Bayesian Hierarchical Network Model for Forecasting Daily River Stage in a River Network

Naman Kishan Rastogi[1,2], Balaji Rajagopalan[1,2]
[1]Department of Civil, Environmental and Architectural Engineering, University of Colorado, Boulder, CO, USA
[2] Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA

Introduction

- **River stage data is crucial for the development of accurate flood forecasting models and early warning systems.**
- **A novel Bayesian Hierarchical Network Model (BHNM) is designed for ensemble predictions of daily river stage, leveraging the spatial interdependence of river networks and hydrometeorological variables from the upstream catchment area between gauge stations.**
- **The model allows parameters to dynamically vary over time, influenced by chosen covariates specific to each day.**
- **Utilizes the river network's structure to integrate flow and stage data from upstream gauges, along with precipitation data, efficiently capturing spatial correlation of stage variations.**

Hierarchical Structure Overview

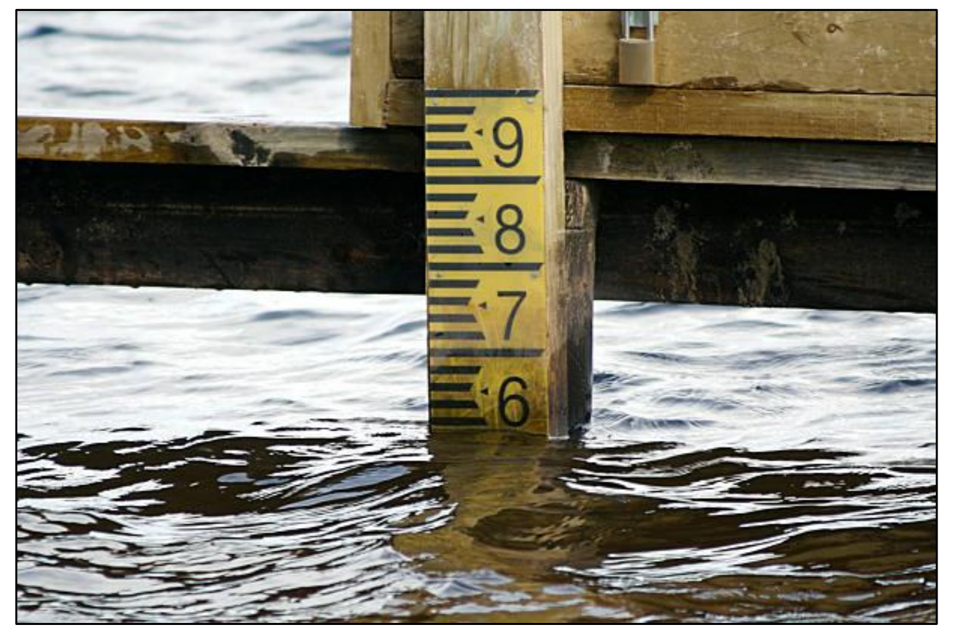


Figure 1. River Stage Measurement
Source: *https://www.istockphoto.com/photos/high-water-level*

- Priors β and $\emptyset$: Multivariate Normal distribution (MVN)
- Posterior distributions estimated via Markov Chain Monte Carlo (MCMC).
- Three chains run, each with a length of 8000.
- First 4000 samples discarded as warmup.
- 3000 samples retained for each parameter.
- Model selection based on the Deviance Information Criterion (DIC), favoring the model with the minimum DIC.
- Convergence assessed using the scale reduction factor $\hat{R}$, with values < 1.1 indicating good convergence.

Stage at each Gauge

Data Layer:

$$Gamma\left(L_t^{(i)} \middle| \alpha_t^{(i)}, \lambda_t^{(i)}, L_{t-k}^{(j\epsilon J(i))}, Q_{t-k}^{(j\epsilon J(i))}, P_{t-k}^{(i)} \right), \qquad i = 1,2,3$$

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

Process Layer: Parameters vary in space and time

$$\mu_t^{(i)} = \begin{cases} \beta_1^{(i)} + \beta_2^{(i)} Q_{t-k}^{(j\epsilon J(i))} + \beta_3^{(i)} L_{t-k}^{(j\epsilon J(i))} & (P_{t-k}^{(i)} = 0) \\ \beta_4^{(i)} + \beta_5^{(i)} Q_{t-k}^{(j\epsilon J(i))} + \beta_6^{(i)} L_{t-k}^{(j\epsilon J(i))} + \beta_7^{(i)} P_{t-k}^{(i)} & (P_{t-k}^{(i)} \neq 0) \end{cases}$$

$$\sigma_t^{(i)} = \begin{cases} \emptyset_1^{(i)} + \emptyset_2^{(i)} Q_{t-k}^{(j\epsilon J(i))} + \emptyset_3^{(i)} L_{t-k}^{(j\epsilon J(i))} & (P_{t-k}^{(i)} = 0) \\ \emptyset_4^{(i)} + \emptyset_5^{(i)} Q_{t-k}^{(j\epsilon J(i))} + \emptyset_6^{(i)} L_{t-k}^{(j\epsilon J(i))} + \emptyset_7^{(i)} P_{t-k}^{(i)} & (P_{t-k}^{(i)} \neq 0) \end{cases}$$

Likelihood:

$$f(\theta \,|\, data) \propto \prod_{t>k}^{T} \prod_{i=1}^{3} f\left(L_t^{(i)} \middle| \theta^{(i)}, L_{t-k}^{(j\epsilon J(i))}, Q_{t-k}^{(j\epsilon J(i))}, P_{t-k}^{(i)} \right) \cdot f\left(\theta^{(i)} \middle| L_{t-k}^{(j\epsilon J(i))}, Q_{t-k}^{(j\epsilon J(i))}, P_{t-k}^{(i)} \right)$$

Priors:

$$\Sigma_\beta^{(i)}\ Inv\ wishart(\nu, A\mathbf{I}); \quad \Sigma_\emptyset^{(i)}\ Inv\ wishart(\nu, B\mathbf{I})$$

$L_t^{(i)}$ is the stage at a downstream gauge, i, on day t is dependent on stage ($L_{t-k}^{(j\epsilon J(i))}$) and flow ($Q_{t-k}^{(j\epsilon J(i))}$) at the most immediate ($i+1$) or second most immediate ($i+2$) upstream feeder gauge at day $t-k$ with $k > 0$ (k represents the lead time of the forecast); 1-day accumulated spatial average precipitation from the area between the station gauges i and $i+1$, $P_{t-k}^{(i)}$

Study area

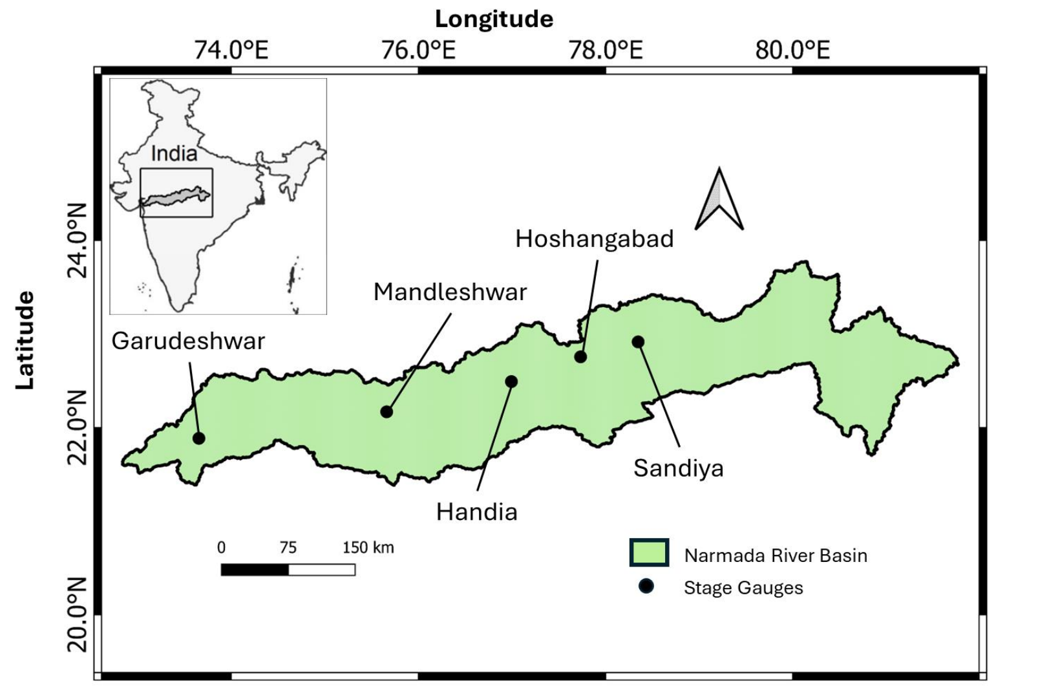


Figure 2. *Narmada basin boundary in India with the stage gauges.*

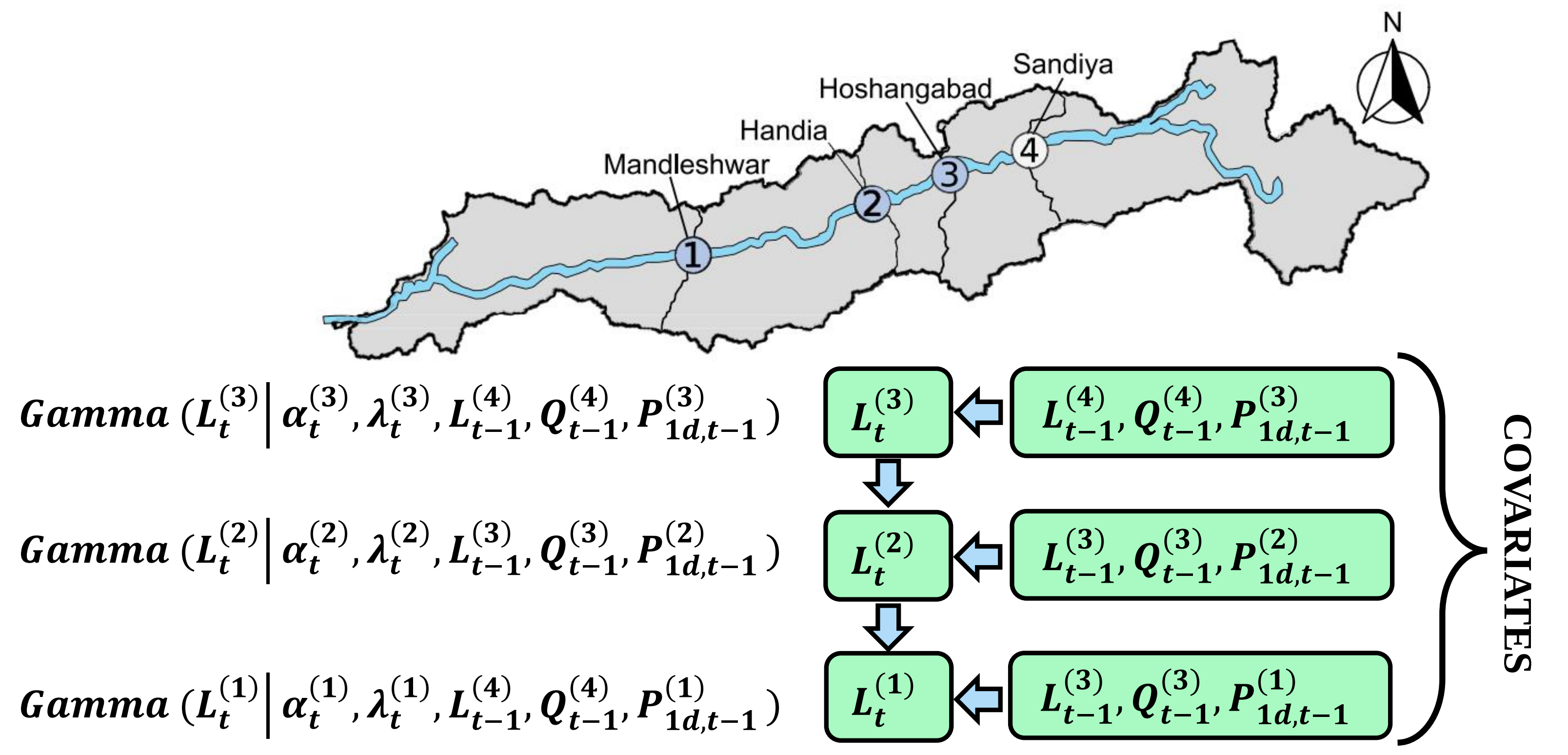
Gauge	Area (km²)	Elevation (m)	Mean seasonal stage (m)	Max. seasonal stage (m)
Mandleshwar	71,739	141	142	156
Handia	51,115	260	263	288
Hoshangabad	44,487	292	287	299
Sandiya	32,495	301	302	315

Table 1. *Dataset Pertaining to Stage Gauges in the Narmada River Basin Investigated in the Study*

Best Model Selection

The best-performing BHNM, identified as model 1, included the stage at the feeder site $L_{t-1}^{(j\epsilon J(i))}$, streamflow at the feeder site $Q_{t-1}^{(j\epsilon J(i))}$ and the 1-day accumulated spatial average precipitation $P_{1d,t-1}^{(i)}$ as covariates for gauge *i*.

BHNM Model: Best Model Selected



$$Gamma\,(L_t^{(3)} \,|\, \alpha_t^{(3)}, \lambda_t^{(3)}, L_{t-1}^{(4)}, Q_{t-1}^{(4)}, P_{1d,t-1}^{(3)})$$

$$Gamma\,(L_t^{(2)} \,|\, \alpha_t^{(2)}, \lambda_t^{(2)}, L_{t-1}^{(3)}, Q_{t-1}^{(3)}, P_{1d,t-1}^{(2)})$$

$$Gamma\,(L_t^{(1)} \,|\, \alpha_t^{(1)}, \lambda_t^{(1)}, L_{t-1}^{(4)}, Q_{t-1}^{(4)}, P_{1d,t-1}^{(1)})$$

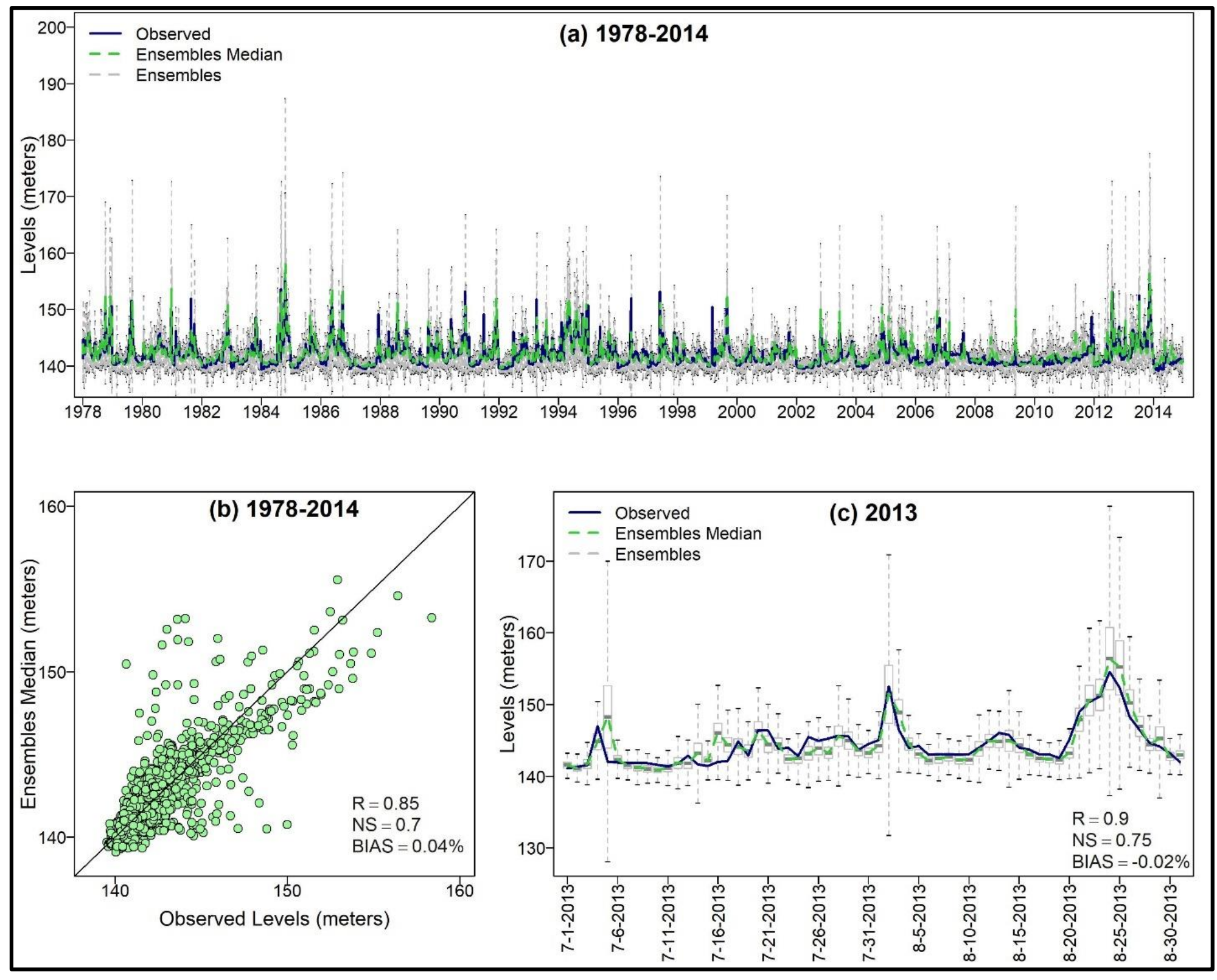
$L_t^{(3)}$ ← $L_{t-1}^{(4)}, Q_{t-1}^{(4)}, P_{1d,t-1}^{(3)}$

$L_t^{(2)}$ ← $L_{t-1}^{(3)}, Q_{t-1}^{(3)}, P_{1d,t-1}^{(2)}$

$L_t^{(1)}$ ← $L_{t-1}^{(3)}, Q_{t-1}^{(3)}, P_{1d,t-1}^{(1)}$

Figure 3. *BHNM Model structure with best set of Covariates*

Results



Gauge	1978-2014			2013		
	R	NSE	BIAS (%)	R	NSE	BIAS(%)
Mandleshwar	0.85	0.70	0.04	0.90	0.75	-0.02
Handia	0.90	0.78	0.03	0.93	0.71	0.08
Hoshangabad	0.93	0.86	0.00	0.94	0.87	-0.16

Table 2. *R, NSE and BIAS Values for the Entire Record (1978–2014) and the High-Flow Year 2013.*

Figure 4. (a) *Ensemble predictions for July–August daily stage,* ***(b)*** *a scatter plot Observed vs. mean stage, 1978–2014,* ***(c)*** *Predictive posterior ensembles, 2013 (high flow year), Mandleshwar gauge*

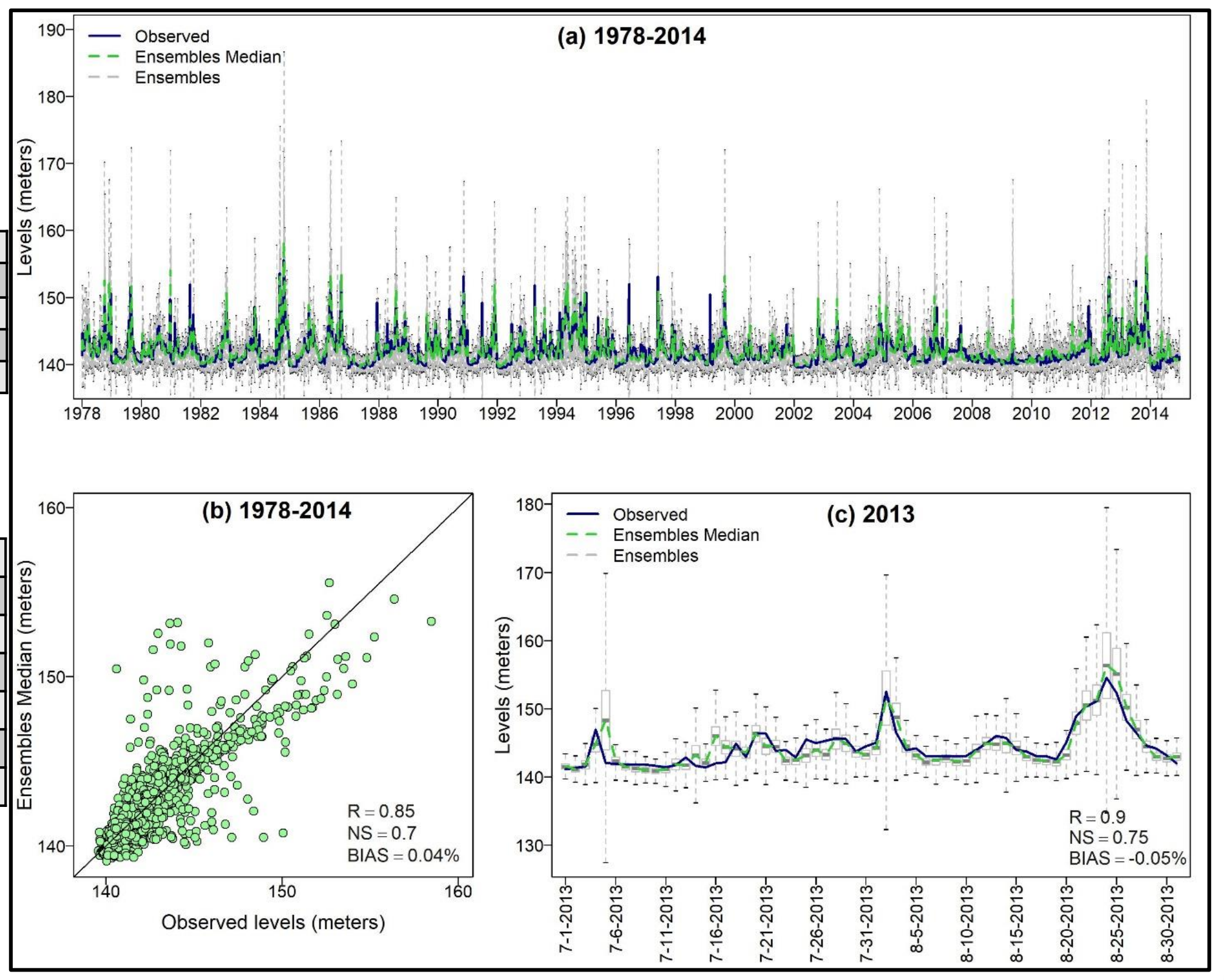


Figure 5. *As in Figure 4, predictive posterior ensemble forecast of July–August daily at Mandleshwar for the cross-validation.*

Gauge	1978-2014			2013		
	R	NSE	BIAS (%)	R	NSE	BIAS(%)
Mandleshwar	0.85	0.70	0.04	0.90	0.75	-0.05
Handia	0.90	0.78	0.03	0.93	0.7	0.08
Hoshangabad	0.93	0.86	0.00	0.94	0.86	-0.17

Table 3. *R, NSE and BIAS Values for the Entire Record (1978–2014) and the High-Flow Year 2013 for Cross-Validation.*

Model	Metric	Mandleshwar	Handia	Hoshangabad
BHNM	R	0.85	0.90	0.93
	NSE	0.70	0.78	0.86
	BIAS (%)	0.04	0.03	0.00
GLM	R	-0.16	-0.07	0.08
	NSE	-0.01	-0.01	-0.02
	BIAS (%)	-0.41	-0.018	-0.016

Table 4. *Statistical metrics for BHNM and GLM for the period 1978-2014.*

Skill Scores

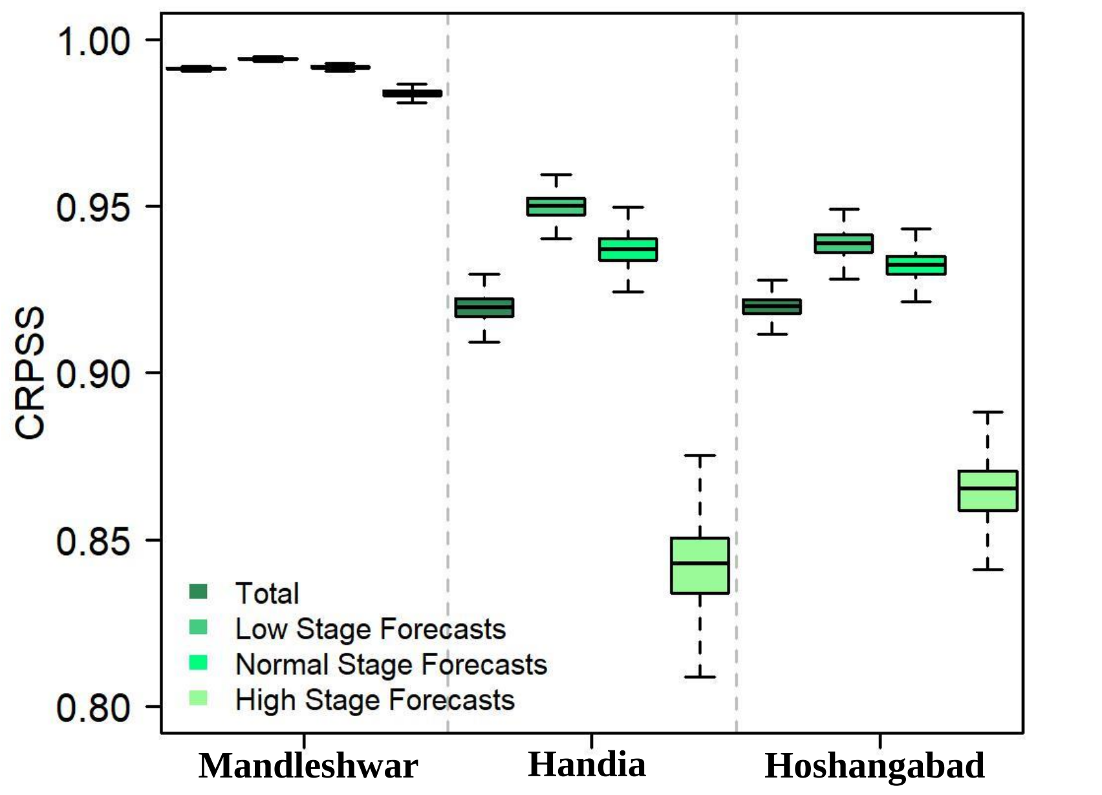


Figure 6. *Boxplots of continuous ranked probability skill score relative to GLM.*

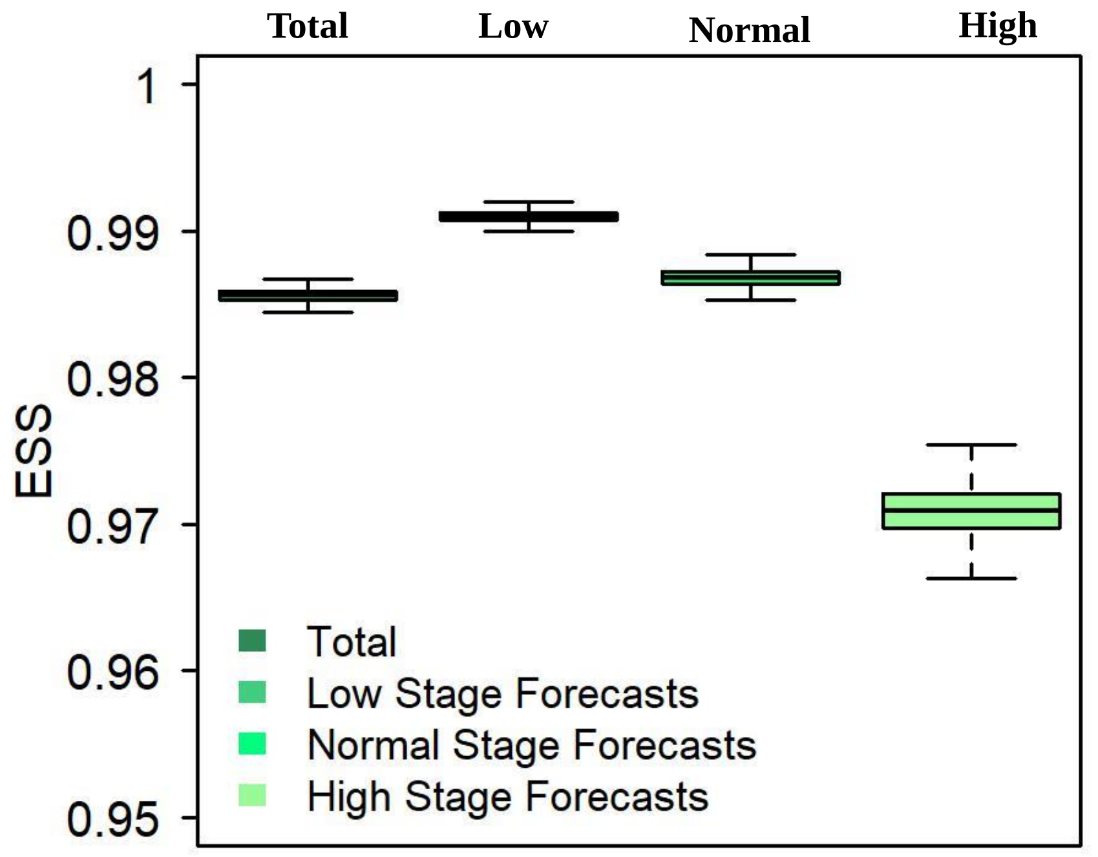


Figure 7. *Boxplots of the energy skill score (ESS) of daily stage ensembles relative to GLM.*

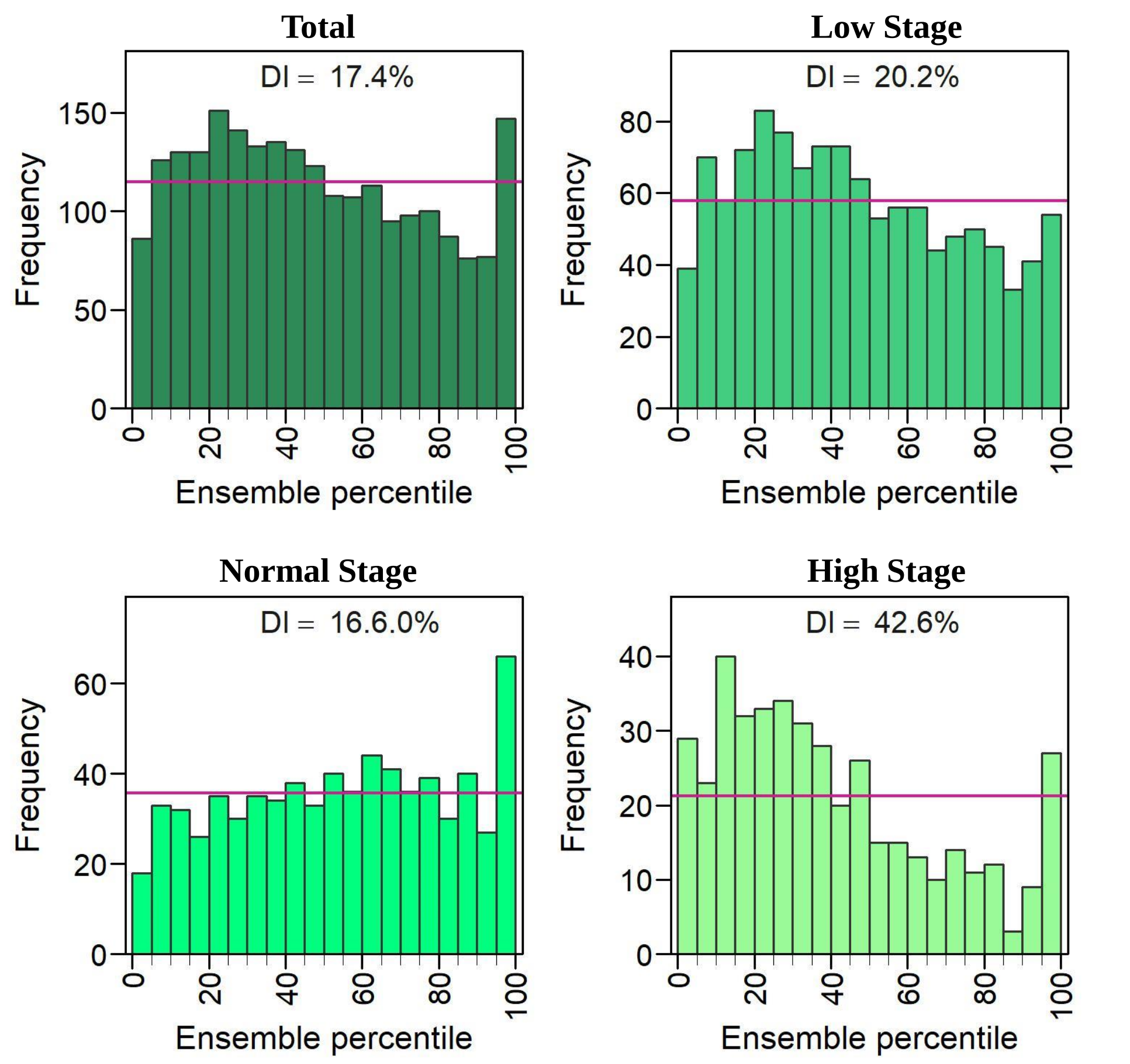


Figure 8. *Rank histograms of July-August daily basin average stage forecasts from Bayesian Hierarchical Network Model ensembles, segmented by stage forecasts relative to GLM.*

Stage Strata	Quantiles
Low stage forecasts	$(\bar{\hat{L}}(t) < L_{50th})$
Normal stage forecasts	$(L_{50th} < \bar{\hat{L}}(t) < L_{80th})$
High stage forecasts	$(\bar{\hat{L}}(t) > L_{80th})$

Table 5. *Quantiles for low, normal, and high flow strata.*

Conclusions

- The selected model incorporated daily stage and streamflow data from upstream feeder gauges and the 1-day accumulated spatial average precipitation between consecutive gauges from the preceding day as covariates.
- The ensemble forecasts from the BHNM exhibited superior skill compared to the GLM both individually at each site and jointly across all sites.
- Model validation indicates high skill relative to a generalized linear regression null-model.
- ESS shows a better joint performance (ESS values higher than CRPSS).

Future Extensions

- Enhanced accuracy by integrating BHNM model forecasts with hydrologic models and meteorological data for real-time river stage forecasting.
- Improved reliability through ensemble forecasting, incorporating multiple model outputs to address uncertainties, particularly effective in complex hydrological systems.

References

❖ **Ossandón, Á., Rajagopalan, B., Lall, U., Nanditha, J. S., & Mishra, V. (2021). A Bayesian hierarchical network model for daily streamflow ensemble forecasting. Water Resources Research, 57, e2021WR029920. https://doi. org/10.1029/2021WR029920**

Contact: Naman Kishan Rastogi
1st Year PhD
Dept. of Civil, Environmental and Architectural Engineering,
University of Colorado, Boulder
✉ naman.rastogi@colorado.edu, namanrastogiiisc@gmail.com