A high-dimensional Bayesian skew-Gaussian spatial model for analyzing rainfall trends

Arghadeep Basu and Arnab Hazra

Abstract Analyzing the effect of climate change on rainfall is crucial from the perspective of agriculture, water resources management, and disaster management. Motivated by the key features present in a gridded rainfall dataset published by the India Meteorological Department (IMD) for the years 1951–2022 at a spatial resolution of $0.25^\circ \times 0.25^\circ$ across India, we propose a high-dimensional Bayesian skew-Gaussian conditional autoregressive model to analyze spatially-varying temporal patterns in Indian monsoon rainfall. We model the mean component in a semiparametric regression framework with spatially-varying coefficients and assume the skewness component to be also spatially-varying. We briefly discuss some theoretical properties of the proposed model. We draw inferences using Gibbs sampling by exploiting the hierarchical definition of a skew-Gaussian distribution as a random location-mixture of a Gaussian distribution. We study the posterior mean and the corresponding T-statistics of the overall change in rainfall amounts during the observational period at each grid cell. Our results indicate a significant negative trend in various parts of the Gangetic River basin, where agriculture is predominantly dependent on the availability of rainwater.

Key words: High-dimensional Bayesian model; Rainfall trends estimation; Skew-Gaussian process

Arghadeep Basu
Department of Statistics, University of Georgia, Athens, Georgia 30602, United States
e-mail: arghadeep.basu@uga.edu

Arnab Hazra
Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur 208016, India
e-mail: ahazra@iitk.ac.in

1 Introduction

Rainfed agriculture refers to a type of farming where rainfall is the primary source of water for crops. In India, rainfed agriculture covers 67% of the total sown areas, the world's largest extent of such farming venkateswarlu2011rainfed. However, future climate assessments suggest that dry regions are gradually becoming drier, which could significantly impact the productivity of rainfed agriculture and exacerbate food security concerns [3]. To address these issues, the Government of India established the National Rainfed Area Authority (NRAA) in 2007 and implemented the Rainfed Area Development Programme (RADP) in 22 states of India during 2012–2013 [11]. Later, in 2015, the government consolidated various irrigation schemes into the Pradhan Mantri Krishi Sinchayee Yojana (PMKSY) to enhance irrigation facilities. However, according to the Economic Survey Report 2017–2018, 52% of the total cropped area still lacks irrigation, relying solely on rainfall. Consequently, the monsoon months (June, July, August, and September; JJAS, henceforth) are critical for farming. Moreover, rainfall anomalies lead to floods and droughts, causing severe crop damage. Accurate assessment of the availability of rainwater is essential to avoid crop failure, and statistical modeling of rainfall data plays a crucial role in agro-meteorology in this context.

We analyze a gridded rainfall dataset for the years 1951–2022 across India. Total rainfall amounts during monsoon months are usually nonzero throughout India, and we ignore a tiny fraction of cases (0.62% in our case) where observed rainfall is zero. We usually model rainfall data for each year using a spatial stochastic process with continuous responses [4, 14]. Gaussian processes (GPs) are prevalent models in spatial geostatistics because of their several favorable theoretical and computational characteristics [6]. However, there are three main issues with fitting a GP model for our case study: (i) because of the large spatial dimension of the dataset, the inference involves computing the determinant and inverse of the large covariance matrix, which is computationally demanding, (ii) a GP is a valid model for point-referenced spatial data, while due to the gridded spatial structure, it is more appropriate to choose areal models like the conditional autoregressive (CAR) model, and (iii) the data exhibit spatially-varying temporal trend and skewness in the marginal and joint distributions, indicating the need for a skewness component to be added to a GP.

The CAR models were first proposed by [2]. They are based on Markovian properties such that the conditional distribution of a component of the random vector (observed data or parameter vector in a Bayesian setting) depends only on a set of neighbors and are particular cases of Markov random fields [17]. They are appropriate models for areal data and allow a sparse inverse covariance matrix; hence, issues (i) and (ii) can be addressed using a CAR model for the spatial fields. The long-term spatially-varying trend can be modeled flexibly using a semiparametric regression framework [7, 8], and the skewness can be modeled using a skew-GP [19, 13]. However, as of our knowledge, no study has considered these aspects simultaneously, which is required to estimate the effects of climate change in a high-dimensional skewed spatial GP.

We propose a Bayesian skew-CAR model to examine the spatially-varying temporal patterns within Indian monsoon rainfall in a high-resolution setting. The study assumes spatially-varying coefficients to model the mean component in a semiparametric regression framework and introduces a spatially-varying skewness component. The paper also outlines the model's theoretical properties. To draw inferences, we utilize Gibbs sampling [5] by leveraging the hierarchical nature of a skew-GP in terms of a random location-mixture of a GP. The study investigates the posterior mean and corresponding T-statistics to analyze the overall changes in rainfall at each pixel during the study period.

The paper is organized as follows. In Section 2, we provide an overview of the Indian monsoon rainfall dataset and present some exploratory analyses. The statistical methodology is described in Section 3. The Bayesian computation for the proposed model is discussed in Section 4. Section 5 contains a discussion of the results, while the paper concludes with some final remarks in Section 6.

2 The Indian Monsoon Rainfall Data

2.1 Data Description and Pre-processing

We obtain daily gridded rainfall data from the India Meteorological Department (IMD) at a resolution of $0.25^\circ \times 0.25^\circ$ across India for 1951–2022. This dataset is prepared by [16] using daily rainfall records from 6995 rain gauge stations in India. Focusing on the analysis of total rainwater availability during the monsoon months, we obtain total rainfall across the months JJAS for each year. A total of 88 grid cells near the northeastern regions are removed due to the inconsistency of the gridded data with the available literature. Finally, data are available at 4876 spatial grid cells over 72 years. The observed data for the years 1951 and 2022 are presented in Figure 1. While higher monsoon rainfall is observable near the Western Ghats and the northeastern Himalayan regions, low rainfall is observed throughout western India except the Western Ghats region for both years. Different spatial patterns are observable near Madhya Pradesh, Uttarakhand, and Arunachal Pradesh for these years.

2.2 Exploratory Data Analysis

We first explore the temporal trend by fitting a simple linear regression model with year as a covariate separately at each grid cell and observe that such a model does not fit the data well (not shown). Besides, assuming a linear trend across 72 years is questionable. Hence, we further replace the simple linear regression model with a semiparametric regression model, where the expected rainfall amounts are assumed to be a smooth function of the years. We rewrite the smooth function as a linear combination of cubic B-splines with equidistant knots, and based on 10-fold

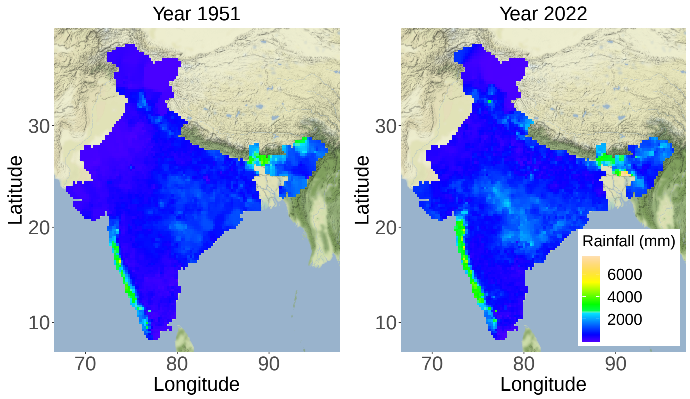

Fig. 1 Gridded monsoon rainfall for the years 1951 (left) and 2022 (right). Both panels share the same legend.

cross-validation and the minimum prediction mean square error criterion, we set the number of splines to be $K = 5$ (not shown). Based on this semiparametric model, the estimated annual rate of change (ROC) for each grid cell is presented in the left panel of Figure 2. Further, after obtaining the residuals, we calculate their marginal empirical skewness and present them in the right panel of Figure 2. While this spatial profile appears to be more jittery than that for ROC, both profiles indicate the need for spatially-varying spline coefficients and skewness. We then explore the presence of multivariate skewness using the multivariate normality test of [12] for a large number of pairs of sites. However, the tests suggest the presence of multivariate skewness in the residuals for most cases. Further, we calculate the empirical semivariance of the regression residuals, the empirical spline coefficients, and the residual skewness and present in Figure 3 after appropriate averaging. All three panels indicate the need for spatially dependent likelihood and priors for the spatial parameters.

3 Methodology

Suppose we denote the monsoon rainfall at a grid cell $\boldsymbol{s}_i$ for the t-th year by $Y_{i,t}$, where $i = 1,\ldots,N$ and $t = 1,\ldots,T$. Further, we denote the response vector for year t by $\boldsymbol{Y}_t = [Y_{1,t},\ldots,Y_{N,t}]'$, the k-th spline by $B_k(\cdot), k = 1,\ldots,K$, the vector of splines evaluated at t by $\boldsymbol{B}_t = [B_1(t),\ldots,B_K(t)]'$, and the spline coefficients for the i-th grid cell by $\boldsymbol{\beta}_i = [\beta_{i,1},\ldots,\beta_{i,K}]'$. We model $\boldsymbol{Y}_t$ hierarchically as

$$\begin{aligned} \boldsymbol{Y}_t | X_t &\overset{\text{Indep}}{\sim} \text{MVN}_N\left(\boldsymbol{\beta}^* \boldsymbol{B}_t + \boldsymbol{\lambda} X_t, \sigma^2[\boldsymbol{M} - \rho \boldsymbol{A}]^{-1}\right), \\ X_t &\overset{\text{IID}}{\sim} \text{Half-Normal}(0, \sigma^2), \;\; t = 1,\ldots,T, \end{aligned} \tag{1}$$

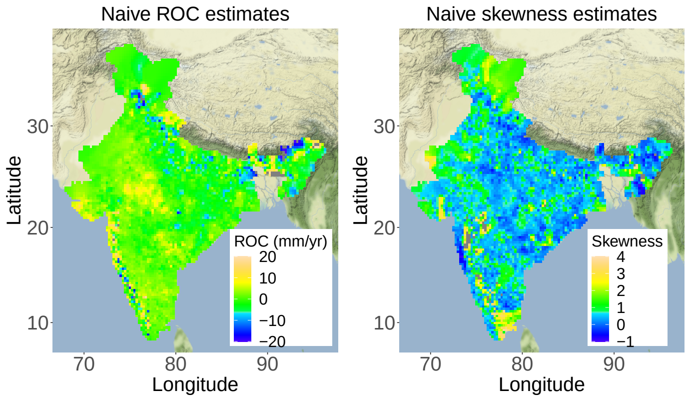

Fig. 2 Naive estimates of the annual rate of change (ROC) in monsoon rainfall (left) and naive estimates of marginal residual skewness (right).

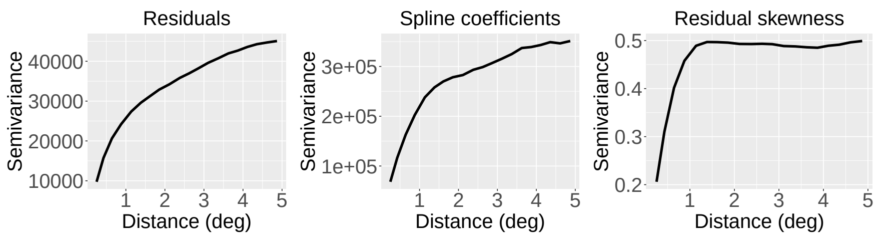

Fig. 3 Average (across years) empirical semivariance for the residuals after fitting a semiparametric mean component (left), average (across splines) empirical semivariance of the naive estimates of the spline coefficients (middle), and the empirical semivariance of the naive estimates of the marginal residual skewness (right).

where MVN_N denotes a N-variate normal distribution, $\boldsymbol{\beta}^*$ is a $(N \times K)$-dimensional matrix with its i-th row being $\boldsymbol{\beta}_i$, $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_N]'$ denotes a skewness-related parameter vector, σ^2 denotes a global variance-related parameter, $\boldsymbol{A}$ denotes a $(N \times N)$-dimensional matrix with its (i,i')-th entry being 1 if $\boldsymbol{s}_i$ and $\boldsymbol{s}_i'$ are neighbors and zero otherwise, $\boldsymbol{M}$ is a diagonal matrix with its diagonal elements being the row-wise sums of $\boldsymbol{A}$, and ρ denotes the spatial autocorrelation parameter. Here, the precision-related matrix $\boldsymbol{M} - \rho\boldsymbol{A}$ is a standard choice for gridded spatial data based on a conditional autoregressive Gaussian graphical model and obtained via Brook's lemma [2]. Thus, from (1), the conditional joint distribution of $\boldsymbol{Y} = [\boldsymbol{Y}_1', \ldots, \boldsymbol{Y}_T']'$ given $\boldsymbol{X} = [X_1, \ldots, X_T]'$ can be written as

$$\boldsymbol{Y}|\boldsymbol{X} \sim \text{MVN}_{NT}\left([\boldsymbol{B} \otimes \boldsymbol{I}_N]\boldsymbol{\beta} + [\boldsymbol{X} \otimes \boldsymbol{\lambda}], \sigma^2 \boldsymbol{I}_T \otimes [\boldsymbol{M} - \rho\boldsymbol{A}]^{-1}\right), \tag{2}$$

where $\boldsymbol{B}$ is a $(T \times K)$-dimensional matrix with its t-th row given by $\boldsymbol{B}_t$ and $\boldsymbol{\beta} = [\boldsymbol{\beta}^{*'}_{\cdot,1}, \ldots, \boldsymbol{\beta}^{*'}_{\cdot,K}]'$ with $\boldsymbol{\beta}^*_{\cdot,k}$ denoting the k-th column of $\boldsymbol{\beta}^*$.

After marginalizing over X_t, matching the notation of [1], we obtain

$$\boldsymbol{Y}_t \overset{\text{Indep}}{\sim} \text{MSN}_N\left(\boldsymbol{\beta}^*\boldsymbol{B}_t, \sigma^2([\boldsymbol{M}-\rho\boldsymbol{A}]^{-1}+\boldsymbol{\lambda}\boldsymbol{\lambda}'), [\boldsymbol{M}-\rho\boldsymbol{A}]\boldsymbol{\lambda}\right), \tag{3}$$

where MSN_N denotes an N-variate skew-normal distribution. When λ_i are equal, say $\lambda_i = \lambda_0$ for each i, (2) reduces to $\boldsymbol{Y}_t \overset{\text{Indep}}{\sim} \text{MSN}_N(\boldsymbol{\beta}^*\boldsymbol{B}_t, \sigma^2([\boldsymbol{M}-\rho\boldsymbol{A}]^{-1} + \lambda_0\mathbf{1}\mathbf{1}'), \lambda_0(1-\rho)\boldsymbol{M}\mathbf{1})$ and we have $\boldsymbol{Y}_t \overset{\text{Indep}}{\sim} \text{MVN}_N(\boldsymbol{\beta}^*\boldsymbol{B}_t, \sigma^2[\boldsymbol{M}-\rho\boldsymbol{A}]^{-1})$ if $\lambda_0 = 0$. Thus, the proposed model is a more general case of the spatial skew-Gaussian process with a fixed skewness-related component and the spatial Gaussian process. The mean vector and the covariance matrix of $\boldsymbol{Y}_t$ are given by

$$E[\boldsymbol{Y}_t] = \boldsymbol{\beta}^*\boldsymbol{B}_t + \sqrt{2/\pi}\sigma\boldsymbol{\lambda}, \;\; V[\boldsymbol{Y}_t] = \sigma^2([\boldsymbol{M}-\rho\boldsymbol{A}]^{-1} + (1-2/\pi)\boldsymbol{\lambda}\boldsymbol{\lambda}'). \tag{4}$$

Further, we assume the priors for the spatially -varying parameter vectors to be $\boldsymbol{\beta} \sim \text{MVN}_{NK}(\mathbf{0}, \sigma^2_\beta[\boldsymbol{M}-\rho_\beta\boldsymbol{A}]^{-1}\otimes\boldsymbol{I}_K)$, $\boldsymbol{\lambda} \sim \text{MVN}_N(\mathbf{0}, \sigma^2_\lambda[\boldsymbol{M}-\rho_\lambda\boldsymbol{A}]^{-1})$, and for the hyperparameters, we assume $\sigma^2, \sigma^2_\beta, \sigma^2_\lambda \overset{\text{IID}}{\sim} \text{Inverse-Gamma}(0.01, 0.01)$.

4 Computation

We draw posterior inferences about the model parameters and hyperparameters using Gibbs sampling. For the sake of feasible computation, it is crucial to assume ρ, ρ_β, and ρ_λ to be known. While drawing inferences for a different combination of choices of these parameters and then comparing the cases based on the deviance information criterion (DIC) or Watanabe Akaike information criterion (WAIC) is recommended [10], for the sake of simplicity, we choose $\rho = \rho_\beta = \rho_\lambda = 0.5$ as suggested by fitting CAR models to the residuals, estimated spline coefficients, and the marginally estimated skewness parameters and obtaining maximum likelihood estimates as a part of exploratory data analysis. While none of such estimates are exactly equal to 0.5, given the high uncertainty in such estimates, we fix them to 0.5, a middle point between spatial independence ($\rho = \rho_\beta = \rho_\lambda = 0$) and intrinsic CAR model scenario ($\rho = \rho_\beta = \rho_\lambda = 1$). The full conditional posterior distributions for the elements of $\Theta = \{\beta, \lambda, \mathbf{X_0}, \sigma^2, \sigma^2_\beta, \sigma^2_\lambda\}$ are as follows. Here, *rest* denotes $\boldsymbol{Y}$ and all the elements of Θ except the one(s) being updated.

- $\pi(\boldsymbol{\beta}|rest) \sim \text{MVN}_{NK}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$, where $\boldsymbol{\Sigma}_\beta^{-1} = [\boldsymbol{M}-\rho\boldsymbol{A}]\otimes[\sigma^{-2}\boldsymbol{B}'\boldsymbol{B} + \sigma_\beta^{-2}\boldsymbol{I}_K]$ and $\boldsymbol{\mu}_\beta = \sigma^{-2}\boldsymbol{\Sigma}_\beta\left([\boldsymbol{B}\otimes\boldsymbol{I}_N]'\left[\boldsymbol{I}_T\otimes(\boldsymbol{M}-\rho\boldsymbol{A})\right]\left[\boldsymbol{Y}-\boldsymbol{X}\otimes\boldsymbol{\lambda}\right]\right)$
- $\pi(\boldsymbol{\lambda}|rest) \sim \text{MVN}_N(\boldsymbol{\mu}_\lambda, \boldsymbol{\Sigma}_\lambda)$, where $\boldsymbol{\Sigma}_\lambda^{-1} = \sigma^{-2}(\sum_{t=1}^T X_t^2)[\boldsymbol{M}-\rho\boldsymbol{A}] + \sigma_\lambda^{-2}[\boldsymbol{M}-\rho_\lambda\boldsymbol{A}]$ and $\boldsymbol{\mu}_\lambda = \sigma^{-2}\boldsymbol{\Sigma}_\lambda([\boldsymbol{M}-\rho\boldsymbol{A}][\sum_{t=1}^T X_t(\boldsymbol{Y}_t - \boldsymbol{\beta}^*\boldsymbol{B}_t)])$
- $\pi(X_t|rest) \overset{\text{Indep}}{\sim} \text{Truncated-N}_{(0,\infty)}\left(\frac{[\boldsymbol{Y}_t-\boldsymbol{\beta}^*\boldsymbol{B}_t]'[\boldsymbol{M}-\rho\boldsymbol{A}]\boldsymbol{\lambda}}{1+\boldsymbol{\lambda}'[\boldsymbol{M}-\rho\boldsymbol{A}]\boldsymbol{\lambda}}, \frac{\sigma^2}{1+\boldsymbol{\lambda}'[\boldsymbol{M}-\rho\boldsymbol{A}]\boldsymbol{\lambda}}\right)$
- $\pi(\sigma^2|\text{rest}) \sim \text{Inverse-Gamma}(0.01 + (N+1)T/2, 0.01 + 0.5\boldsymbol{\varepsilon}'(\boldsymbol{I}_T\otimes[\boldsymbol{M}-\rho_\lambda\boldsymbol{A}])\boldsymbol{\varepsilon} + 0.5\sum_{t=1}^T X_t^2)$, where $\boldsymbol{\varepsilon} = \boldsymbol{Y} - [\boldsymbol{B}\otimes\boldsymbol{I}_N]\boldsymbol{\beta} - [\boldsymbol{X}\otimes\boldsymbol{\lambda}]$.
- $\pi(\sigma^2_\beta|\text{rest}) \sim \text{Inverse-Gamma}(0.01 + NK/2, 0.01 + \boldsymbol{\beta}'([\boldsymbol{M}-\rho\boldsymbol{A}]\otimes\boldsymbol{I}_k)\boldsymbol{\beta})$
- $\pi(\sigma^2_\lambda|\text{rest}) \sim \text{Inverse-Gamma}\left(0.01 + N/2, 0.01 + \boldsymbol{\lambda}'[\boldsymbol{M}-\rho_\lambda\boldsymbol{A}]\boldsymbol{\lambda}\right)$

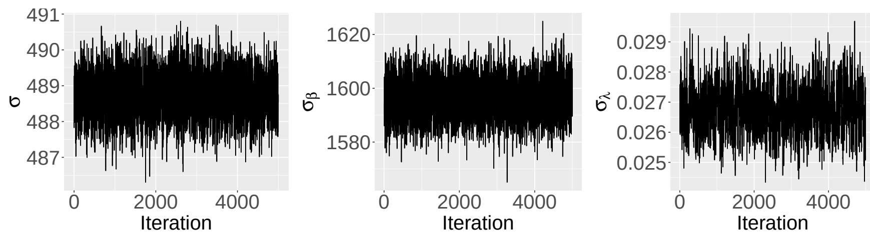

Fig. 4 Trace plots of the variance-related parameters and hyperparameters.

We run the Gibbs sampling for 10,000 iterations, and discard first 5,000 iterations as burn-in. The chains are monitored through trace plots; the post-burn-in samples for σ^2, σ_β^2 and σ_λ^2 are presented in Figure 4. All the panels show good convergence and mixing of the Gibbs sampler. The computing time is 402 minutes on a desktop with AMD Ryzen 9 5900x processor and 64GB RAM.

5 Results and Discussions

We calculate the posterior mean and standard deviation (SD) profiles of the annual ROC in monsoon rainfall based on the Markov chain Monte Carlo (MCMC) samples obtained in Section 4. We calculate the T-statistic, i.e., the ratio of posterior mean ROC and the corresponding posterior SD, for every grid cell; an absolute T-statistic value larger than two indicates significant changes and its sign indicates whether the change is positive or negative. We present the spatial maps of the posterior means and the T-statistics for ROC in Figure 5. For a total of 410 grid cells, the T-statistic turns out to be significantly negative, while it is significantly positive for 457 grid cells. Comparing with the empirical ROC estimates presented in the left panel of Figure 2, we observe that the posterior mean and the empirical estimates are close to each other, indicating a good model fit. From the plot for T-statistics, we observe that the estimated ROC is significantly negative for parts of Haryana and Punjab, some parts around the lower Gangetic River basin as well as in Arunachal Pradesh and Assam. This is concerning as a significant portion of these regions lies among the most fertile and rainfed regions of India. Besides, the estimated ROC is significantly positive, mainly for parts of Madhya Pradesh, Gujarat, Uttarakhand, and Meghalaya.

Further, the spatial maps of the posterior mean and T-statistic of the skewness-related parameter $\boldsymbol{\lambda}$ (not shown) show significant positive values near the Western Ghats. The posterior mean and SD of σ are 488.72 mm and 0.63 mm, respectively. For σ_β, they are 1596.44 mm and 7.32 mm, respectively, while for σ_λ, they are 0.03 mm and 7×10^{-4} mm, respectively. The posterior median, 0.025^{th} and 0.975^{th} quantiles of the random effects X_t are presented in Figure 6.

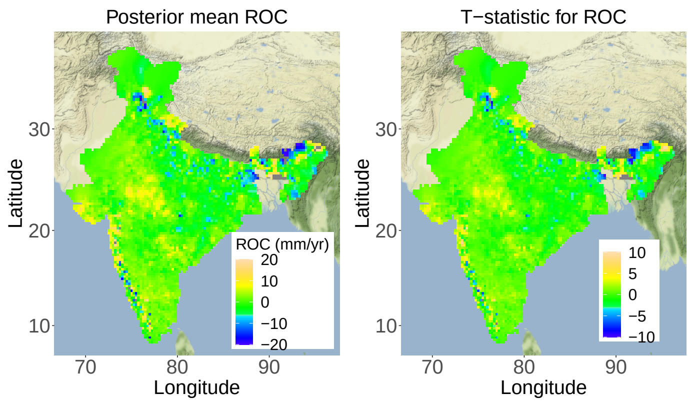

Fig. 5 Posterior mean profile the annual rate of change (ROC) in monsoon rainfall (left) and the corresponding T-statistics values across the grid cells (right).

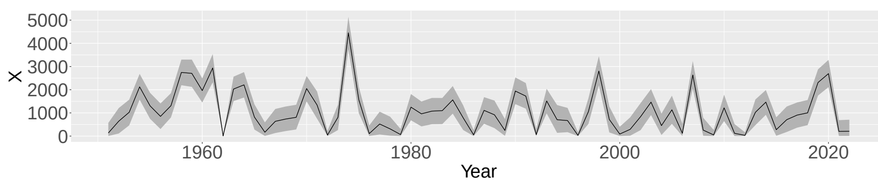

Fig. 6 Posterior median (solid line), and the 0.025-th and 0.975-th quantiles (grey ribbon) of the random effects X_t. A high value (during 1974, for example) indicates an event of country-wide higher rainfall on average.

6 Conclusions

We present a novel Bayesian model that analyzes spatially-varying temporal patterns using a high-dimensional skew-Gaussian CAR model. Our model, initially motivated by Indian monsoon rainfall features, is also applicable for analyzing long-term trends in other climatological variables. Traditional linear regression models with only year as a covariate are unsuitable for meteorological parameters due to multiple linear or nonlinear, micro-scale or macro-scale factors (e.g., El Nino); however, data for all such potential covariates are often unavailable. To address this, we use B-splines in a semiparametric Bayesian regression framework to model the mean component, which enables the model to explain local temporal behaviors in a data-driven approach. Additionally, our model assumes spatially-varying spline and skewness coefficients. The model allows posterior inference using Gibbs sampling. However, our model requires spatial autocorrelation-related parameters and hyperparameters to be pre-fixed due to computational constraints. For an exact Bayesian inference, one possible remedy is discretizing the parameter space of the spatial autocorrelation-related parameters, pre-calculating the necessary quadratic forms and determinants of the final covariance matrices over the parameter space, and using the stored information within MCMC; such an idea is explored by [15]. Future endeavors would be to extend our model to fit winter rainfall data with a large proportion of nil rainfall cases and explore faster computational schemes like Max-and-Smooth [9].

From an application perspective, analyzing the impact of climate change on rainfall is crucial for agriculture, water resource management, and disaster management. Our results highlight a significant negative trend in various parts of the Gangetic River basin, where agriculture relies heavily on rainwater availability.

Acknowledgements The authors would like to thank Laleh Tafakori from RMIT University, Australia, and an anonymous reviewer for their several important suggestions, which improved the flow and the content of the paper. The research of the second author is partially supported by the Startup Research Grant, Science and Engineering Research Board, India, with Award No. SERB-MATH-2023632.

References

1. Azzalini, A.: The skew-normal and related families. Cambridge University Press, Cambridge, UK (2014)
2. Besag, J.: Spatial interaction and the statistical analysis of lattice systems. J. R. Stat. Soc. Ser. B Methodol. **36**, 192–225 (1974)
3. Chauhan, B. S., Mahajan, G., and Randhawa, R. K., Singh, H., Kang, M. S.: Global warming and its possible impact on agriculture in India. Adv. Agron. **123**, 65–121 (2014)
4. Christakos, G.: Random field models in earth sciences. Academic Press, San Diego, USA (1992)
5. Gelfand, A. E.: Gibbs sampling. J. Am. Stat. Assoc. **95**, 1300–1304 (2000)
6. Gelfand, A. E., Schliep, E. M.: Spatial statistics and Gaussian processes: A beautiful marriage. Spat. Stat. **18**, 86–104 (2016)
7. Hazra, A., Reich, B. J., Staicu, A-M: A multivariate spatial skew-t process for joint modeling of extreme precipitation indexes. Environmetrics. **31**, e2602 (2020)
8. Hazra, A., Huser, R.: Estimating high-resolution Red Sea surface temperature hotspots, using a low-rank semiparametric spatial model. Ann. Appl. Stat. **15**, 572–596 (2021)
9. Hazra, A., Huser, R., Jóhannesson, A. V.: Bayesian latent Gaussian models for high-dimensional spatial extremes. In: Statistical Modeling Using Bayesian Latent Gaussian Models: With Applications in Geophysics and Environmental Sciences. (2023)
10. Hazra, A., Huser, R., Bolin, D.: Efficient modeling of spatial extremes over large geographical domains. J. Comput. Graph. Stat. (2024) doi: 10.1080/10618600.2024.2409784
11. Katyaini, S., Barua, A.: Assessment of interstate virtual water flows embedded in agriculture to mitigate water scarcity in India (1996–2014). Water Resour. Res. **53**, 7382–7400 (2017)
12. Mardia, K. V.: Measures of multivariate skewness and kurtosis with applications. Biometrika. **57**, 519–530 (1970)
13. Mastrantonio, G., Gelfand, A. E., Jona Lasinio, G.: The wrapped skew Gaussian process for analyzing spatio-temporal data. Stoch. Environ. Res. Risk Assess. **30**, 2231–2242 (2016)
14. Olea, R. A.: Geostatistics for engineers and earth scientists. Springer, New York, USA (1999)
15. Pakrashi, A., Hazra, A., Raveendran, S. M., Balakrishnan, K.: Approximate Bayesian inference for high-resolution spatial disaggregation using alternative data sources. J. Agric. Biol. Environ. Stat. (2025) doi: 10.1007/s13253-025-00695-5
16. Pai, D. S., Rajeevan, M., Sreejith, O. P., Mukhopadhyay, B., Satbha, N. S.: 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 (2014)
17. Rue, H., Held, L.: Gaussian Markov random fields: theory and applications. CRC Press, New York, USA (2005)
18. Venkateswarlu, B.: Rainfed agriculture in India: issues in technology development and transfer. In: Model training course on “impact of climate change in rainfed agriculture and adaptation strategies” (2011) via ICAR-Central Research Institute for Dryland Agriculture.
19. Zhang, H., El-Shaarawi, A.: On spatial skew-Gaussian processes and applications. Environmetrics. **21**, 33–47 (2010)