2.5, 7.5
3-month	EEPCT	252.5, 282.5	2.5, 17.5

regression coefficients suggest model convergence. Consequently, the posterior distributions of the GEV regression coefficients, the Gaussian copula dependence matrix, and the predictive posterior distributions of monsoon daily maximum streamflow consisted of 3,000 ensembles. The predictive posterior distributions of maximum spring streamflow (ensembles) for all years were estimated according to Section 2.3. The best BHM was selected for each lead time using the Watanabe-Akaike information criterion (AIC; Watanabe, 2010). This metric can be viewed as an improvement on the deviance information criterion (DIC; Spiegelhalter et al., 2002) because it is fully Bayesian (uses the entire posterior distribution) and does not produce negative estimates of a model's effective number of parameters (Vehtari et al., 2017). For each candidate BHM, we computed the WAIC using the R package loo (Vehtari et al., 2020) and selected the model with the lowest WAIC value. In addition, the significance of the GEV regression coefficients was checked for each lead time (posterior PDFs do not contain zero in the 90% credible interval). A similar procedure was followed to fit the candidate BHMs for NEETM, but omitting the Gaussian copula and estimating the NEETM ensembles according to Section 2.4.

3.5. Model Cross-Validation and Verification Metrics

To test the out-of-sample predictability of the model, we performed the leave-one-year-out cross-validation by dropping 1 year from the record (1978–2018), and the BHM was fitted using the remaining (training) years. Then, the fitted model is applied to estimate for the dropped year. We repeated the cross-validation procedure 41 times.

To assess the at-site performance of the proposed BHM, we considered deterministic metrics—that is, the Pearson correlation coefficient (R) and the relative bias (BIAS)—and probabilistic verification measures, including the continuous ranked probability skill score (CRPSS) and PIT plot (Wang et al., 2009). Also, to assess the joint performance of the proposed BHM, we computed the energy skill score (ESS). The energy score (ES) assesses the probabilistic forecasts of a multivariate quantity (Gneiting et al., 2007, 2008) compared to a reference forecast model. For both CRPSS and ESS, we considered climatology as the reference model. A detailed description of the verification methods is provided the Supporting Information S1.

4. Results

4.1. Selection of the Best Model for Each Lead Time

We calibrated different candidate BHMs for each lead time from 1978 to 2018. Then, based on the lowest value of the Watanabe-Akaike information criterion (WAIC; Watanabe, 2010), we selected the best BHM. The WAIC values for the best model for each lead time are displayed in Table 3. The best BHM considered the new indices (i.e., WIO, EWIO, EPCT, and EEPCT) as the covariate for all lead times. However, for longer lead times of 2 and 3 months, the BHM includes the PWPR as an additional covariate. The inclusion of an additional covariate could be due to the decrease in the predictive skill of the new indices at longer lead times; thus, the model includes the traditional large-scale climate indices to boost the predictive skill. The decline in predictive skill as lead time increases is apparent when comparing the WAIC values with the WAIC of the stationary BHM.

Table 3
WAIC Values for the Best Model for Each Lead Time

Lead time	Model	Covariates	WAIC
0-month	Non-stationary	WIO	249.6
1-month	Non-stationary	EWIO	252.2
2-month	Non-stationary	EPCT, PWPR	264.3
3-month	Non-stationary	EEPCT, PWPR	267.8
–	Stationary	–	271.2

Note. The same covariates for the location parameter are considered at all gauges for each model.

4.2. Cross-Validation

Figure 6 shows cross-validated daily monsoon maximum streamflow forecast ensembles from the best BHM at the Handia gauge for different lead times. For 0- and 1-month lead times (Figures 6a and 6b), all the observed values are captured by the ensemble spread (95% credible interval), most of them fall inside or close to the interquartile range of the ensembles, and the ensemble's median captures some of the observed values. For 2- and 3-month lead times (Figures 6c and 6d), ensembles show similar features overall than for
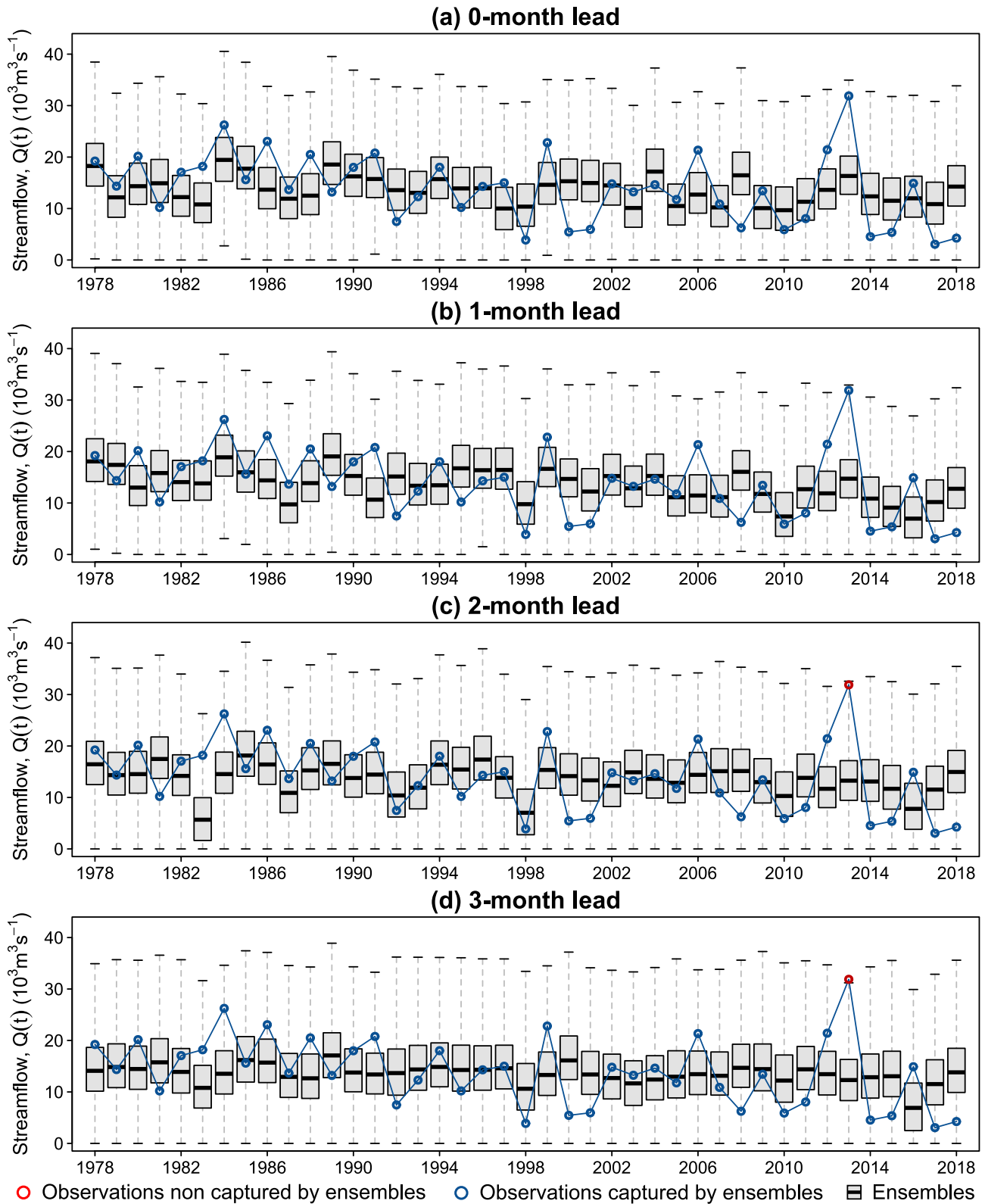
Figure 6. Time series of cross-validated daily monsoon maximum streamflow forecast ensembles from the best BHM at the Handia gauge for (a) 0-month lead time, (b) 1- month lead time, (c) 2-month lead time, (d) 3-month lead time. Blue and red points indicate when observed values are captured (or not) by the ensemble's variability. Whiskers denote the 95% credible intervals, boxes the interquartile range, and horizontal lines inside the boxes the median. Outliers are not displayed.
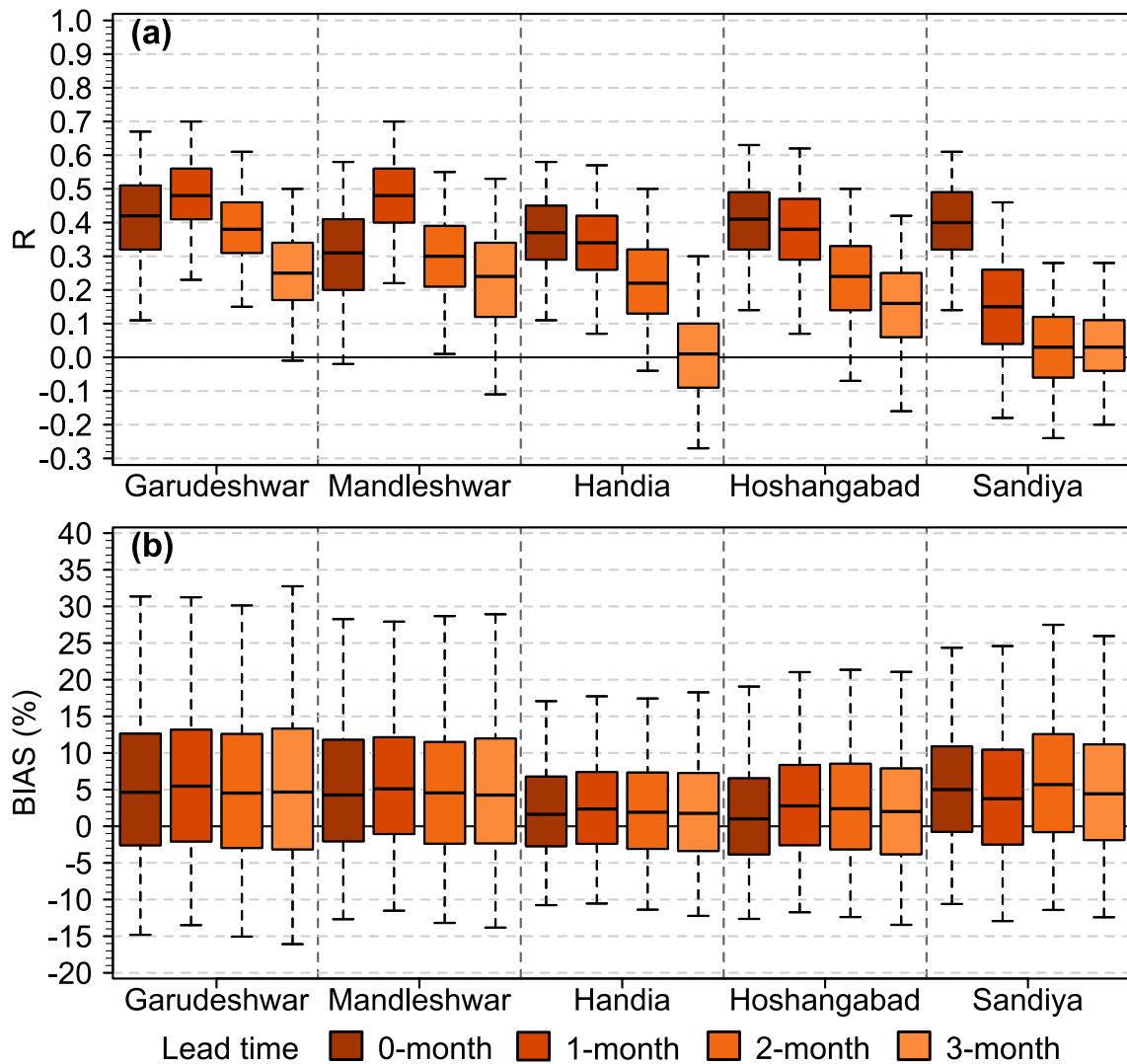
Figure 7. Boxplots of (a) Pearson correlation coefficient (*R*) and (b) BIAS at the five gauges of the NRB for different lead times (orange color scheme) from the cross-validation. The bold horizontal line in the boxplot is the median, the box of a boxplot shows the interquartile range (Q25–Q75), and the whiskers represent the sample 95%. Each boxplot comprises results from 3,000 sample bootstraps.

0- and 1-month lead except for 2013 (the highest flow in the record), where the ensemble spread does not capture the observed flow value. Also, ensembles have less temporal variability for a 3-month lead time (Figure 6d), indicating that it is getting closer to a stationary model or climatology (same ensemble forecast every year). This feature can be attributed to a reduction in the skill of the covariates as lead time increases and, as a result, an increase in the corresponding WAIC values (Table 3). Similar performance is observed for the other gauges (Figures S9–S12 in Supporting Information S1).

4.3. Model Performance at the Site for Different Lead Times

Figure 7 displays boxplots with deterministic performance metrics (*R* and BIAS) of BHM for different lead times. In the case of correlation (Figure 7a), the correlation decreases as the lead time increases, except at Garudeshwar and Mandleshwar, with median *R* values greater than 0.35 at all gauges and most 95% of their distribution values above 0 for a 0-month lead time. For the first two gauges (Garudeshwar and Mandleshwar), *R* increases from 0- to 1-month lead with median values and 95% intervals above 0.48 and 0.2, respectively. *R* distributions for 0 and 2-month leads are similar. Handia and Hoshangabad show similar correlations up to 1-month leads and not a high reduction for 2-month leads. Sandiya gauge presents the lowest performance in terms of correlation. A
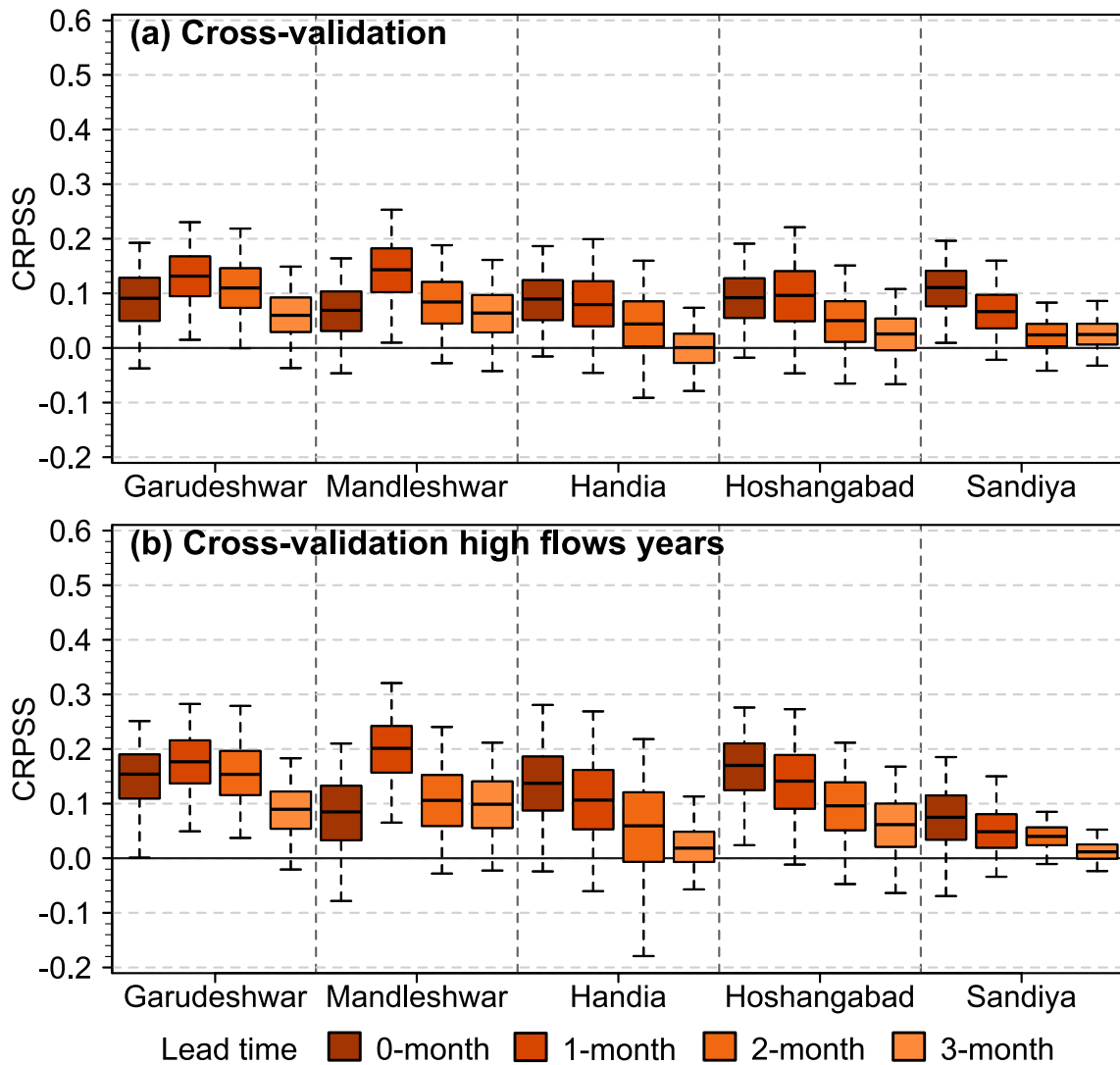
Figure 8. Boxplots of CRPSS at the five gauges of the Narmada River basin for different lead times (orange color scheme) from the cross-validation for (a) the entire period (1978–2018) and (b) only high flow years (higher than the 60th quantile). The bold horizontal line in the boxplot is the median, the box of a boxplot shows the interquartile range (Q25–Q75), and the whiskers represent the sample 95%. Each boxplot comprises results from 3,000 sample bootstraps. Climatology is considered the reference model.

substantial R reduction for 3-month lead times is reported for all the gauges. The BIAS reveals a good performance (Figure 7a) with median values below 7% and their 95% intervals width ranging between 29% and 49% at all the gauges of the NRB and all lead times, which highlights the ability of BHM to provide low BIAS (Ossandón, Nanditha, et al., 2022). Handia gauge shows the lowest median values and variability (95% intervals width lower than 31%) of the BIAS for all the lead times.

Figure 8 shows the CRPSS distribution of BHM for different lead times from the cross-validation for the entire period (1978–2018) and only years with high observed maximum flows. For whole period (Figure 8a), Garudeshwar and Mandleshwar gauges show an increase in skill for a 1-month lead compared to a 0-month lead, with mean values above 0.12 and almost 75% interval width higher than 0.1. Skill for 0- and 2-month leads are similar but slightly better for a 2-month lead with median values close to 0.1. Handia and Hoshangabad gauges present similar skill up to 1-month leads in median values (0.09–0.1) but an increase in the 95% interval width for 1-month leads. Even though there is a reduction in skill for a 2-month lead, it is still better than climatology (75% interval width above 0). In the case of the headwater gauge (Sandiya), it is seen that the skill decreases as the lead time increases (significant reduction after a 1-month lead), indicating a decrease in model skill with increasing lead
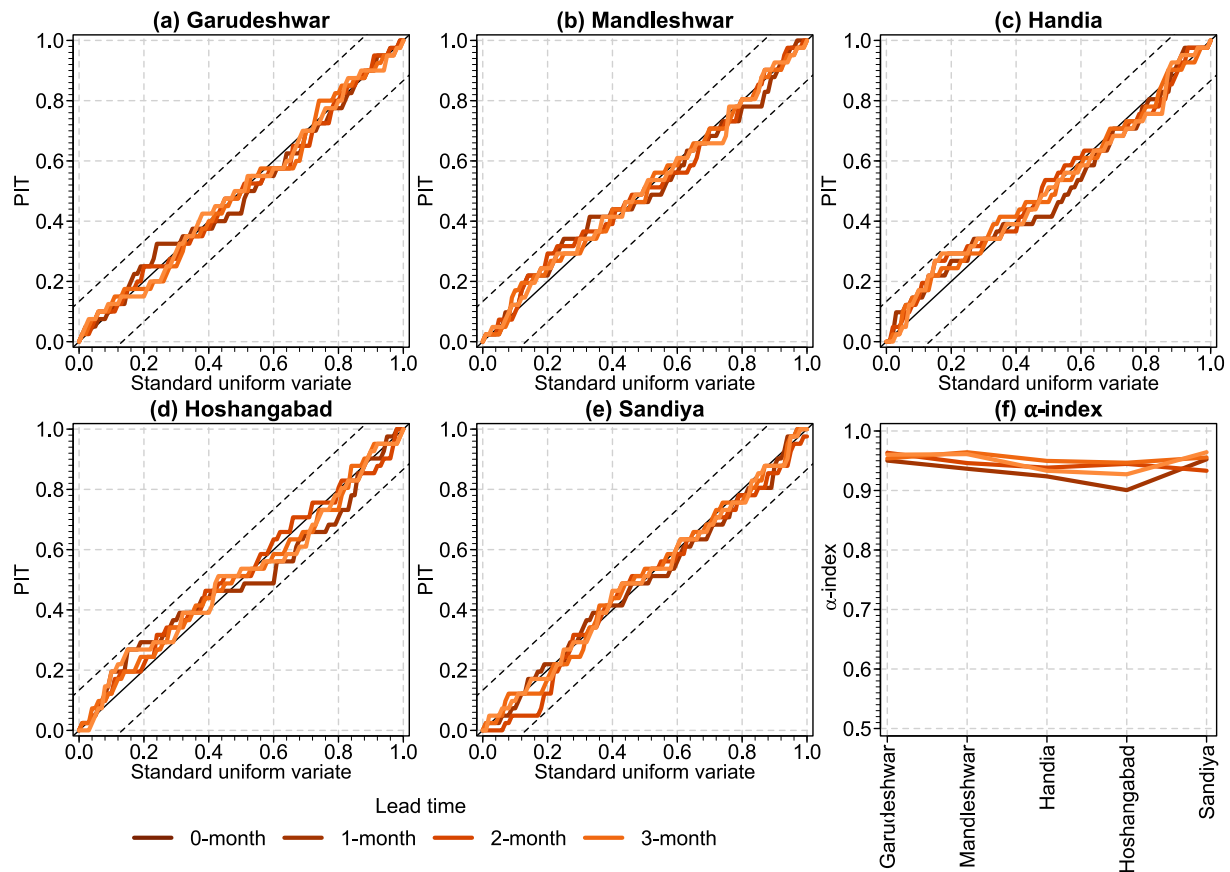
Figure 9. PIT plots for daily monsoon maximum streamflow forecast ensembles from the best BHM for different lead times at (a) Garudeshwar, (b) Mandleshwar, (c) Handia, (d) Hoshangabad, and (e) Sandiya gauges. Panel (f) displays the alpha index as a function of the gauge for different lead times. Dashed lines in panels (a–e) give 95% Kolmogorov-Smirnoff confidence intervals.

time. Although there is a noticeable reduction in skill for a 3-month lead time, forecasts (CRPSS median values close to 0) for all the gauges are "coherent," that is, they are always at least as skillful as climatology forecasts (Bennett et al., 2021; Krzysztofowicz, 1999). In the case of the CRPSS for only high flow years (Figure 8b), similar features to the entire period are seen, but with an enhancement in the skill, especially for a 0- and 1-month lead time.

Figure 9 displays PIT plots and the α-index of the BHM forecast for different lead times and five gauges of the NRB. For the first four gauges (Figures 9a–9d), the PIT plots are close to the 1:1 line for all gauges but slightly overestimate the forecast uncertainty for different lead times (orange gradient scheme). Sandiya gauge exhibits (Figure 9e) similar plots for most of the lead times except for a 1-month lead time that underestimates the forecast uncertainty (s-shape), with low PIT values close to 95% Kolmogorov-Smirnoff confidence intervals. Still, the rest of the PIT values are close to the 1:1 line. These findings and α-index values above 0.9 for all lead times at different gauges indicate excellent forecast reliability for the four lead times.
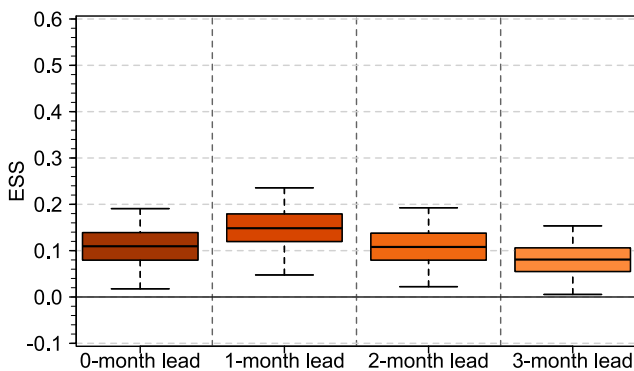

Figure 10. Boxplots of energy skill score of the Narmada River basin for different lead times (orange color scheme) from the cross-validation. The bold horizontal line in the boxplot is the median, the box of a boxplot shows the interquartile range (Q25–Q75), and the whiskers represent the sample 95%. Each boxplot comprises results from 3,000 sample bootstraps. Climatology is considered the reference model.

4.4. Probabilistic Multivariate Performance Metrics

Figure 10 displays the cross-validation energy skill score (ESS) distribution for different lead times. An increase in the ESS is seen from 0- to 1-month leads, which can be explained by the lower skill at-site (CRPSS) obtained for
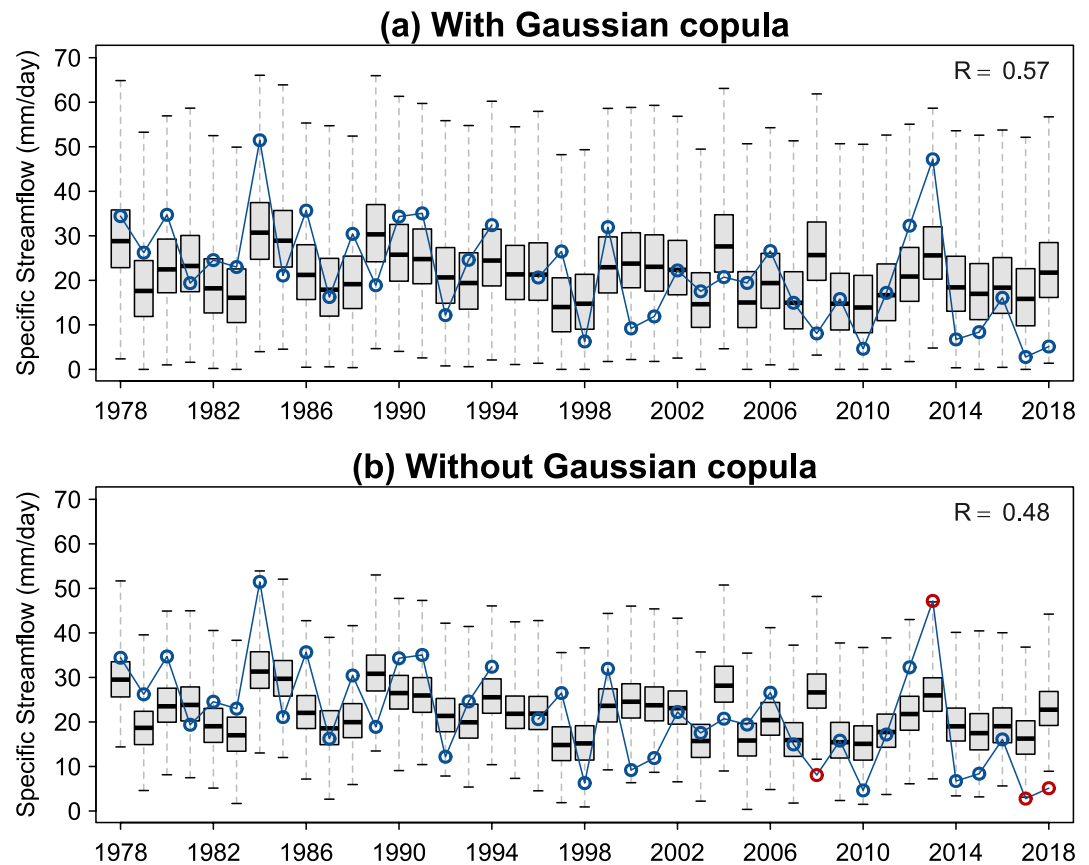
Figure 11. Time series of average forecasted daily monsoon maximum specific streamflow ensembles over all five gauges of the Narmada River basin for the best model with a 0-month lead time (b) with a Gaussian copula and (b) without a Gaussian copula. Blue and red points indicate when observed values are captured (or not) by the ensemble's variability, respectively. Whiskers denote the 95% credible intervals, boxes the interquartile range, and horizontal lines inside the boxes the median. Outliers are not displayed. *R* corresponds to the correlation between the posterior median and observations.

0-month compared to 1-month leads at Garudeshwar and Mandleshwar. After 1-month leads, ESS decreases as the lead time increases, indicating a decrease in model skill with increasing lead time. However, the four lead times capture the multivariate dependence better than the climatology forecast (ESS values above 0). In addition, ESS values above 0 were obtained if we used the BHM without the copula as the reference model (Figure S13 in Supporting Information S1), which implies that the BHM with a copula also captures the multivariate dependence better than BHM without.

To highlight the ability of Gaussian copula to capture the multivariate dependence, we computed the time series of the spatially averaged forecast of the monsoon daily maximum specific streamflow over all five gauges of the NRB. Figure 11 shows the time series obtained from the cross-validation process for the best model at a 0-month lead time with and without a Gaussian copula. Results show that by adding a Gaussian copula to the BHM, there is an increase in the correlation between the posterior median and observations that the ensemble's interquartile range (Q25–Q75, Figure 11a) captures the correlations of the observations very well. Further, the ensembles represent a higher variability, which allows for the capture of some observations that are not well captured without the copula (Figure 11b). Similar results are exhibited for other lead times (Figures S14–S16 in Supporting Information S1).

4.5. NEETM Forecast

We obtained the observed values of NEETM considering as a threshold 80th observed quantile of the daily monsoon maximum streamflow observed (Q_{80th}). For the NEETM time series: the percentage of years with
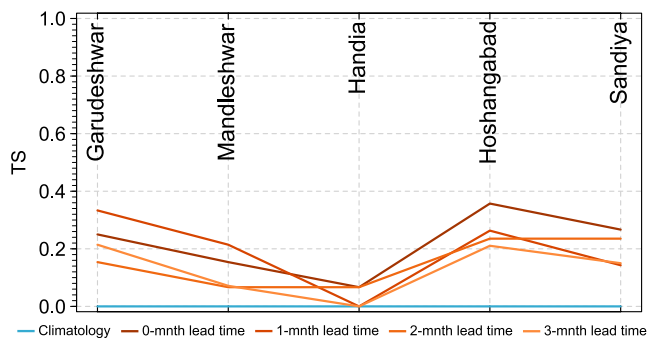
Figure 12. TS values of forecasted NEETM at the gauges of the Narmada River basin for different lead times from the cross-validation for the category NEETM >0.

high flow events (NEETM > 0) for the NRB gauges range from 19% to 22%. The time series of NEETM are displayed in Figure S17 in Supporting Information S1.

We applied the BHM to forecast NEETM for 0- to 3- month lead times as it was described in Section 2.1. We considered the same covariates used for maximum flows (Section 3.2), and the best BHM was selected based on the lowest WAIC value. WAIC values for the best model for each lead time are displayed in Table S2 in Supporting Information S1. We also checked the significance of the Poisson regression coefficients at a 90% confidence level for each lead time. Then, we performed a leave-one-year-out cross-validation by dropping one year from the record (1978–2018). To assess the accuracy of the BHM forecast, we used the Threat Score (DIC; Gilbert, 1984). The threat score corresponds to the number of correct "yes" forecasts divided by the number of occasions when the event was forecast or observed. The worst possible threat score is zero, and the best possible threat score is one (Wilks, 2011). We considered the median NEETM ensembles for each lead time, classified them into two categories (NEETM = 0 and NEETM > 0), and computed the TS for the NEETM > 0. We did not calculate the TS for NEETM = 0 since we are interested in the occurrence of high events during the peak monsoon. Also, the climatological forecast is expected to predict this category (NEETM = 0) better, given its overpopulation in the observed data (see Figure S17 in Supporting Information S1). Figure 12 shows the results obtained for the cross-validation. Although TS values are not so high (below 0.4), the BHM model at different lead times adds some skill compared to climatology at almost all gauges and lead times. The exception is Handia for 1- and 3- month lead times with similar TS as climatology. The low increase in the TS can be explained by the low number of observed nonzero NEETM values at different gauges of the NRB, which biases the model's fit toward the non-occurrence of high events. The BHM calibration results show a slight increase in TS compared to the cross-validation for almost all the gauges and lead times (Figure S18 in Supporting Information S1).

5. Summary and Discussion

This study presented a Bayesian hierarchical model (BHM) to forecast two flood risk attributes - seasonal daily maximum flow and the number of events that exceed a threshold during a season (NEETM)—and demonstrated it in a rainfed river NRB basin in India for 0–3 months lead times. The flood risk attributes at several gauges in a basin are modeled using a Gaussian copula and generalized extreme value (GEV; magnitude) or Poisson (frequency) marginals with non-stationary parameters and without a Gaussian copula. These parameters are modeled as a linear function of suitable covariates. As potential covariates, we used indices of large-scale climate teleconnections, that is, Niño 3.4, Niño 1+2, IOD, PWPR, and the new indices WIO, EWIO, EPCT, and EEPCT, defined at this basin for 0-, 1-, 2-, and 3-month lead time.

Two criteria were adopted for selecting the best model for different lead times: the WAIC (Watanabe, 2010) and the significance of the GEV (Poisson) regression coefficients for the daily monsoon maximum streamflow (NEETM). The Higher WAIC value for the stationary model compared to the best BHM model for each lead time (Table 3 and Table S2 in Supporting Information S1) and the significance of the regression coefficients for the selected models illustrate the need for the time-varying model. This need is further bolstered by the fact that stationary models, such as climatology, are insufficient for probabilistic forecasting because they generate the same ensemble forecast every time, disregarding the temporal variability.

The selection of the best candidate Bayesian hierarchical models (BHMs) for daily monsoon maximum streamflow revealed that WIO, EWIO, EPCT, and EEPCT were selected as a covariate at the corresponding lead times, with traditional large-scale climate indices included as additional covariate only for a 2- and 3-month lead times (PWPR). These results suggest that although several authors have reported correlation between large-scale climate indices and seasonal rainfall over India (Curtis et al., 2001; Kumar et al., 2006; Saji et al., 1999), they do not necessarily imply strong teleconnection with maximum streamflow. Thus, conducting detailed climate diagnostics to find skillful covariates for daily seasonal maximum streamflow is highly recommended. Nonetheless, using just the large-scale climate indices as covariates, the BHM provides skillful forecasts of seasonal flow extremes compared to climatology, with higher skill at shorter lead times. This feature is a powerful aspect

of the proposed method. This study is one of the first attempts at long-lead forecasting of attributes of seasonal streamflow extremes in India. It offers increased prospects of application in other basins of the country and other parts of the world. Seasonal basin average precipitation was found to correlate strongly with the flow extremes; however, spatial precipitation forecasts are not available at longer lead times. However, seasonal climate forecasts from Indian Meteorological Department (IMD) and other agencies could be exploited in a follow-up study. In addition, we should remark that our BHM provides a correlation skill of 0.4–0.5 of seasonal streamflow maximum at some gauges much ahead of the actual occurrence of the extreme. These skill are comparable to the ISMR forecast across India for a 0-month lead time for different studies ($r < 0.46$; Delsole & Shukla, 2012; Jain et al., 2019; Köhn-Reich & Bürger, 2019; Rajeevan et al., 2012). However, the ISMR forecast skill is for a single seasonal value for India, which is of limited help in flood management.

Overall, the BHM framework presented has the following benefits.

- Considering non-stationary parameters of marginal distributions by linear regressions of them with time-dependent covariates allow capturing the temporal variability of streamflow extremes.
- Provides "coherent" forecast, that is, never less accurate than climatology forecast.
- Including a copula in the BHM enables capturing spatiotemporal dependence of streamflow extremes, which is not entirely possible with independent marginal models.

The comparative analysis for four different lead times revealed that increasing the lead time from 0- to 2- months only weakly decreases model skill for the first four gauges (Garudeshwar, Mandleshwar, Handia, and Hoshangabad), which are the closest gauges to the high population center of the basin. Even though a higher skill is reported for the 1-month lead relative to the 0-month lead at the two lower gauges (Garudeshwar and Mandleshwar), these findings imply that the framework proposed could be helpful for the early implementation of a robust system of flood risk adaptation and preparedness strategies. For example, this medium-term forecast of seasonal maximum (up to 2-month lead) can help policymakers immensely to devise risk-based mitigation and adaptation strategies before the start of the season by considering early alerts based on the exceedance of a flow threshold by specific ensemble forecast quantiles (e.g., 50th to 75th; Ossandón, Brunner, et al., 2022). That early flooding alert can complement the short-term daily forecast model implemented by Ossandón, Rajagopalan, et al. (2022) and provide a multi-step flood risk system alarm. Such a communication strategy could help decision-makers implement adaptation strategies that address the spatial dimension of flooding.

In the case of NEETM, the results of the brief demonstration implemented in this study are promising, but further work and analysis need to be conducted to get better results. Notably, a trade-off analysis between the ability of the BHM to capture nonzero NEETM (times series with a higher number of nonzero values) and the setting of a threshold high enough to be useful for risk mitigation planning. Also, using covariates skillful for NEETM instead of just relying on the skill of the same covariates to predict the maximum flow magnitude could help improve the performance of BHM for NEETM forecasting.

Finally, this framework's future extensions include using long-term forecasts of seasonal precipitation as potential covariates, its extension to model and forecast the river stage, valuable input for evacuation plans, and its application to other rainfed river basins across the world.

Data Availability Statement

The data set used in this study, which consists of time series of potential covariates, seasonal daily maximum streamflow and NEETM at five station gauges in the Narmada River basin, and the scripts for the implementation of the BHM proposed here, can be downloaded from https://zenodo.org/badge/latestdoi/593792803.

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*(14), 6972–6982. https://doi.org/10.1029/2018GL078689

Anghileri, D., Voisin, N., Castelletti, A., Pianosi, F., Nijssen, B., & Lettenmaier, D. P. (2016). Value of long-term streamflow forecasts to reservoir operations for water supply in snow-dominated river catchments. *Water Resources Research*, *52*(6), 4209–4225. https://doi.org/10.1002/2015WR017864

Bennett, J. C., Wang, Q. J., Li, M., Robertson, D. E., & Schepen, A. (2016). Reliable long-range ensemble streamflow forecasts: Combining calibrated climate forecasts with a conceptual runoff model and a staged error model. *Water Resources Research*, *52*(10), 8238–8259. https://doi.org/10.1002/2016WR019193

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

Bracken, C., Holman, K. D., Rajagopalan, B., & Moradkhani, H. (2018). A Bayesian hierarchical approach to multivariate nonstationary hydrologic frequency analysis. *Water Resources Research*, *54*(1), 243–255. https://doi.org/10.1002/2017WR020403

Bracken, C., Rajagopalan, B., Cheng, L., Kleiber, W., & Gangopadhyay, S. (2016). Spatial Bayesian hierarchical modeling of precipitation extremes over a large domain. *Water Resources Research*, *52*(8), 6643–6655. https://doi.org/10.1002/2016WR018768

Brunner, M. I., Papalexiou, S., Clark, M. P., & Gilleland, E. (2020). How probable is widespread flooding in the United States? *Water Resources Research*, *56*(10), e2020WR028096. https://doi.org/10.1029/2020WR028096

Chevuturi, A., Turner, A. G., Johnson, S., Weisheimer, A., Shonk, J. K., Stockdale, T. N., & Senan, R. (2021). Forecast skill of the Indian monsoon and its onset in the ECMWF seasonal forecasting system 5 (SEAS5). *Climate Dynamics*, *56*(9–10), 2941–2957. https://doi.org/10.1007/S00382-020-05624-5/FIGURES/9

Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., & Wilby, R. (2004). The Schaake Shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields. *Journal of Hydrometeorology*, *5*(1), 243–262. https://doi.org/10.1175/1525-7541(2004)005⟨0243:TSSAMF⟩2.0.CO;2

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer.

Coles, S. G., & Tawn, J. A. (1996). Modelling extremes of the areal rainfall process. *Journal of the Royal Statistical Society: Series B*, *58*(2), 329–347. https://doi.org/10.1111/j.2517-6161.1996.tb02085.x

Cooley, D., Nychka, D., & Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, *102*(479), 824–840. https://doi.org/10.1198/016214506000000780

Cooley, D., & Sain, S. R. (2010). Spatial hierarchical modeling of precipitation extremes from a regional climate model. *Journal of Agricultural, Biological, and Environmental Statistics*, *15*(3), 381–402. https://doi.org/10.1007/s13253-010-0023-9

Curtis, S., Adler, R., Huffman, G., Nelkin, E., & Bolvin, D. (2001). Evolution of tropical and extratropical precipitation anomalies during the 1997–1999 ENSO cycle. *International Journal of Climatology*, *21*(8), 961–971. https://doi.org/10.1002/JOC.643

Delsole, T., & Shukla, J. (2012). Climate models produce skillful predictions of Indian summer monsoon rainfall. *Geophysical Research Letters*, *39*(9). https://doi.org/10.1029/2012GL051279

De Perez, E. C., Stephens, E., Bischiniotis, K., Van Aalst, M., Van Den Hurk, B., Mason, S., et al. (2017). Should seasonal rainfall forecasts be used for flood preparedness? *Hydrology and Earth System Sciences*, *21*(9), 4517–4524. https://doi.org/10.5194/HESS-21-4517-2017

Field, C., Barros, V., Stocker, T., Qin, D., Dokken, D., Ebi, K., et al. (2012). *Managing the risks of extreme events and disasters to advance climate change adaptation: Special report of the Intergovernmental panel on climate change (Technical report)*. Cambridge University Press.

Gao, S., Liu, P., & Lall, U. (2021). Seasonal precipitation predictability for the northern Hemisphere using concurrent and preseason atmospheric water vapor transport and sea surface temperature. *Journal of Hydrometeorology*, *22*(1), 183–199. https://doi.org/10.1175/JHM-D-20-0107.1

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

Ghosh, S., & Mallick, B. K. (2011). A hierarchical Bayesian spatio-temporal model for extreme precipitation events. *Environmetrics*, *22*(2), 192–204. https://doi.org/10.1002/ENV.1043

Gilbert, G. F. (1984). Finley's tornado predictions. *American Meteorological Journal*, *1*, 166–172.

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

Gneiting, T., Stanberry, L. I., Grimit, E. P., Held, L., & Johnson, N. A. (2008). Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *Test*, *17*(2), 211–235. https://doi.org/10.1007/s11749-008-0114-x

Govindaraju, R. S. (2000). Artificial neural networks in hydrology. I: Preliminary concepts. *Journal of Hydrologic Engineering*, *5*(2), 115–123. https://doi.org/10.1061/(asce)1084-0699

Grantz, K., Rajagopalan, B., Clark, M., & Zagona, E. (2005). A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts. *Water Resources Research*, *41*(10), 10410. https://doi.org/10.1029/2004WR003467

Hadi, S. J., & Tombul, M. (2018). Forecasting daily streamflow for basins with different physical characteristics through data-driven methods. *Water Resources Management*, *32*(10), 3405–3422. https://doi.org/10.1007/s11269-018-1998-1

He, J., Anderson, A., & Valeo, C. (2015). Bias compensation in flood frequency analysis. *Hydrological Sciences Journal*, *60*(3), 381–401. https://doi.org/10.1080/02626667.2014.885651

Hochrainer-Stigler, S. (2020). *Extreme and systemic risk analysis* (Integrated ed.). Springer. https://doi.org/10.1007/978-981-15-2689-3

Hoffman, M. D., & Gelman, A. (2014). The No-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*(47), 1593–1623.

Hrudya, P. H., Varikoden, H., & Vishnu, R. (2020). A review on the Indian summer monsoon rainfall, variability and its association with ENSO and IOD. *Meteorology and Atmospheric Physics*, *133*(1), 1–14. https://doi.org/10.1007/S00703-020-00734-5

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

Jain, S., Scaife, A. A., & Mitra, A. K. (2019). Skill of Indian summer monsoon rainfall prediction in multiple seasonal prediction systems. *Climate Dynamics*, *52*(9–10), 5291–5301. https://doi.org/10.1007/S00382-018-4449-Z/FIGURES/8

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

Kaplan, A., Cane, M. A., Kushnir, Y., Clement, A. C., Blumenthal, M. B., & Rajagopalan, B. (1998). Analyses of global sea surface temperature 1856–1991. *Journal of Geophysical Research*, *103*(C9), 18567–18589. https://doi.org/10.1029/97JC01736

Katz, R. W. (2013). Statistical methods for nonstationary extremes. In A. AghaKouchak, D. Easterling, K. Hsu, S. Schubert, & S. Sorooshian (Eds.), *Extremes in a changing climate* (65th ed., pp. 15–37). Springer. https://doi.org/10.1007/978-94-007-4479-0_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

Köhn-Reich, L., & Bürger, G. (2019). Dynamical prediction of Indian monsoon: Past and present skill. *International Journal of Climatology*, *39*(8), 3574–3581. https://doi.org/10.1002/JOC.6039

Koster, R. D., Mahanama, S. P., Livneh, B., Lettenmaier, D. P., & Reichle, R. H. (2010). Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow. *Nature Geoscience*, *3*(9), 613–616. https://doi.org/10.1038/ngeo944

Krishnaswamy, J., Vaidyanathan, S., Rajagopalan, B., Bonell, M., Sankaran, M., Bhalla, R. S., & Badiger, S. (2015). Non-stationary and non-linear influence of ENSO and Indian Ocean Dipole on the variability of Indian monsoon rainfall and extreme rain events. *Climate Dynamics*, *45*(1–2), 175–184. https://doi.org/10.1007/S00382-014-2288-0/FIGURES/5

Krzysztofowicz, R. (1999). Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resources Research*, *35*(9), 2739–2750. https://doi.org/10.1029/1999WR900099

Kumar, K. K., Rajagopalan, B., Hoerling, M., Bates, G., & Cane, M. (2006). Unraveling the mystery of Indian monsoon failure during El Niño. *Science*, *314*(5796), 115–119. https://doi.org/10.1126/SCIENCE.1131152/SUPPL_FILE/KRISHNAKUMAR_SOM.PDF

Kurian, C., Sudheer, K. P., Vema, V. K., & Sahoo, D. (2020). Effective flood forecasting at higher lead times through hybrid modelling framework. *Journal of Hydrology*, *587*, 124945. https://doi.org/10.1016/j.jhydrol.2020.124945

Kwon, H.-H., Brown, C., & Lall, U. (2008). Climate informed flood frequency analysis and prediction in Montana using hierarchical Bayesian modeling. *Geophysical Research Letters*, *35*(5), L05404. https://doi.org/10.1029/2007GL032220

Kwon, H. H., Brown, C., Xu, K., & Lall, U. (2009). Seasonal and annual maximum streamflow forecasting using climate information: Application to the Three Gorges Dam in the Yangtze River basin, China. *Hydrological Sciences Journal*, *54*(3), 582–595. https://doi.org/10.1623/hysj.54.3.582

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001. https://doi.org/10.1016/J.JMVA.2009.04.008

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

Lima, C. H., & Lall, U. (2010). Climate informed monthly streamflow forecasts for the Brazilian hydropower network using a periodic ridge regression model. *Journal of Hydrology*, *380*(3–4), 438–449. https://doi.org/10.1016/j.jhydrol.2009.11.016

Livneh, B., & Badger, A. M. (2020). Drought less predictable under declining future snowpack. *Nature Climate Change*, *10*(5), 452–458. https://doi.org/10.1038/s41558-020-0754-8

López, J., & Francés, F. (2013). Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates. *Hydrology and Earth System Sciences*, *17*(8), 3189–3203. https://doi.org/10.5194/HESS-17-3189-2013

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

McInerney, D., Thyer, M., Kavetski, D., Lerat, J., & Kuczera, G. (2017). Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. *Water Resources Research*, *53*(3), 2199–2239. https://doi.org/10.1002/2016WR019168

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*(7), 6030–6052. https://doi.org/10.1002/2014WR015426

Najafi, M. R., & Moradkhani, H. (2014). A hierarchical Bayesian approach for the analysis of climate change impact on runoff extremes. *Hydrological Processes*, *28*(26), 6292–6308. https://doi.org/10.1002/hyp.10113

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*, *1*(9–10), 1–21. https://doi.org/10.1007/S00382-022-06244-X

Neiman, P. J., Ralph, F. M., Wick, G. A., Kuo, Y. H., Wee, T. K., Ma, Z., et al. (2008). Diagnosis of an intense atmospheric river impacting the Pacific Northwest: Storm summary and offshore vertical structure observed with COSMIC satellite retrievals. *Monthly Weather Review*, *136*(11), 4398–4420. https://doi.org/10.1175/2008MWR2550.1

Ossandón, Á., Brunner, M. I., Rajagopalan, B., & Kleiber, W. (2022). A space–time Bayesian hierarchical modeling framework for projection of seasonal maximum streamflow. *Hydrology and Earth System Sciences*, *26*(1), 149–166. https://doi.org/10.5194/HESS-26-149-2022

Ossandón, Á., Nanditha, J. S., Mendoza, P. A., Rajagopalan, B., & Mishra, V. (2022). A Bayesian hierarchical framework for postprocessing 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., & Kleiber, W. (2021). Spatial-temporal multivariate semi-Bayesian hierarchical framework for extreme precipitation frequency analysis. *Journal of Hydrology*, *600*, 126499. https://doi.org/10.1016/j.jhydrol.2021.126499

Ossandón, Á., Rajagopalan, B., Tiwari, A. D., & Mishra, V. (2022). A Bayesian hierarchical model combination framework for real-time daily ensemble streamflow forecasting across a rainfed river basin (in rewiew). *Earth's Future*, *10*(12). https://doi.org/10.1029/2022ef002958

Pagano, T. C., Garen, D. C., Perkins, T. R., & Pasteris, P. A. (2009). Daily updating of operational statistical seasonal water supply forecasts for the Western U.S. *JAWRA Journal of the American Water Resources Association*, *45*(3), 767–778. https://doi.org/10.1111/j.1752-1688.2009.00321.x

Pai, D., Sridhar, L., Rajeevan, M., Sreejith, O. P., Satbhai, N. S., & Mukhopadhyay, B. (2014). Development of a new high spatial resolution $(0.25° \times 0.25°)$ 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

Papacharalampous, G. A., & Tyralis, H. (2018). Evaluation of random forests and Prophet for daily streamflow forecasting. *Advances in Geosciences*, *45*, 201–208. https://doi.org/10.5194/adgeo-45-201-2018

Papalexiou, S. M., & Montanari, A. (2019). Global and regional increase of precipitation extremes under global warming. *Water Resources Research*, *55*(6), 4901–4914. https://doi.org/10.1029/2018WR024067

Parker, D. E., Jones, P. D., Folland, C. K., & Bevan, A. (1994). Interdecadal changes of surface temperature since the late nineteenth century. *Journal of Geophysical Research*, *99*(D7), 14373–14399. https://doi.org/10.1029/94JD00548

Penn, C. A., Clow, D. W., Sexstone, G. A., & Murphy, S. F. (2020). Changes in climate and land cover affect seasonal streamflow forecasts in the Rio Grande headwaters. *Journal of the American Water Resources Association*, *56*(5), 882–902. https://doi.org/10.1111/1752-1688.12863

Rajagopalan, B., & Molnar, P. (2012). Pacific Ocean sea-surface temperature variability and predictability of rainfall in the early and late parts of the Indian summer monsoon season. *Climate Dynamics*, *39*(6), 1543–1557. https://doi.org/10.1007/S00382-011-1194-Y/FIGURES/13

Rajeevan, M., Unnikrishnan, C. K., & Preethi, B. (2012). Evaluation of the ENSEMBLES multi-model seasonal forecasts of Indian summer monsoon variability. *Climate Dynamics*, *38*(11–12), 2257–2274. https://doi.org/10.1007/S00382-011-1061-X/FIGURES/13

Regonda, S. K., Rajagopalan, B., Clark, M., & Zagona, E. (2006). A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin. *Water Resources Research*, *42*(9), 9404. https://doi.org/10.1029/2005WR004653

Reis, D. S., & Stedinger, J. R. (2005). Bayesian MCMC flood frequency analysis with historical information. *Journal of Hydrology*, *313*(1–2), 97–116. https://doi.org/10.1016/J.JHYDROL.2005.02.028

Renard, B. (2011). A Bayesian hierarchical approach to regional frequency analysis. *Water Resources Research*, *47*(11). https://doi.org/10.1029/2010WR010089

Renard, B., & Lang, M. (2007). Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. *Advances in Water Resources*, *30*(4), 897–912. https://doi.org/10.1016/j.advwatres.2006.08.001

Reynolds, R. W., & Smith, T. M. (1994). Improved global sea surface temperature analyses using optimum interpolation. *Journal of Climate*, *7*(6), 929–948. https://doi.org/10.1175/1520-0442(1994)007<0929:igssta>2.0.co;2

Reza Najafi, M., & Moradkhani, H. (2013). Analysis of runoff extremes using spatial hierarchical Bayesian modeling. *Water Resources Research*, *49*(10), 6656–6670. https://doi.org/10.1002/wrcr.20381

Robertson, D. E., & Wang, Q. J. (2012). A Bayesian approach to predictor selection for seasonal streamflow forecasting. *Journal of Hydrometeorology*, *13*(1), 155–171. https://doi.org/10.1175/JHM-D-10-05009.1

Ruiz, J. E., Cordery, I., & Sharma, A. (2007). Forecasting streamflows in Australia using the tropical Indo-Pacific thermocline as predictor. *Journal of Hydrology*, *341*(3–4), 156–164. https://doi.org/10.1016/j.jhydrol.2007.04.021

Saji, N. H., Goswami, B. N., Vinayachandran, P. N., & Yamagata, T. (1999). A dipole mode in the tropical Indian Ocean. *Nature*, *401*(6751), 360–363. https://doi.org/10.1038/43854

Sankarasubramanian, A., & Lall, U. (2003). Flood quantiles in a changing climate: Seasonal forecasts and causal relations. *Water Resources Research*, *39*(5), 1134. https://doi.org/10.1029/2002WR001593

Sivakumar, B. (2016). *Chaos in hydrology: Bridging determinism and stochasticity*. Springer. https://doi.org/10.1007/978-90-481-2552-4

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B*, *64*(4), 583–639. https://doi.org/10.1111/1467-9868.00353

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/

Steinschneider, S., & Lall, U. (2016). El Niño and the U.S. precipitation and floods: What was expected for the January–March 2016 winter hydroclimate that is now unfolding? *Water Resources Research*, *52*(2), 1498–1501. https://doi.org/10.1002/2015WR018470

Steirou, E., Gerlitz, L., Apel, H., Sun, X., & Merz, B. (2019). Climate influences on flood probabilities across Europe. *Hydrology and Earth System Sciences*, *23*(3), 1305–1322. https://doi.org/10.5194/HESS-23-1305-2019

Steirou, E., Gerlitz, L., Sun, X., Apel, H., Agarwal, A., Totz, S., & Merz, B. (2022). Towards seasonal forecasting of flood probabilities in Europe using climate and catchment information. *Scientific Reports*, *12*(1), 1–10. https://doi.org/10.1038/s41598-022-16633-1

Tanoue, M., Hirabayashi, Y., & Ikeuchi, H. (2016). Global-scale river flood vulnerability in the last 50 years. *Scientific Reports*, *6*(1), 1–9. https://doi.org/10.1038/srep36021

Tiwari, A. D., Mukhopadhyay, P., & Mishra, V. (2021). Influence of bias correction of meteorological and streamflow forecast on hydrological prediction in India. *Journal of Hydrometeorology*. https://doi.org/10.1175/JHM-D-20-0235.1

Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P.-C., Paananen, T., & Gelman, A. (2020). *loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models*. (R package version 2.4.1). Stan Development Team. Retrieved from https://mc-stan.org/loo/

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

Villarini, G., Smith, J. A., Serinaldi, F., Ntelekos, A. A., & Schwarz, U. (2012). Analyses of extreme flooding in Austria over the period 1951–2006. *International Journal of Climatology*, *32*(8), 1178–1192. https://doi.org/10.1002/JOC.2331

Wallemacq, P., & House, R. (2018). *Economic losses, poverty and disasters: 1998-2017 (Technical report)*. UNISDR and CRED.

Wang, Q. J., & Robertson, D. E. (2011). Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences. *Water Resources Research*, *47*(2), 2546. https://doi.org/10.1029/2010WR009333

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*(5), 5407. 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

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571–3594.

Werner, K., & Yeager, K. (2013). Challenges in forecasting the 2011 runoff season in the Colorado basin. *Journal of Hydrometeorology*, *14*(4), 1364–1371. https://doi.org/10.1175/JHM-D-12-055.1

Wijayarathne, D. B., & Coulibaly, P. (2020). Identification of hydrological models for operational flood forecasting in St. John's, Newfoundland, Canada. *Journal of Hydrology: Regional Studies*, *27*, 100646. https://doi.org/10.1016/j.ejrh.2019.100646

Wikle, C. K., Berliner, L. M., & Cressie, N. (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, *5*(2), 117–154. https://doi.org/10.1023/A:1009662704779

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

Wood, A. W., Hopson, T., Newman, A., Brekke, L., Arnold, J., & Clark, M. (2016). Quantifying streamflow forecast skill elasticity to initial condition and climate prediction skill. *Journal of Hydrometeorology*, *17*(2), 651–668. https://doi.org/10.1175/JHM-D-14-0213.1

Yan, H., & Moradkhani, H. (2015). A regional Bayesian hierarchical model for flood frequency analysis. *Stochastic Environmental Research and Risk Assessment*, *29*(3), 1019–1036. https://doi.org/10.1007/s00477-014-0975-3

Yan, H., & Moradkhani, H. (2016). Toward more robust extreme flood prediction by Bayesian hierarchical and multimodeling. *Natural Hazards*, *81*(1), 203–225. https://doi.org/10.1007/s11069-015-2070-6

Zanardo, S., Nicotina, L., Hilberts, A. G., & Jewson, S. P. (2019). Modulation of economic losses from European floods by the North Atlantic Oscillation. *Geophysical Research Letters*, *46*(5), 2563–2572. https://doi.org/10.1029/2019GL081956

Zhao, T., Schepen, A., & Wang, Q. J. (2016). Ensemble forecasting of sub-seasonal to seasonal streamflow by a Bayesian joint probability modelling approach. *Journal of Hydrology*, *541*, 839–849. https://doi.org/10.1016/J.JHYDROL.2016.07.040