A Bayesian Hierarchical Network Model for Daily Streamflow Forecasting

Álvaro Ossandón[1], Balaji Rajagopalan[2], Upmanu Lall[3], Vimal Mishra[4], and Nanditha J S[5]

[1]University of Colorado Boulder
[2]University of Colorado at Boulder
[3]Columbia University
[4]Indian Institute of Technology Gandhinagar
[5]Indian Institute of Technology, Gandhinagar

November 23, 2022

Abstract

We developed a novel Bayesian Hierarchical Network Model (BHNM) for daily streamflow, which uses the spatial dependence induced by the river network topology, and average daily precipitation from the upstream contributing area between station gauges. In this, daily streamflow at each station is assumed to be distributed as Gamma distribution with temporal non-stationary parameters. The mean and standard deviation of the Gamma distribution for each day are modeled as a linear function of suitable covariates. The covariates include daily streamflow from upstream gauges or from the gauge above of the upstream gauges depending on the travel times, and daily, 2-day, or 3-day precipitation from the area between two stations that attempts to reflect the antecedent land conditions. Intercepts and slopes of the mean and standard deviation parameters are modeled as a Multivariate Normal distribution (MVN) to capture their dependence structure. To ensure that the covariance matrix of MVN is positive definite, it is model as an Inverse Wishart distribution. Non-informative priors for each parameter were considered. Using the network structure in incorporating flow information from upstream gauges and precipitation from the immediate contributing area as covariates, enables to capture the spatial correlation of flows simultaneously and parsimoniously. The posterior distribution of the model parameters and, consequently, the predictive posterior Gamma distribution of the daily streamflow at each station and for each day are obtained. The model is demonstrated by its application to daily summer (July-August) streamflow at 4 gauges in the Narmada basin network in central India for the period 1978 – 2014. The skill of the probabilistic forecast is carried out by rank histograms and the Continuous Ranked Probability Score (CRPS). The model validation indicates that the model is highly skillful relative to climatology and relative to a null-model of linear regression. The forecasts present an adequate spread of uncertainty and non-bias. Since flooding is of major concern in this basin, we applied the BHNM in a cross-validated mode on two high flooding years – in that, the model was fitted on other years, and forecasts were made for the dropped-out high flooding year. The skill of the model in forecasting the high flood events was very good across the network – in both the timing and magnitude of the events. The model will be of immense help to policy makers in risk-based flood mitigation. The BHNM framework is general in nature and can be applied to any river network with other covariates as appropriate.

A Bayesian Hierarchical Network Model for Daily Streamflow Forecasting

Álvaro Ossandón (1,2), Balaji Rajagopalan (1,3), Upmanu Lall (4), Vimal Mishra (5), and J. S. Nanditha (5)

(1) CEAE, University of Colorado, Boulder CO, (2) Department of Civil Engineering, Santa Maria University,Valparaiso, Chile, (3) CIRES, University of Colorado, Boulder CO, (4) Department of Earth andEnvironmental Engineering, Columbia Water Center, The Earth Institute, Columbia University, New York,NY, USA, (5) Civil Engineering, Indian Institute of Technology, Gandhinagar, India





PRESENTED AT:

AGU FALL MEETING
Online Everywhere | 1–17 December 2020

INTRODUCTION

- In India, Riverine floods are the major cause of the destruction of property and loss of life, each year.
- The floods occur mostly during the summer monsoon season of June - September when more than 80% of annual rainfall arrives over India. The extreme rainfall events which produce the floods are a result of synoptic-scale cyclonic depressions.
- While forecast of precipitation is increasingly becoming skillful, forecasts of streamflow and consequently, floods, are not skillful and vary widely across River Basins. This need motivates the proposed research.
- We propose a Novel Bayesian Hierarchical Network Model (BHNM) for daily streamflow forecast, which uses the spatial dependencies induced by the river network topology, and antecedent hydroclimate information from upstream. The hierarchical aspect and the Bayesian framework, Together it captures the spatial correlation in the streamflow on the river network and provides robust estimates of uncertainties.
- The Narmada River Basin in West Central India is used as a testbed to develop and demonstrate this model.

STUDY REGION AND DATA

Narmada Basin, India

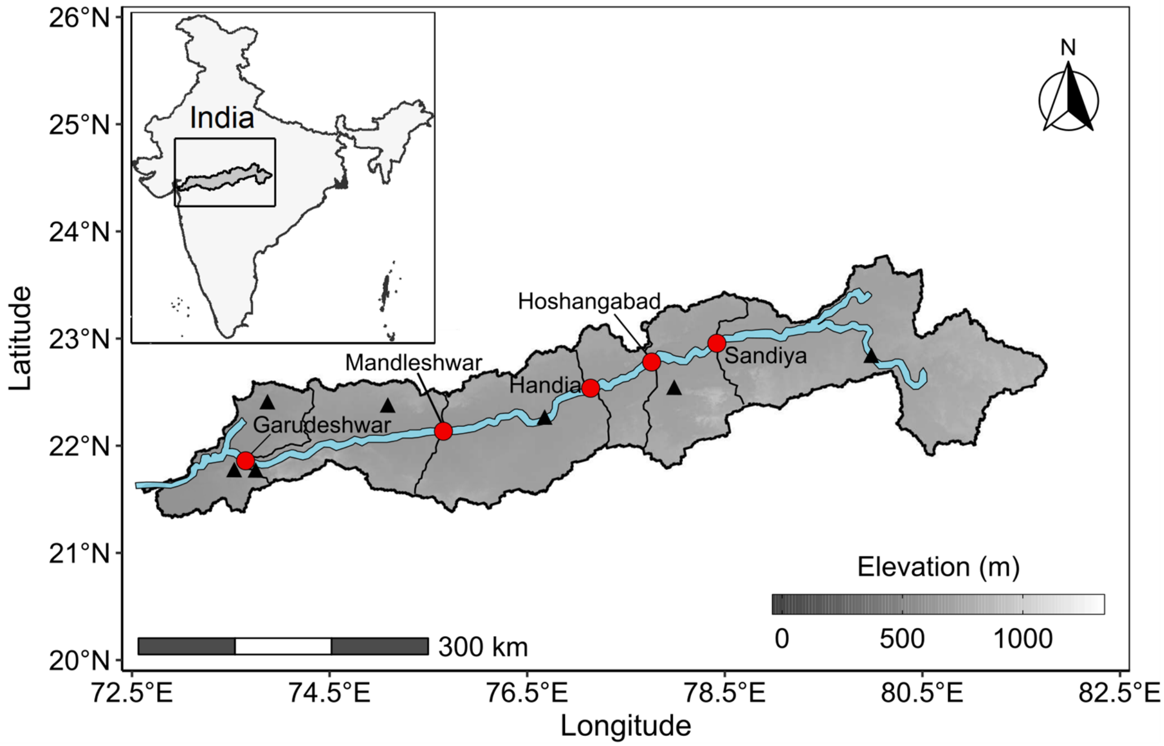

Figure 1. Map of the Narmada basin boundary in India showing the digital elevation model of the basin (SRTM DEM); the locations of five sub-basin outlets: Sandiya, Hoshangabad, Handia, Mandleshwar and Garudeshwar; and some of the major dams in the basin are marked: Bargi, Tawa, Indirasagar, Jobat, and Sardar Sarovar (from upstream to downstream direction).

- Narmada River originates from the Amarkantak hills in Madhya Pradesh and drains into the Gulf of Cambay in the Arabian Sea, flowing from the east to west direction.
- Narmada is the fifth largest river in India and the largest west flowing river in the country.
- The Narmada River basin has an area of 98,796 km, and it extends 953 km in the east-west direction.

Data

Streamflow

- Observed daily summer (July-August) streamflow at four gauge stations in the Narmada basin: Sandiya, Handia, Hoshangabad, and Mandleshwar were obtained from India Water Resource Information System (IWRIS) (*Figure 1*)
- Period 1978 – 2014
- Garudeshwar gauge station was not considered in this study since it had longer missing periods.

Hydro-Meteorological Variable

- Gridded daily summer (July-August) precipitation from the India Meteorology Department (IMD)
- 0.25° spatial resolution
- Period 1978 – 2014

MODEL STRUCTURE

Covariates

As potential covariates, we considered:

- Daily streamflow from upstream gauges or from the gauge above of the upstream gauges depending on effective travel times
- 1-day, 2-day, or 3-day spatial average precipitation from the area between two stations that attempts to reflect the antecedent land conditions.
- These variables are considered at lag -1 day, i.e., we have 1-day lead time for the forecast.
- The best set of covariates for each station gauges were obtained based on the highest linear correlation coefficient, R (*Figure 2*).

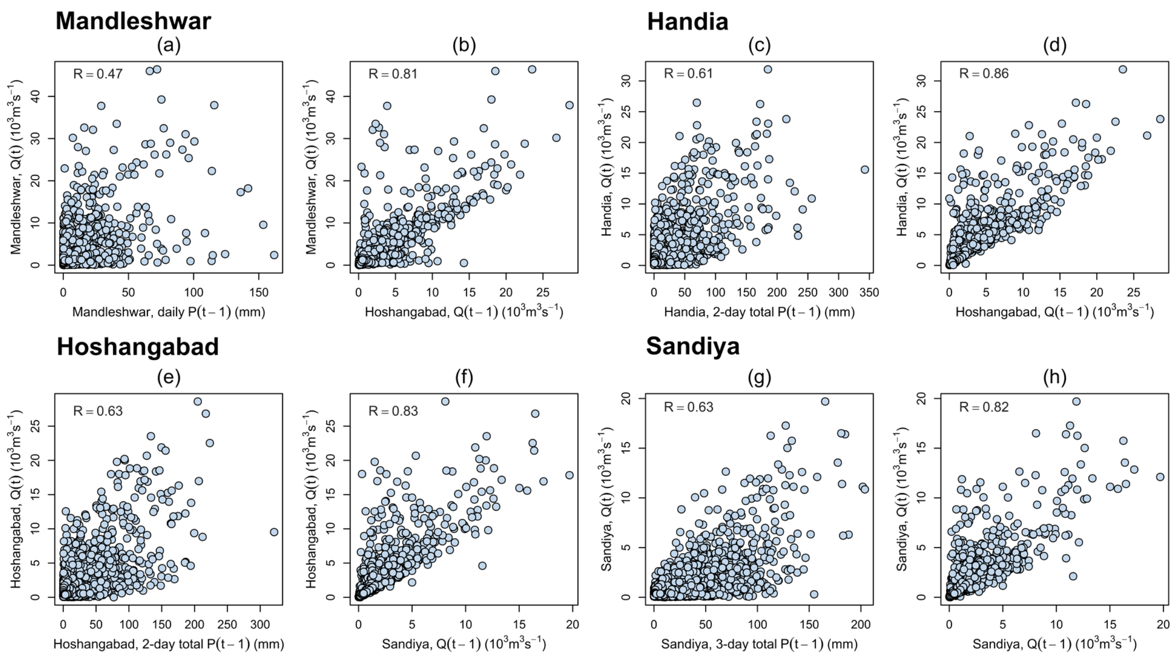

Figure 2. Scatter plots of daily streamflow vs. lag -1 day covariates selected for each station gauge: Mandleshwar streamflow vs. (a) daily spatial average precipitation, (b) and daily Hoshangabad streamflow; Handia streamflow vs. (c) 2-day spatial average precipitation, (d) and daily Hoshangabad streamflow; Hoshangabad streamflow vs. (e) 2-day spatial average precipitation, (f) and daily Sandiya streamflow; Sandiya streamflow vs. (g) 3-day spatial average precipitation, (h) and lag -1 day daily Sandiya streamflow. All Pearson correlation coefficients, R, are significant (P-value<0.1).

Model Structure for Narmada Basin

For the structure of the *Bayesian Hierarchical Network Model* (BHNM) for the Narmada basin, we considered that streamflow at each gauge station follows a gamma distribution. *Figure 3* displays the conceptual sketch of the network Bayesian model implemented here.

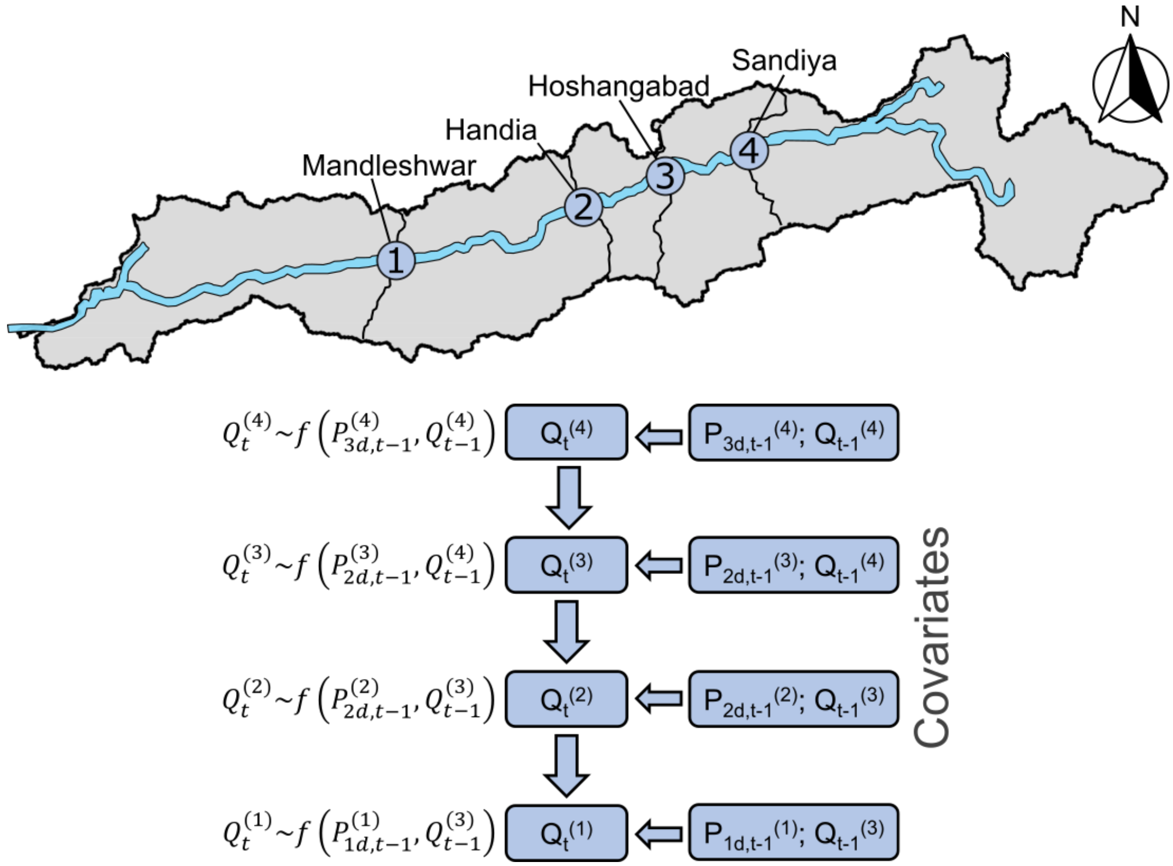

Figure 3. Conceptual sketch of the network Bayesian model for the Narmada basin. $Q_t^{(i)}$ corresponds to the observed streamflow at gauge *i* and day *t*, and $P_{xd,t-1}^{(i)}$ to x-day spatial average precipitation from the area between stations *i* and *i+1* at day *t-1*.

We incorporated the covariates showed in *Figure 2*, which give the model structure showed in *Figure 3* and represented by the following equations

$$Q_t^{(i)} \sim Gamma\left(r_t^{(i)}, \gamma_t^{(i)}\right) \quad i = 1,2,3,4$$

$$\gamma_t^{(i)} = \frac{\mu_t^{(i)}}{\left(\sigma_t^{(i)}\right)^2}; \quad r_t^{(i)} = \frac{\left(\mu_t^{(i)}\right)^2}{\left(\sigma_t^{(i)}\right)^2}; \quad i = 1,2,3,4$$

Mandleshwar:

$$\mu_t^{(1)} = \beta_0^{(1)} + \beta_1^{(1)} P_{1d,t-1}^{(1)} + \beta_2^{(1)} Q_{t-1}^{(3)}$$
$$\sigma_t^{(1)} = \phi_0^{(1)} + \phi_1^{(1)} P_{1d,t-1}^{(1)} + \phi_2^{(1)} Q_{t-1}^{(3)}$$

Handia:

$$\mu_t^{(2)} = \beta_0^{(2)} + \beta_1^{(2)} P_{2d,t-1}^{(2)} + \beta_2^{(2)} Q_{t-1}^{(3)}$$

$$\sigma_t^{(2)} = \phi_0^{(2)} + \phi_1^{(2)} P_{2d,t-1}^{(2)} + \phi_2^{(2)} Q_{t-1}^{(3)}$$

Hoshangabad:

$$\mu_t^{(3)} = \beta_0^{(3)} + \beta_1^{(3)} P_{2d,t-1}^{(3)} + \beta_2^{(3)} Q_{t-1}^{(4)}$$

$$\sigma_t^{(3)} = \phi_0^{(3)} + \phi_1^{(3)} P_{2d,t-1}^{(3)} + \phi_2^{(3)} Q_{t-1}^{(4)}$$

Sandiya:

$$\mu_t^{(4)} = \beta_0^{(4)} + \beta_1^{(4)} P_{3d,t-1}^{(4)} + \beta_2^{(4)} Q_{t-1}^{(4)}$$

$$\sigma_t^{(4)} = \phi_0^{(4)} + \phi_1^{(4)} P_{3d,t-1}^{(4)} + \phi_2^{(4)} Q_{t-1}^{(4)}$$

- Posterior distributions of the parameters and streamflow (ensembles) were estimated using the Gibbs sampling algorithm for the Markov Chain Monte Carlo method.
- The priors of $\boldsymbol{\beta}^{(i)}$ and $\boldsymbol{\phi}^{(i)}$ for each gauge station were considered Multivariate Normal distribution (MVN) to capture their dependence structure.

RESULTS

Calibration

3000 simulations from posterior distributions of the model parameters, and consequently, streamflow ensembles were obtained.

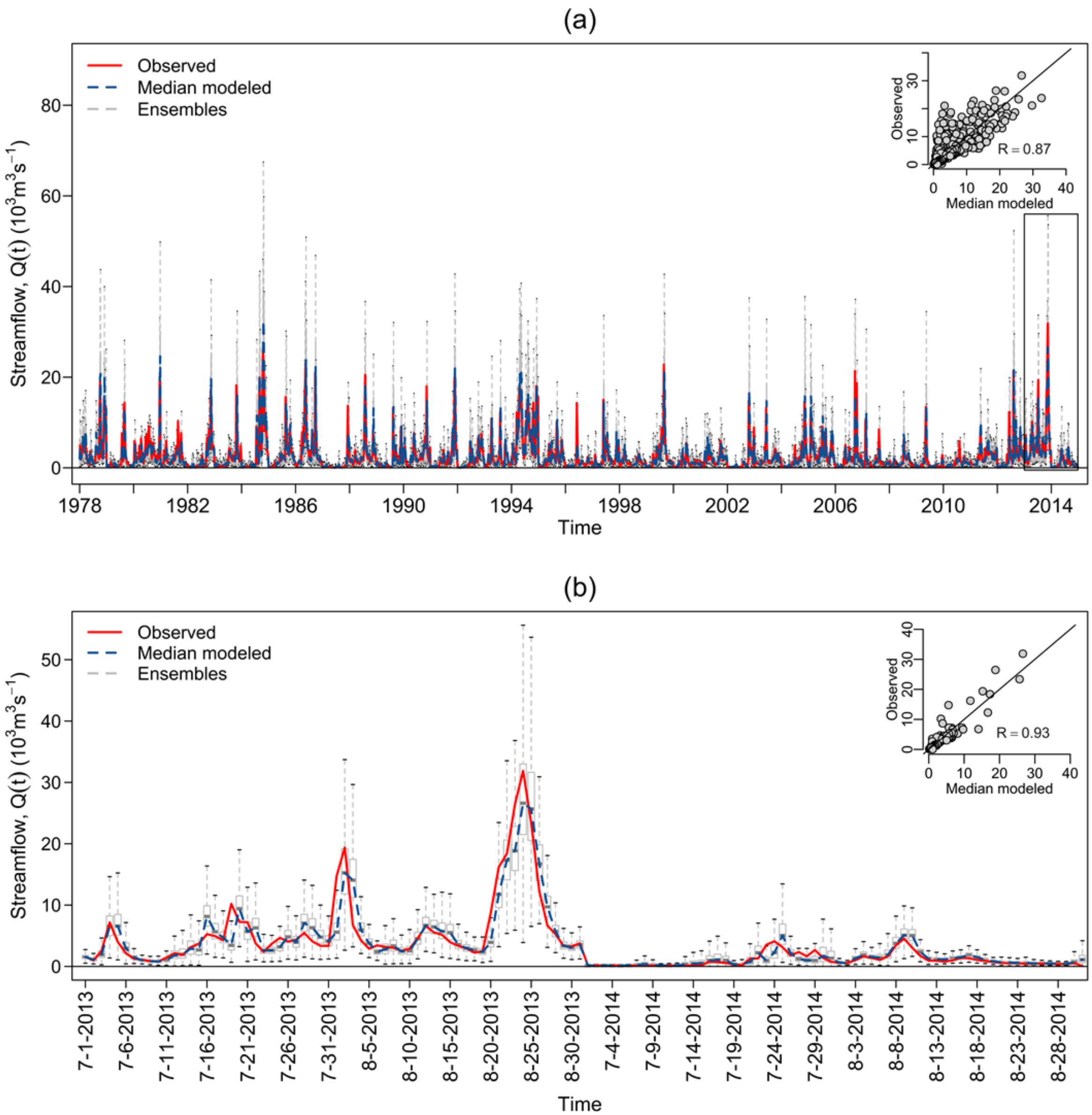

Figure 4. Ensembles of simulated July August daily streamflow for the Handia gauge station presented as boxplot time series for (a) entire record (1978-2014) and (b) 2013-2014. The boxplots represent the posterior distribution estimates of the daily streamflow. Red lines correspond to the observed daily streamflow and blue-dashed lines to the posterior median daily streamflow. The medians of these boxplots/distributions are considered to be the actual simulated values when computing R, which are displayed on the scatter plots on the upper right of each panel. R values are significant (P value<0.1). The black box in panel a shows the temporal windows for time series in panel b.

- All the observed values are captured by the ensembles variability
- The timing of the streamflow peaks is captured by the ensembles
- The performance for high flow years is even better (R values).

Cross-Validation

We applied the BHNM in a cross-validated mode on two high flooding years – in that, the model was fitted on other years, and forecasts were made for the dropped-out high flooding years. *Figure 5* shows the four two high flooding years validation periods considered.

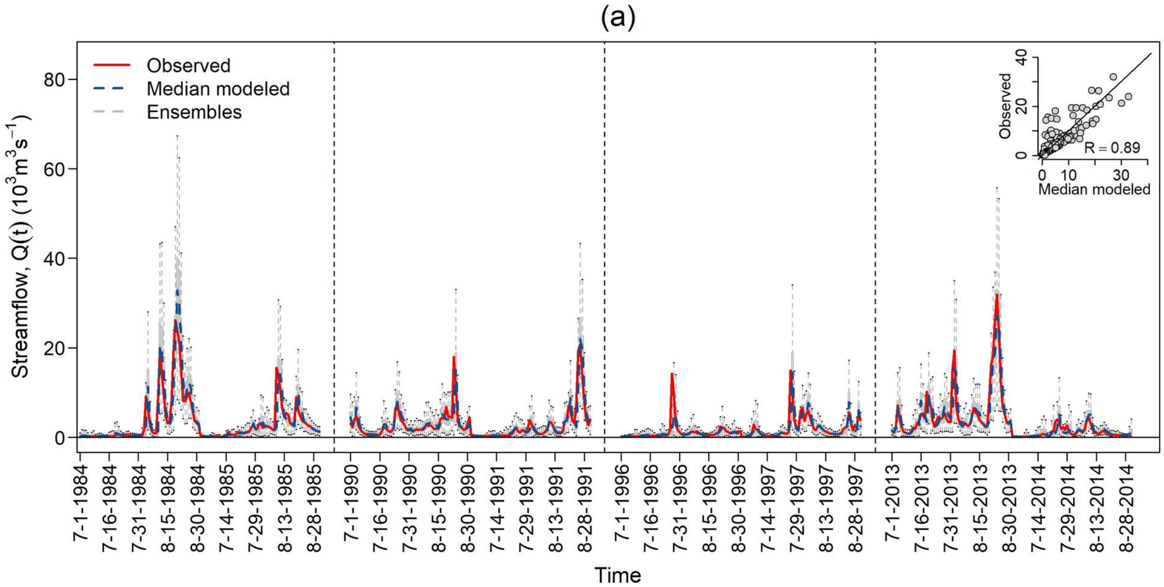

Figure 5. Ensembles forecast of July-August daily streamflow presented as boxplot time series for the four validation periods (1984 1985, 1990-1991, 1996-1997, and 2013 2014) at Handia gauge station. The boxplots represent the posterior distribution estimates of the daily streamflow. Red lines correspond to the observed daily streamflow and blue-dashed lines to the posterior median daily streamflow. The medians of these boxplots/distributions are considered to be the actual forecast values when computing R, which are displayed on the scatter plots on the upper right of each panel. R values are significant (P value<0.1). black-dashed vertical lines indicate the division between validation periods.

- Most of the observed values are captured by the ensembles forecast variability
- The correlation obtained is higher than the one for the whole calibration period.

Statistical Consistency

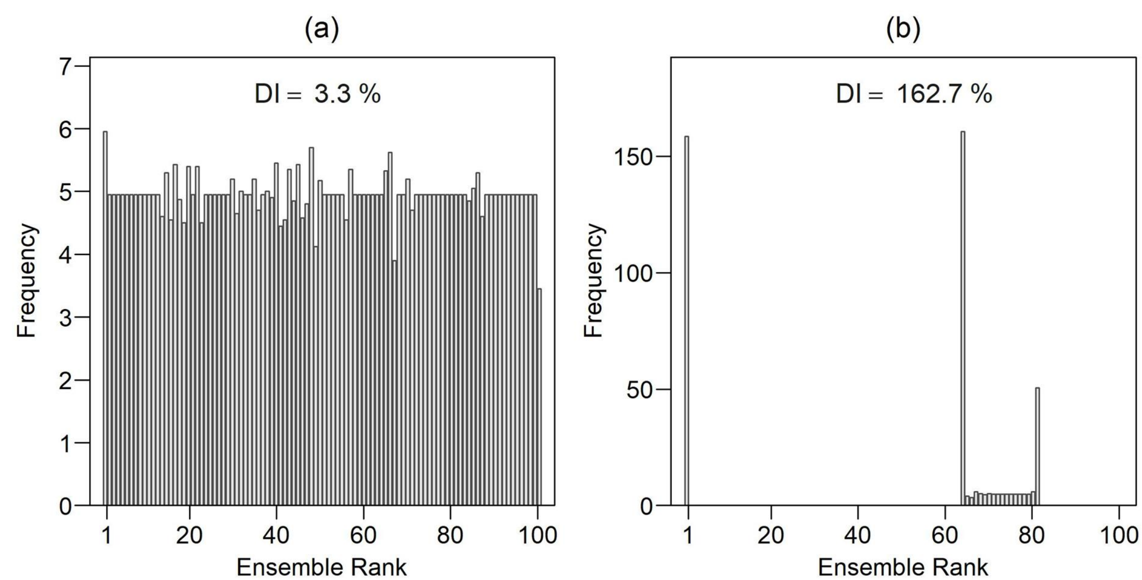

Figure 6. Rank histograms of the ensembles forecast of July-August daily streamflow during cross-validation periods for (a) the Bayesian Hierarchical Network Model (BHNM) and (b) the Linear Model (LM) at Handia gauge station. DI denotes the discrepancy index.

- A better spread is generated using the ensembles forecast of the BHNM since its rank histogram of the BHNM is almost uniform (non-bias) and shows low DI values
- The U-shaped rank histogram for the Linear Model (LM) indicates a lack of variability in the ensembles.

At Site Probabilistic Skill

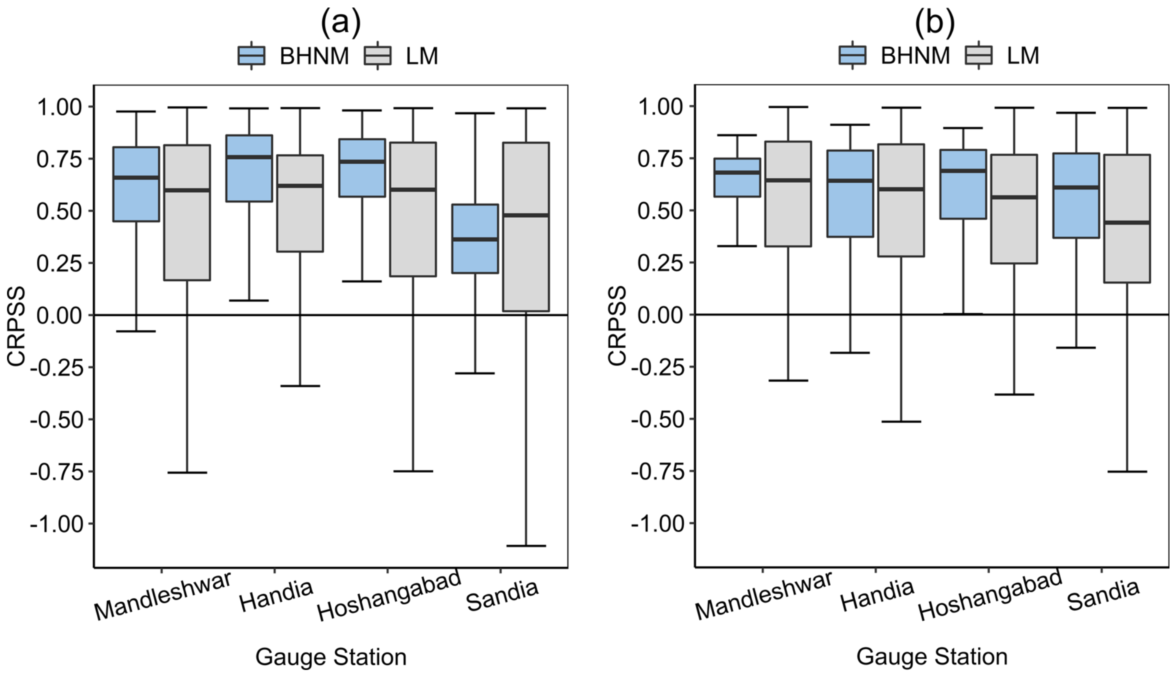

Figure 7. The cross-validation distributions for the continuous rank probability score (CRPSS) statistic of BHNM (sky blue boxes) and LM (gray boxes) models for (a) 496 days of the four validation periods and (b) days with high flows. Climatology was considered as the reference forecast model. CRPSS values above zero and closest to one indicates a better skill.

- For both models, and BHNM presents better performance than LM and climatology with the exception of Sandiya (median of the distribution is lower for BHNM compared to LM, *Figure 7a*)
- For high flow days, BHNM presents a better overall performance than LM and climatology for all the gauges (*Figure 7b*).

CONCLUSIONS

The proposed Bayesian Hierarchical Network Model has benefits when is compared to stationary, at site Bayesian and non-Bayesian models:

- By incorporating flow information from upstream gauges and precipitation from the immediate contributing area as covariates, enables to capture the spatial correlation of flows simultaneously and parsimoniously.
- Can be applied to basins with non-natural flow regimes since by incorporating the right gauge feeder, the effect of some human interventions such as dams can be replicated by the model.
- It is not as computationally exhaustive as other models that consider the spatial correlation of the flow.

ACKNOWLEDGEMENTS

This project was funded by the Monsoon Mission project of the Ministry of Earth Sciences, India. We also acknowledge the support from a Fulbright fellowship and CONICYT PFCHA/DOCTORADO BECAS CHILE/2015-56150013 to the first author.

सत्यमेव जयते
Ministry of Earth Sciences
Government of India

FULBRIGHT

BECAS CHILE

ABSTRACT

We developed a novel Bayesian Hierarchical Network Model (BHNM) for daily streamflow, which uses the spatial dependence induced by the river network topology, and average daily precipitation from the upstream contributing area between station gauges. In this, daily streamflow at each station is assumed to be distributed as Gamma distribution with temporal non-stationary parameters. The mean and standard deviation of the Gamma distribution for each day are modeled as a linear function of suitable covariates. The covariates include daily streamflow from upstream gauges or from the gauge above of the upstream gauges depending on the travel times, and daily, 2-day, or 3-day precipitation from the area between two stations that attempts to reflect the antecedent land conditions. Intercepts and slopes of the mean and standard deviation parameters are modeled as a Multivariate Normal distribution (MVN) to capture their dependence structure. To ensure that the covariance matrix of MVN is positive definite, it is model as an Inverse Wishart distribution. Non-informative priors for each parameter were considered. Using the network structure in incorporating flow information from upstream gauges and precipitation from the immediate contributing area as covariates, enables to capture the spatial correlation of flows simultaneously and parsimoniously. The posterior distribution of the model parameters and, consequently, the predictive posterior Gamma distribution of the daily streamflow at each station and for each day are obtained. The model is demonstrated by its application to daily summer (July-August) streamflow at 4 gauges in the Narmada basin network in central India for the period 1978 – 2014. The skill of the probabilistic forecast is carried out by rank histograms and the Continuous Ranked Probability Score (CRPS). The model validation indicates that the model is highly skillful relative to climatology and relative to a null-model of linear regression. The forecasts present an adequate spread of uncertainty and non-bias. Since flooding is of major concern in this basin, we applied the BHNM in a cross-validated mode on two high flooding years – in that, the model was fitted on other years, and forecasts were made for the dropped-out high flooding year. The skill of the model in forecasting the high flood events was very good across the network – in both the timing and magnitude of the events. The model will be of immense help to policy makers in risk-based flood mitigation. The BHNM framework is general in nature and can be applied to any river network with other covariates as appropriate.

REFERENCES

Ali, H., Modi, P., & Mishra, V. (2019). Increased flood risk in Indian sub-continent under the warming climate. Weather and Climate Extremes, 25, 100212. https://doi.org/10.1016/j.wace.2019.100212 (https://doi.org/10.1016/j.wace.2019.100212)

Banerjee, R., & Trust, A. (2009). Review of water governance in the Narmada river basin. Delle Monache, L., Hacker, J. P., Zhou, Y., Deng, X., & Stull, R. B. (2006). Probabilistic aspects of meteorological and ozone regional ensemble forecasts. Journal of Geophysical Research, 111(D24), D24307.

Garg, S., & Mishra, V. (2019). Role of Extreme Precipitation and Initial Hydrologic Conditions on Floods in Godavari River Basin, India. Water Resources Research, 55(11), 9191–9210. https://doi.org/10.1029/2019WR025863 (https://doi.org/10.1029/2019WR025863)

Gelman, A., & Hill, J. (2006). Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511790942 (https://doi.org/10.1017/CBO9780511790942)

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 (https://doi.org/10.1214/ss/1177011136)

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association, 102(477), 359–378. https://doi.org/10.1198/016214506000001437 (https://doi.org/10.1198/016214506000001437)

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 (https://doi.org/10.1007/s11749-008-0114-x)

Hamill, T. M. (2001). Interpretation of rank histograms for verifying ensemble forecasts. Monthly Weather Review, 129(3), 550–560. https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2 (https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2)

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather and Forecasting, 15(5), 559–570. https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2 (https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2)

Jensen, F. V., & Nielsen, T. D. (2007). Bayesian Networks and Decision Graphs. New York, NY: Springer New York. https://doi.org/10.1007/978-0-387-68282-2 (https://doi.org/10.1007/978-0-387-68282-2)

Mendoza, P. A., Rajagopalan, B., Clark, M. P., Ikeda, K., & Rasmussen, R. M. (2015). Statistical postprocessing of high-resolution regional climate model output. Monthly Weather Review, 143(5), 1533–1553. https://doi.org/10.1175/MWR-D-14-00159.1 (https://doi.org/10.1175/MWR-D-14-00159.1)

Pai, D., Sridhar, L., Rajeevan, M., Sreejith, O. P., Satbhai, N. S., & Mukhopadhyay, B. (2014). Development of a new high spatial resolution (0.25° × 0.25°) long period (1901-2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. MAUSAM, 65(1), 18.

Plummer, M. (2003). Proceedings of the 3rd international workshop on distributed statistical computing. JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling, 124(125.10), 1–10.

Plummer, M. (2019). rjags: Bayesian graphical models using MCMC. R Package Version, 4(10), 19.

Ravindranath, A., Devineni, N., Lall, U., Cook, E. R., Pederson, G., Martin, J., & Woodhouse, C. (2019). Streamflow Reconstruction in the Upper Missouri River Basin Using a Novel Bayesian Network Model. Water Resources Research, 55(9), 7694–7716. https://doi.org/10.1029/2019WR024901 (https://doi.org/10.1029/2019WR024901)

Robert, C., & Casella, G. (2011). A short history of Markov Chain Monte Carlo: Subjective recollections from incomplete data. Statistical Science, 26(1), 102–115. https://doi.org/10.1214/10-STS351 (https://doi.org/10.1214/10-STS351)

Shah, R. D., & Mishra, V. (2016). Utility of global ensemble forecast system (GEFS) reforecast for medium-range drought prediction in India. Journal of Hydrometeorology, 17(6), 1781–1800. https://doi.org/10.1175/JHM-D-15-0050.1 (https://doi.org/10.1175/JHM-D-15-0050.1)

Team, R. C. (2017). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Wilks, & Daniel. (2011). Statistical Methods in the Atmospheric Sciences, Volume 100 - 3rd Edition. Academic Press Inc.