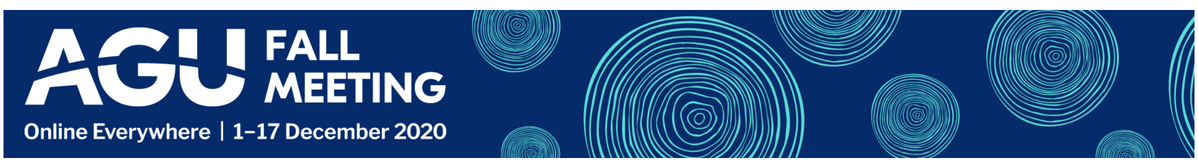
A Coupled Physical-Statistical Model for Daily Streamflow Forecasting

Nanditha J. S. , Álvaro Ossandón, Balaji Rajagopalan, and Vimal Mishra

Civil Engineering, Indian Institute of Technology, Gandhinagar, India
Department of Civil, Environmental and Architectural Engineering, University of Colorado at Boulder, Boulder, CO, USA
Departamento de Obras Civiles, Universidad Técnica Santa María, Valparaíso, Chile
CIRES, University of Colorado at Boulder, Boulder, CO, USA




PRESENTED AT:



INTRODUCTION

Improving the streamflow forecast skill is of huge importance in flood-prone river basins, especially basins driven by extreme rainfall events. Here, we present a coupled physical-statistical model for modelling flood events in the Narmada basin in India.

Physical hydrological models suffer from model parameterization and regionalization errors, which may result in biases in the model output. One needs to adopt a combination of post-processing methods and data assimilation to overcome these errors. Here, we combine the outputs of a semi-distributed hydrological model with a *Bayesian Hierarchical Network Model (BHNM)*.

The BHNM models the river network as a single system and uses precipitation and physical model outputs as potential covariates to improve streamflow skill. In addition, the Markov Chain Monte Carlo (MCMC) approach helps in reducing parameter uncertainty by providing an ensemble forecast.

STUDY REGION AND DATA

Narmada Basin, India

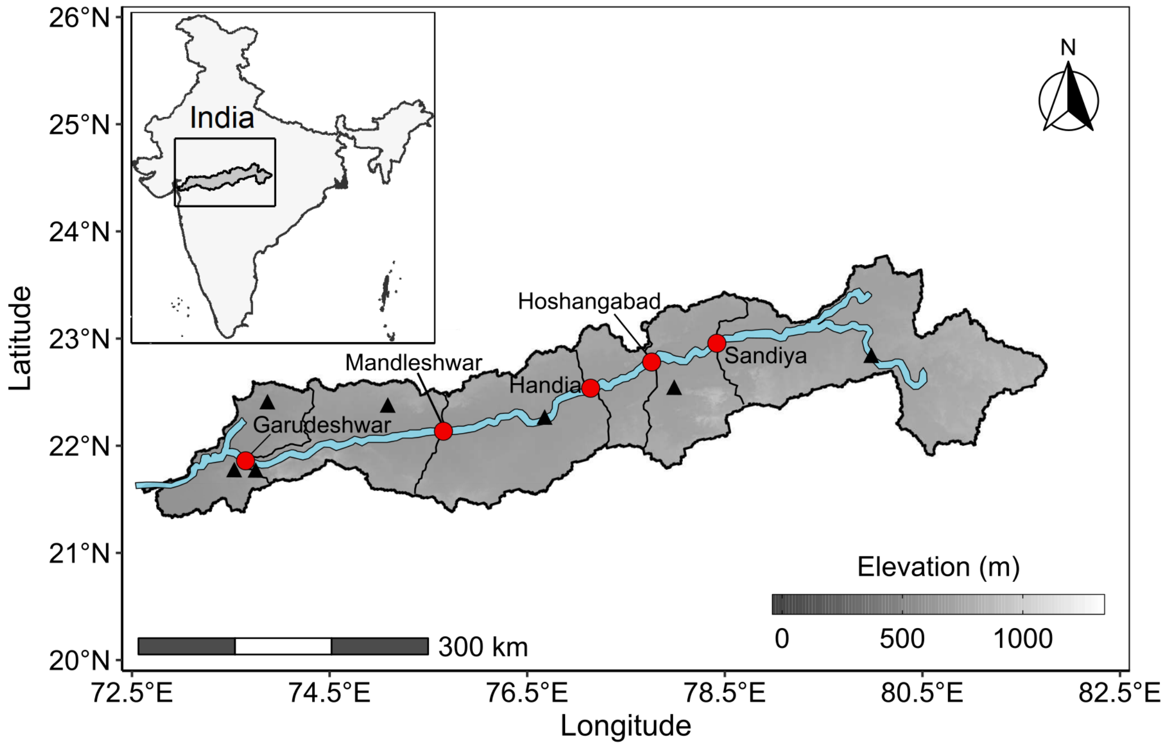


Figure 1. The Narmada basin in India. The figure shows the digital elevation model of the basin (SRTM DEM); the locations of five sub-basin outlets: Sandiya, Hoshangabad, Handia, Mandleshwar and Garudeshwar; and some of the major dams in the basin: Bargi, Tawa, Indirasagar, Jobat, and Sardar Sarovar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

- Narmada is a west-flowing river that originates from the Amarkantak ranges in Madhya Pradesh.
- The basin has an area of 98,796 km^2, and it extends 953 km in the east-west direction.

Discharge Data

- We obtained daily observed summer discharge data from 1978-2014 at five gauge stations in the Narmada basin: **Sandiya, Handia, Hoshangabad, Mandleshwar and Garudeshwar** from India Water Resource Information System (IWRIS).

Meteorological Variables

- We used 0.25° gridded daily precipitation, temperature and wind data from 1951-2018 from the India Meteorological Department (IMD) (Pai et al., 2014; Srivastava et al., 2009).

VIC Hydrological Model

- We used the Variable Infiltration Capacity (VIC) hydrological model (Liang et al., 1994, 1996) at 0.25° resolution for modelling the Narmada basin.
- VIC is a semi-distributed hydrological model that solves water and energy balance at each grid cell.
- It generates 3-layer soil moisture and streamflow values at each grid point.
- We used a routing model (Lohmann et al., 1996) to simulate daily streamflow values at specific locations.
- We used a reservoir module to obtain naturalized observations to calibrate the VIC model.

METHODS

VIC model calibration

We calibrated the VIC model at three gauge stations using naturalized observations.

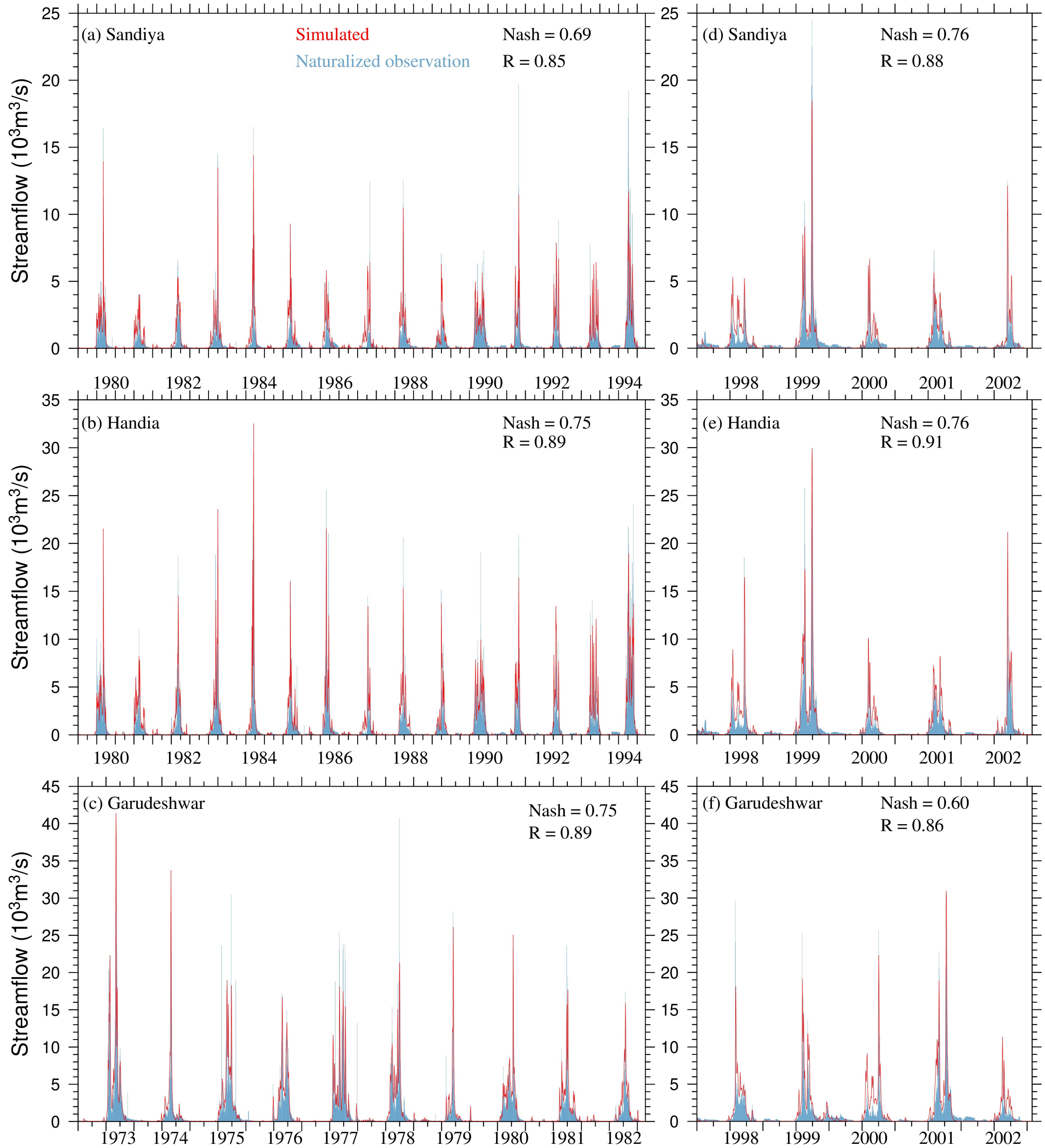


Figure 2. Left panel (a-c) shows the result of calibration of the VIC model from 1980 to 1994 for the stations Sandiya, Handia and Garudeshwar (1973-1982). Right panel (d-f) shows the result of model validation from 1998 to 2002. The Nash Sutcliffe efficiency and correlation coefficient, R2 are given in each plot.

Potential Covariates

As potential covariates we considered:

- 1-day, 2-day, and 3-day spatial average **precipitation and soil moisture** from the area between two successive stations.
- VIC streamflow simulation at each gauge point.

- 1-day lagged VIC streamflow simulation at the upstream gauge.
- The product between VIC Streamflow simulation and 2-day spatial average precipitation from the area between two stations. This nonlinear variable allows correcting the nonsystematic bias and timing issue present in the VIC streamflow (*Figure 3*)

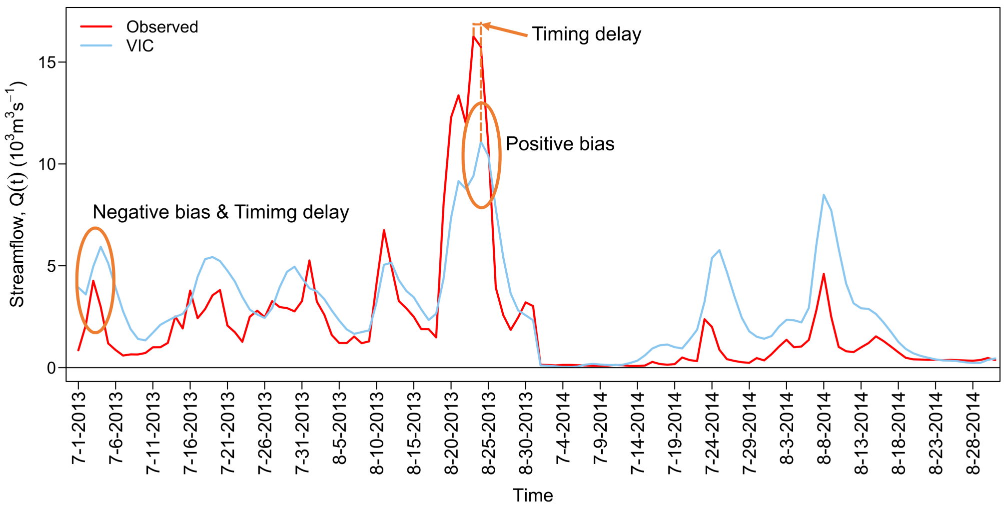


Figure 3. Nonsystematic bias and timing delay of the VIC streamflow at the Sandiya station gauge for 2013-2014.

Model Structure

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 4* displays the conceptual sketch of the network Bayesian model implemented here.

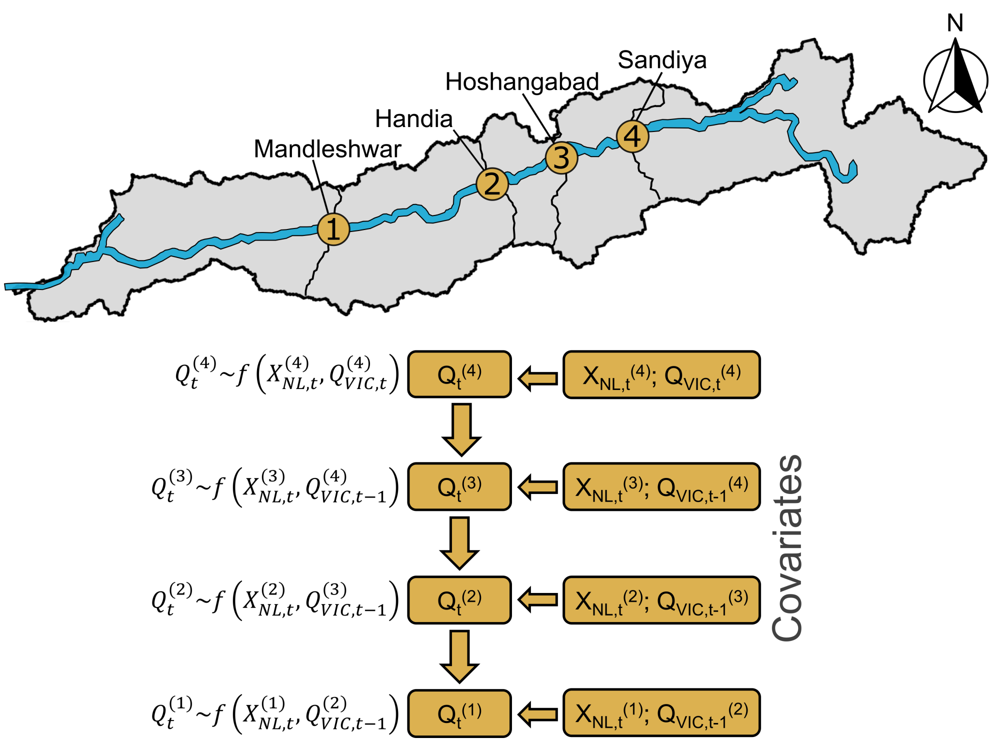


Figure 4. Conceptual sketch of the network Bayesian model for the Narmada basin. $Q_t^{(i)}$ corresponds to the observed streamflow at gauge *i* and day *t*, $X_{LN,t}^{(i)}$ to the product daily VIC Streamflow at gauge *i* and 2-day spatial average precipitation from the area between stations *i* and *i+1* at day *t*, and $Q_{VIC,t}^{(i)}$ to the daily VIC streamflow at gauge *i* at day *t*.

We select the best covariates for each gauge based on the lowest value of the Deviance Information Criterium (DIC). This gives 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:

$$\mu_t^{(1)} = \beta_0^{(1)} + \beta_1^{(1)} X_{NL,t}^{(1)} + \beta_2^{(1)} Q_{VIC,t-1}^{(2)}$$

$$\sigma_t^{(1)} = \phi_0^{(1)} + \phi_1^{(1)} X_{NL,t}^{(1)} + \phi_2^{(1)} Q_{VIC,t-1}^{(2)}$$

Handia:

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

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

Hoshangabad:

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

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

Sandiya:

$$\mu_t^{(4)} = \beta_0^{(4)} + \beta_1^{(4)} X_{NL,t}^{(4)} + \beta_2^{(4)} Q_{VIC,t}^{(4)}$$

$$\sigma_t^{(4)} = \phi_0^{(4)}$$

- 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 β(i) and ϕ(i) for each gauge station were considered Multivariate Normal distribution (MVN) to capture their dependence structure.

RESULTS

BHNM Calibration

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

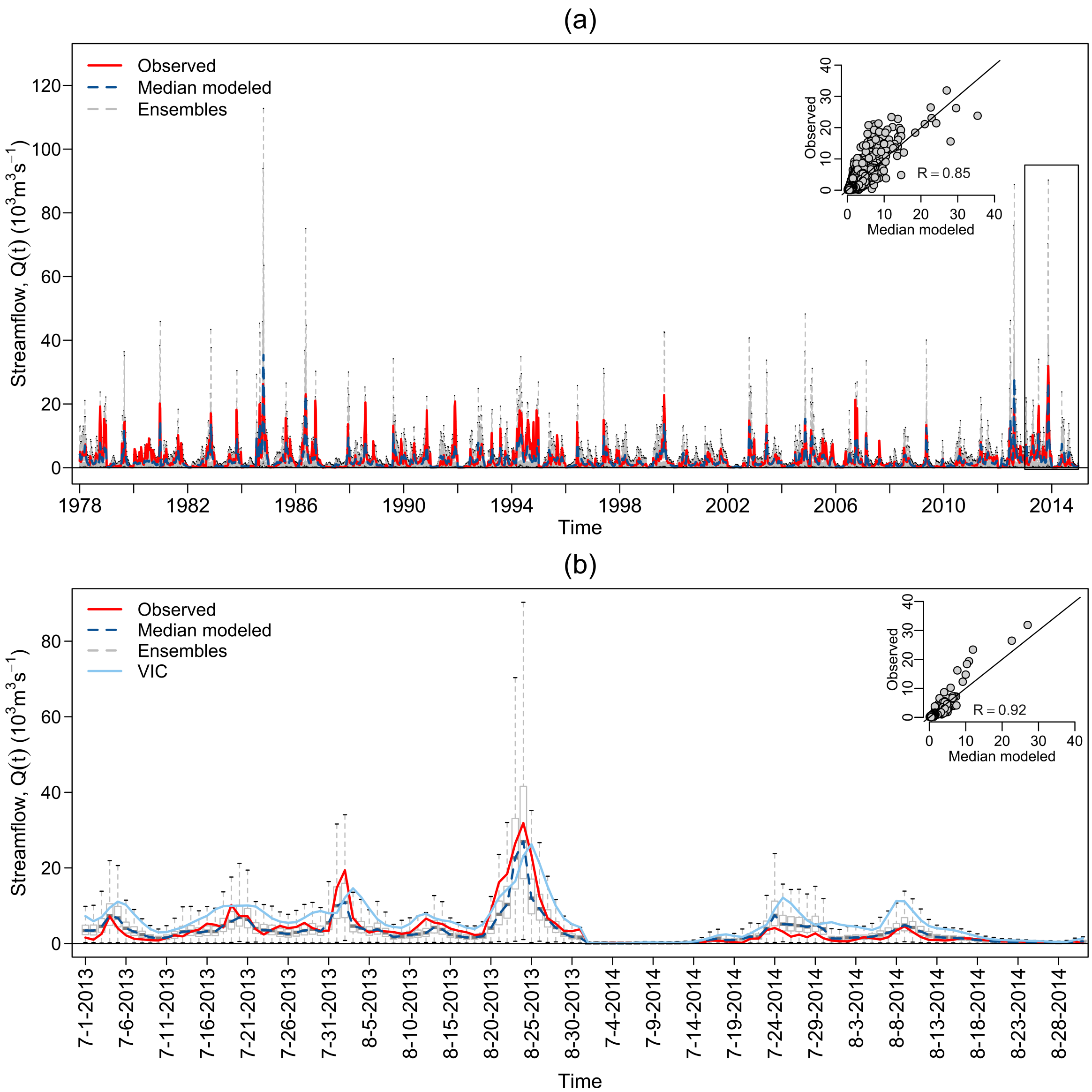


Figure 5. Ensembles of simulated July-August daily streamflow for the Handia gauge station presented as boxplot time series for (a) entire record (1978-2014) and (b) 2013-2014. The boxplots represent the posterior distribution estimates of the daily streamflow. Red lines correspond to the observed daily 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.

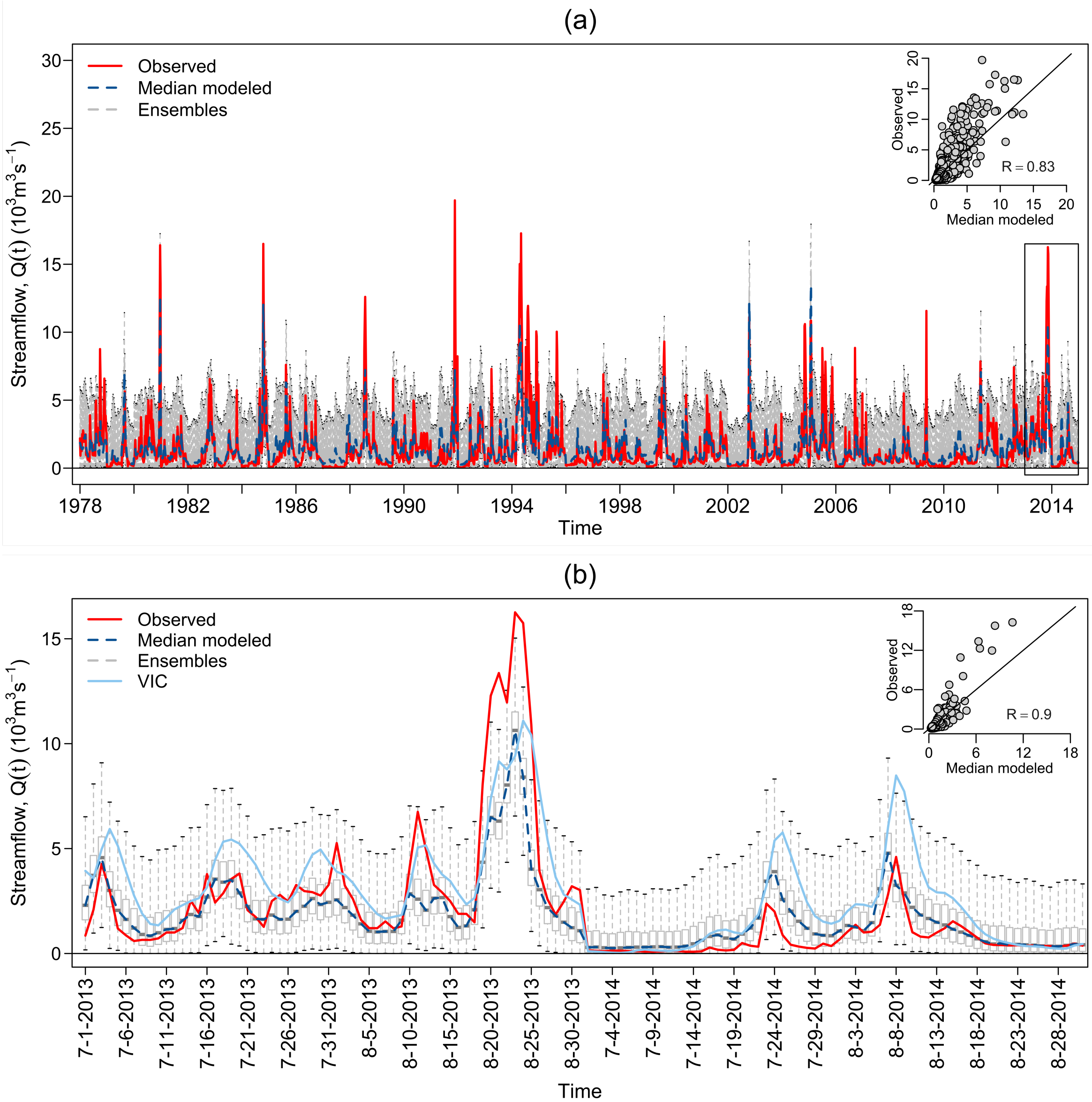


Figure 6. Similar to figure 5, but for the Sandiya gauge station. R values are significant (P-value<0.1).

- All the observed values are captured by the ensembles variability for Mandleshwar, Handia (Figure 4a), and Hoshangabad gauges
- The timing of the streamflow peaks is captured by the ensembles for all the gauges (see Figures 4b and 4b)
- Peaks of observed streamflow for Sandiya gauge station cannot be captured by the ensembles variability

BHNM Validation

The BHNM was fitted for the years 1978-2012, and forecasts were made for the years 2013-2014 (high streamflow values occur in 2013).

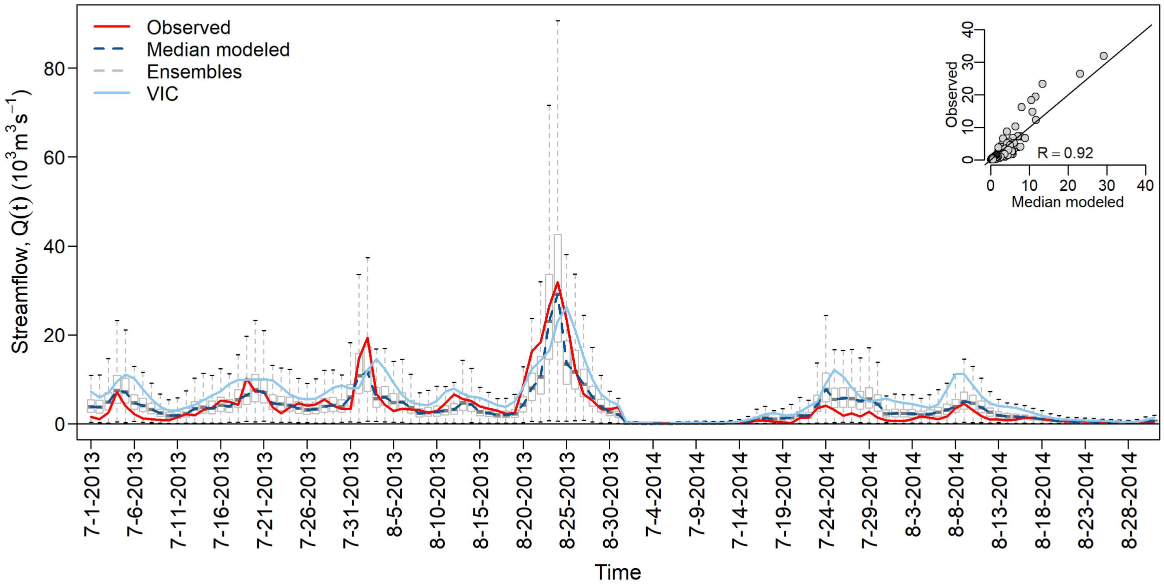


Figure 7. Ensembles forecast of July-August daily streamflow presented as boxplot time series for the years 2013-2014 at Handia gauge station. The boxplots represent the posterior distribution estimates of the daily streamflow. Red lines correspond to the observed daily streamflow and blue-dashed lines 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.

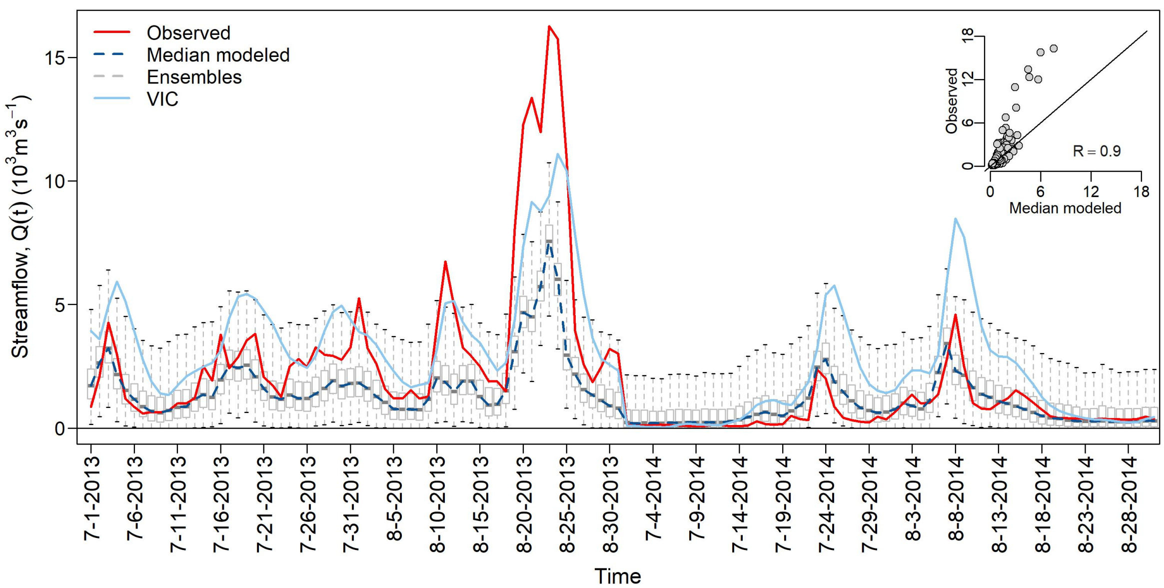


Figure 8. Similar to figure 7, but for the Sandiya gauge station. R values are significant (P-value<0.1).

- There is no reduction of the skill compared to the calibration for Handia
- The coupled model is able to reduce the negative bias and correct the timing delay for all the gauges
- For Sandiya there is a reduction of the skill for extreme flows compared to the calibration period.

CONCLUSIONS

- The incorporation of nonlinear covariates in the BHNM model enabled to correct the timing delay and nonsystematic bias present in the VIC model streamflow.
- There is an increase in the skill (R) compared to the VIC streamflow simulations.
- There is no significant reduction of the skill in a forecast mode compared to the calibration, except for Sandiya.
- The BHNM model is a potential post-processing tool for improving the streamflow simulations and could be tested in real-time.
- This model can be easily applied to other regions where physical-based models are implemented for improving the forecast performance.

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 second author.

ABSTRACT

Skillful forecasts of daily streamflow on river networks are crucial for flood mitigation, especially in rainfall-driven river basins. This is of acute importance on the Narmada River Basin in Central India, which is driven by the summer monsoon rainfall, and floods lead to heavy loss of life and infrastructure. Physical hydrologic models based on the land surface model – Variable Infiltration Capacity (VIC), have been developed in an experimental mode to model and forecast the hydrologic system, which includes – storage, soil moisture and runoff – by incorporating rainfall data from the India Meteorological Department (IMD). To enhance the forecast skill by model-combination, we propose a coupled physical-statistical modeling framework. In this, we couple the VIC model with a novel Bayesian Hierarchical Network Model (BHNM) for daily streamflow forecasts that uses the network topology to capture the spatial dependence. The daily streamflow at each station is modelled as Gamma distribution with time-varying parameters. The distribution parameters for each day are modeled as a linear function of covariates, which include antecedent streamflow from upstream gauges and, daily 2-day, or 3-day precipitation from the upstream contributing areas - that reflect the antecedent land conditions. The posterior distribution of the model parameters and, consequently, the predictive posterior distribution of the daily streamflow at each station and for each day are obtained. To the BHNM model, we will couple the VIC model by including the hydrologic forecasts – especially soil moisture and storage – as additional covariates. The coupled model will be demonstrated by its application to daily summer (July-August) streamflow at 4 gauges in the Narmada basin network for the period 1978 – 2014. The model skill will also be tested on one high flooding event on both the timing and magnitude. These model combinations will enable to combine the strengths of the individual models in capturing the hydrologic processes, biases and nonstationary relationships, to provide skillful daily streamflow forecasts. This will be of immense help in flood mitigation.

REFERENCES

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)

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)

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. Journal of Geophysical Research, 99(D7), 14415. https://doi.org/10.1029/94JD00483 (https://doi.org/10.1029/94JD00483)

Liang, X., Wood, E. F., & Lettenmaier, D. P. (1996). Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification. Global and Planetary Change, 13(1), 195–206.

Lohmann, D., Nolte-Holube, R., & Raschke, E. (1996). A large-scale horizontal routing model to be coupled to land surface parametrization schemes. Tellus A: Dynamic Meteorology and Oceanography, 48(5), 708–721. https://doi.org/10.3402/tellusa.v48i5.12200 (https://doi.org/10.3402/tellusa.v48i5.12200)

Pai, D. S., Sridhar, L., Rajeevan, M., Sreejith, O. P., Satbhai, N. S., & Mukhopadyay, 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), 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/10.1029/2019WR024901)

Srivastava, A. K., Rajeevan, M., & Kshirsagar, S. R. (2009). Development of a high resolution daily gridded temperature data set (1969-2005) for the Indian region. Atmospheric Science Letters, 10(4), 249–254. https://doi.org/10.1002/asl.232 (https://doi.org/10.1002/asl.232)

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