Estimating Bayesian Model Averaging Weights and Variances of Ensemble Flood Modeling Using Multiple Markov Chains Monte Carlo

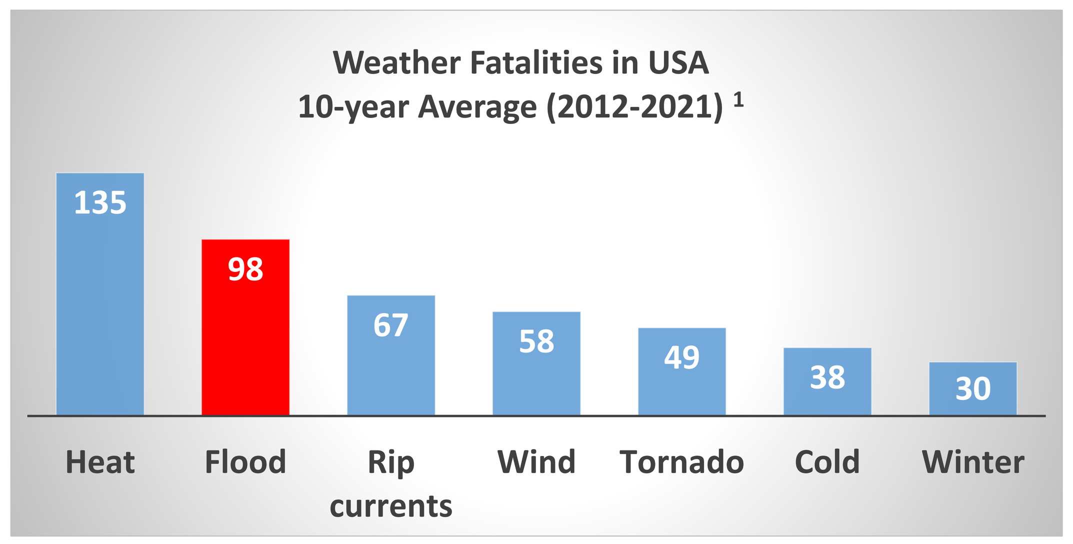
Tao Huang and Venkatesh Merwade

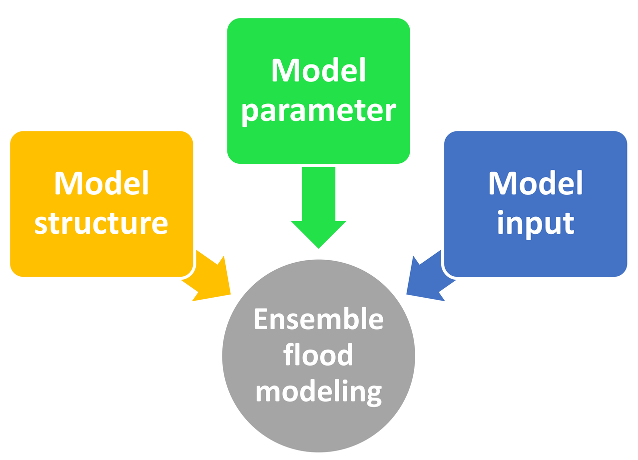
Lyles School of Civil Engineering, Purdue University, USA

2022 AGU Fall Meeting

NG35B-0458

Introduction





Multi-model method

- physics-based and data-driven models
- No "perfect" single model
- "Equifinality" in hydrologic & hydraulic modeling
- Fixed estimates cannot fit all kinds of scenarios

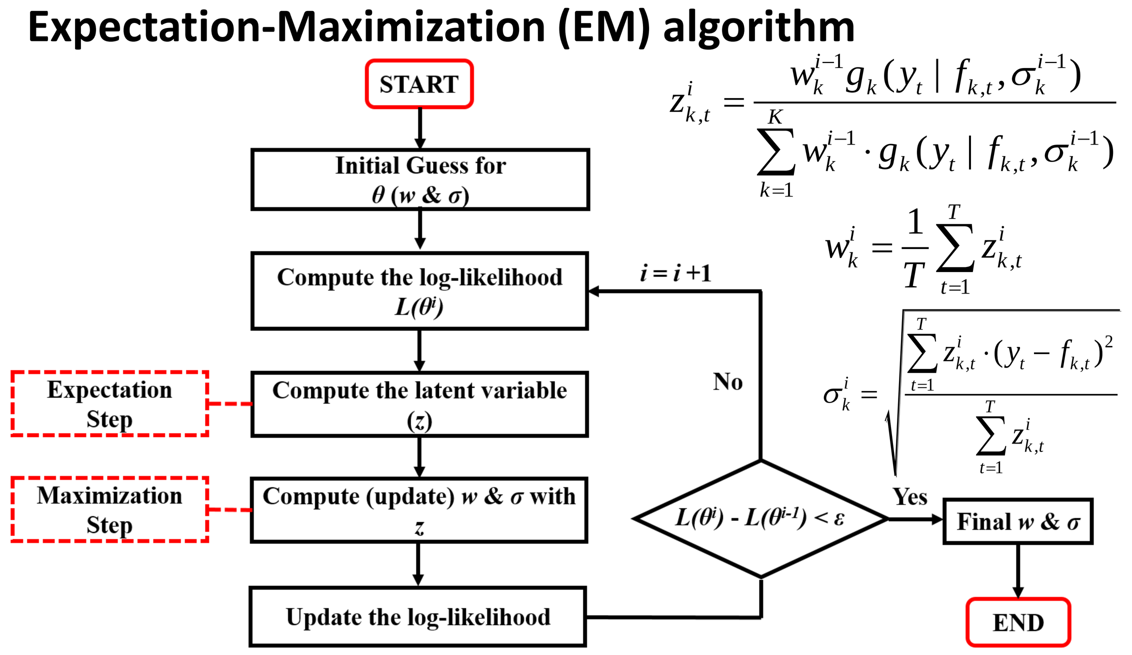
Bayesian Model Averaging (BMA)[2]

- **Law of total probability**

$$p(y \mid D) = \sum_{k=1}^{K} p(f_k \mid D) \cdot p_k(y \mid f_k, D) = \sum_{k=1}^{K} w_k \cdot p_k(y \mid f_k, D) \qquad \sum_{k=1}^{K} w_k = 1$$

$$L(\theta) = \sum_{t=1}^{T} \log\left(\sum_{k=1}^{K} w_k \cdot p_k(y_t \mid f_{k,t}, \sigma_k)\right) = \sum_{t=1}^{T} \log\left(\sum_{k=1}^{K} w_k \cdot \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y_t - f_{k,t}}{\sigma_k}\right)}\right)$$

- **Expectation-Maximization (EM) algorithm**



$$z_{k,t}^{i} = \frac{w_k^{i-1} g_k(y_t \mid f_{k,t}, \sigma_k^{i-1})}{\sum_{k=1}^{K} w_k^{i-1} \cdot g_k(y_t \mid f_{k,t}, \sigma_k^{i-1})}$$

$$w_k^{i} = \frac{1}{T} \sum_{t=1}^{T} z_{k,t}^{i}$$

$$\sigma_k^{i} = \sqrt{\frac{\sum_{t=1}^{T} z_{k,t}^{i} \cdot (y_t - f_{k,t})^2}{\sum_{t=1}^{T} z_{k,t}^{i}}}$$

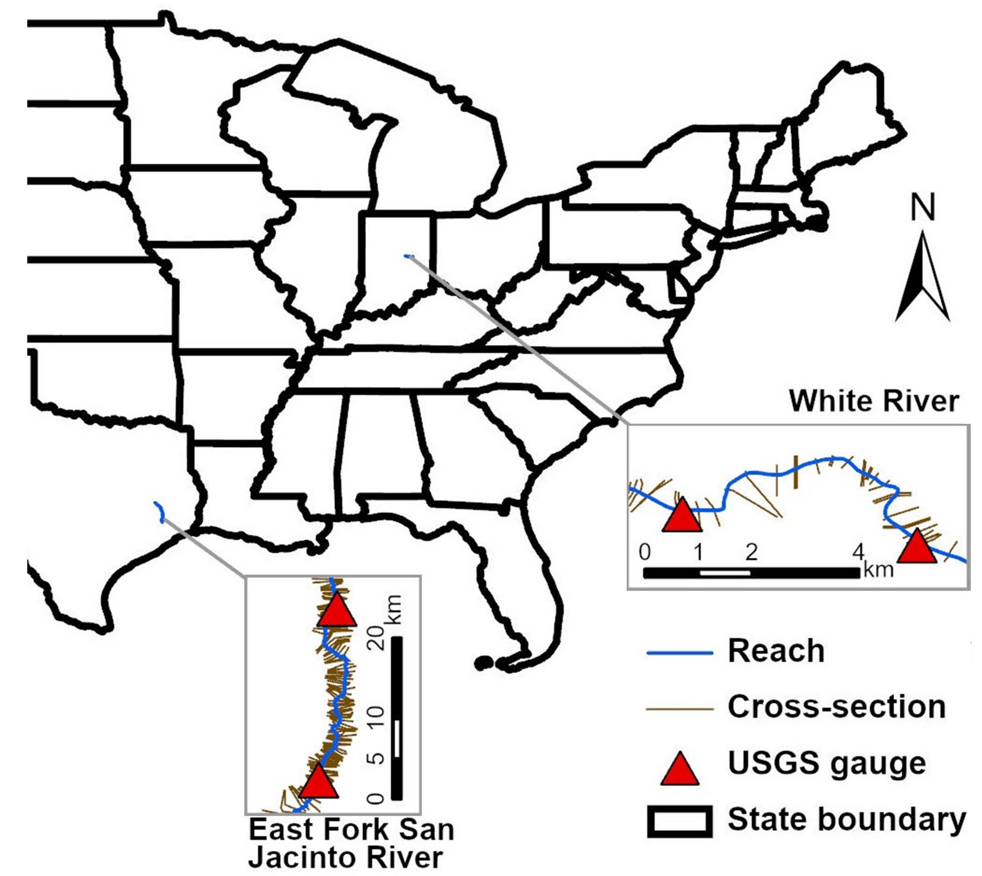
✓	✗
• combine estimations from multiple competing models • BMA weights are interpretable • provide a prediction distribution of variables of interest	• EM algorithm yielded fixed BMA weights and variances • Conditional PDF is limited by Gaussian assumption • Markov Chain Monte Carlo (MCMC) is rarely examined in BMA analysis

Research Objectives

Quantify the uncertainty in BMA parameters (weight & variance):

- estimate the BMA parameters using different numbers of samples in each Markov chain with the Metropolis-Hastings (M-H) algorithm
- compare the performance of EM and M-H MCMC algorithms for estimating the BMA parameters
- estimate the BMA weights using different proposal distributions in the M-H MCMC algorithm
- investigate the impact of different conditional PDFs of the predictor variable on the BMA parameters

Study Area and Data



Study stream	Channel length (km)	Average channel width (m)	Channel slope (%)
White	6.76	64	0.0631
East Fork San Jacinto	50.11	76	0.0438

Study stream	Upstream USGS gauge	Downstream USGS gauge	Simulation Period (100 days)
White	03348000	03348130	2021-3-15 to 2021-6-22
East Fork San Jacinto	08070000	08070200	2021-4-15 to 2021-7-23

Methodology

- **Numerical experiment (ensemble of 10 members)**

D = 100 days of daily water stage data (in meters)	
$f_1 = D + \varepsilon$, where $\varepsilon \sim N(0, 0.06^2)$	$f_2 = D + \varepsilon$, where $\varepsilon \sim N(0, 0.06^2)$
$f_3 = D + \varepsilon$, where $\varepsilon \sim N(0, 0.12^2)$	$f_4 = D + \varepsilon$, where $\varepsilon \sim N(0, 0.12^2)$
$f_5 = D + \varepsilon$, where $\varepsilon \sim N(0, 0.18^2)$	$f_6 = D + \varepsilon$, where $\varepsilon \sim N(0, 0.18^2)$
$f_7 = D + \varepsilon$, where $\varepsilon \sim N(0, 0.24^2)$	$f_8 = D + \varepsilon$, where $\varepsilon \sim N(0, 0.24^2)$
$f_9 = D + \varepsilon$, where $\varepsilon \sim N(0, 0.30^2)$	$f_9 = D + \varepsilon$, where $\varepsilon \sim N(0, 0.30^2)$

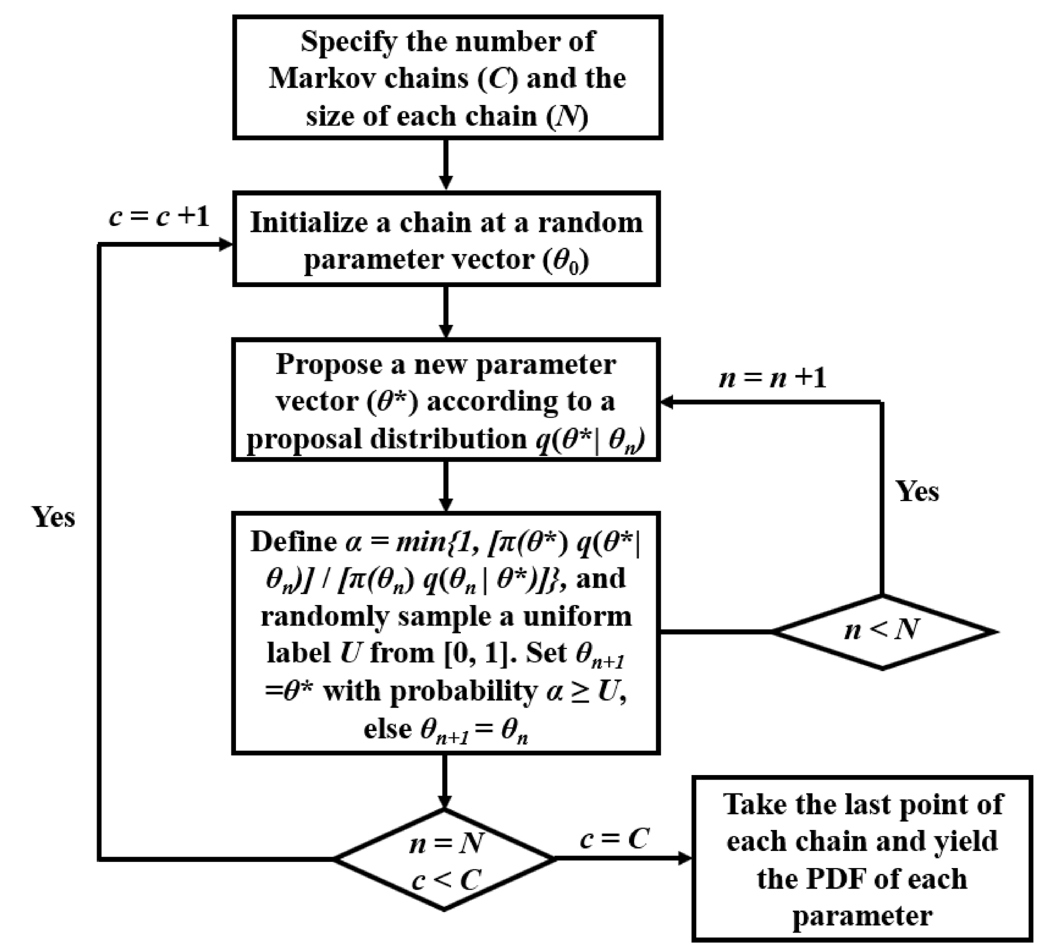
- **Ensemble flood modeling in 1D HEC-RAS**

No.	Channel Roughness	Upstream Flow Input	HEC-RAS Plan Files
1	0.8n	0.8Q	g01 & u01
2	0.8n	Q	g01 & u02
3	0.8n	1.2Q	g01 & u03
4	n	0.8Q	g02 & u01
5	n	Q	g02 & u02
6	n	1.2Q	g02 & u03
7	1.2n	0.8Q	g03 & u01
8	1.2n	Q	g03 & u02
9	1.2n	1.2Q	g03 & u03
10	Average of simulations from No.1-No.9		

Note: n is the Manning's n value for the main channel in the original HEC-RAS models, Q is the streamflow from USGS gauge stations, g** represents a geometry file of a HEC-RAS project, and u** represents a flow data file of a HEC-RAS project.

Methodology (cont.)

- **BMA analysis & Metropolis-Hastings (M-H) algorithm**



$$L(\theta) = \sum_{t=1}^{T} \log\left(\sum_{k=1}^{K} w_k \cdot p_k(y_t \mid f_{k,t}, \sigma_k)\right) = \sum_{t=1}^{T} \log\left(\sum_{k=1}^{K} w_k \cdot \frac{1}{\Gamma(\alpha)\beta^{\alpha}} y_t^{\alpha-1} e^{-\frac{y_t}{\beta}}\right)$$

$$\alpha = f_{k,t}^{\ 2} / \sigma_k^{\ 2} \qquad \beta = \sigma_k^{\ 2} / f_{k,t}$$

- **Evaluation metrics for model performance[3]**

$$UC1 = \frac{N_{obs-90\%}}{n} \cdot 100\%$$

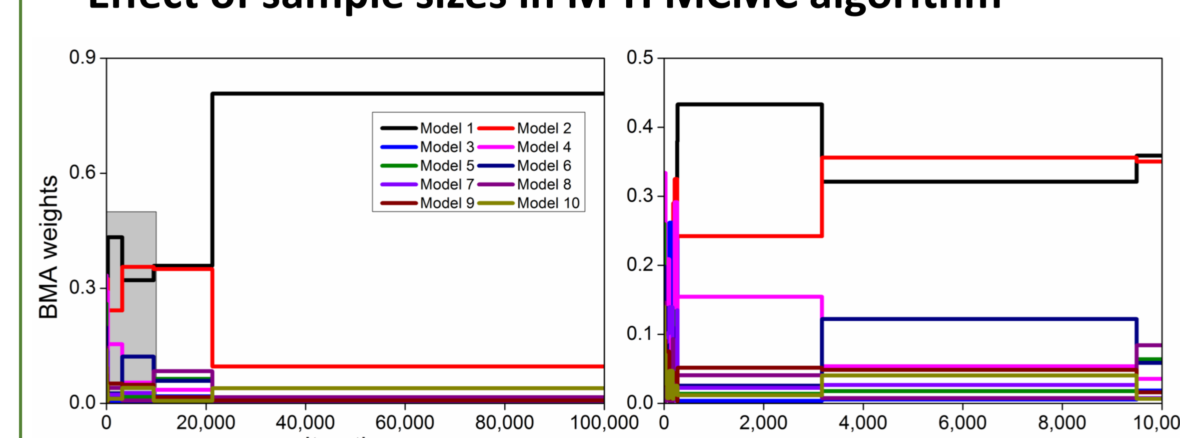
$$UC2 = 1 - NSE = \frac{RMSE^2}{\sigma_{obs}^{\ 2}} \cdot 100\% \qquad RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_{obs,i} - y_{BMA,i})^2}{n}}$$

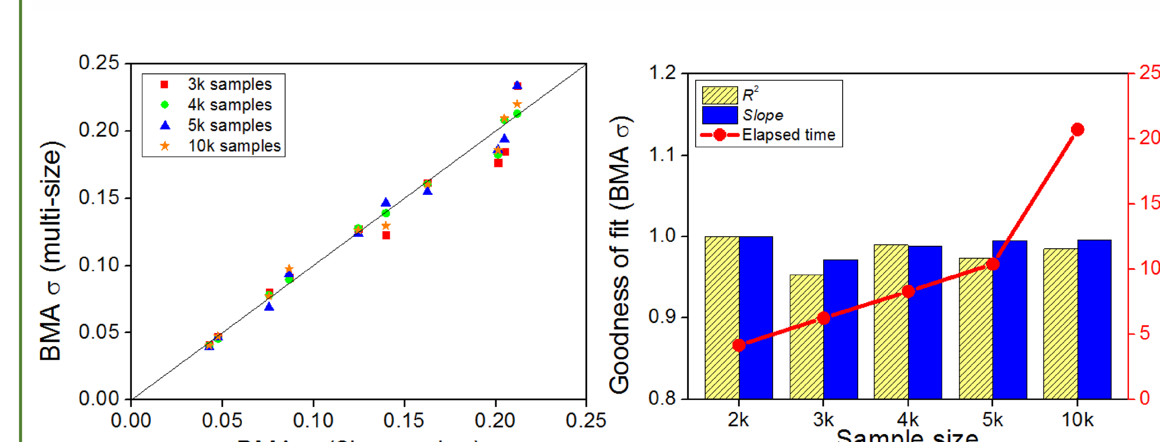
$$UC3 = 1 - KGE = \sqrt{(r-1)^2 + (\frac{\sigma_{BMA}}{\sigma_{obs}} - 1)^2 + (\frac{\bar{y}_{BMA}}{\bar{y}_{obs}} - 1)^2} \cdot 100\%$$

$$UC4 = (1 - R^2 + |1 - Slope|) \cdot 100\%$$

Results and Discussion

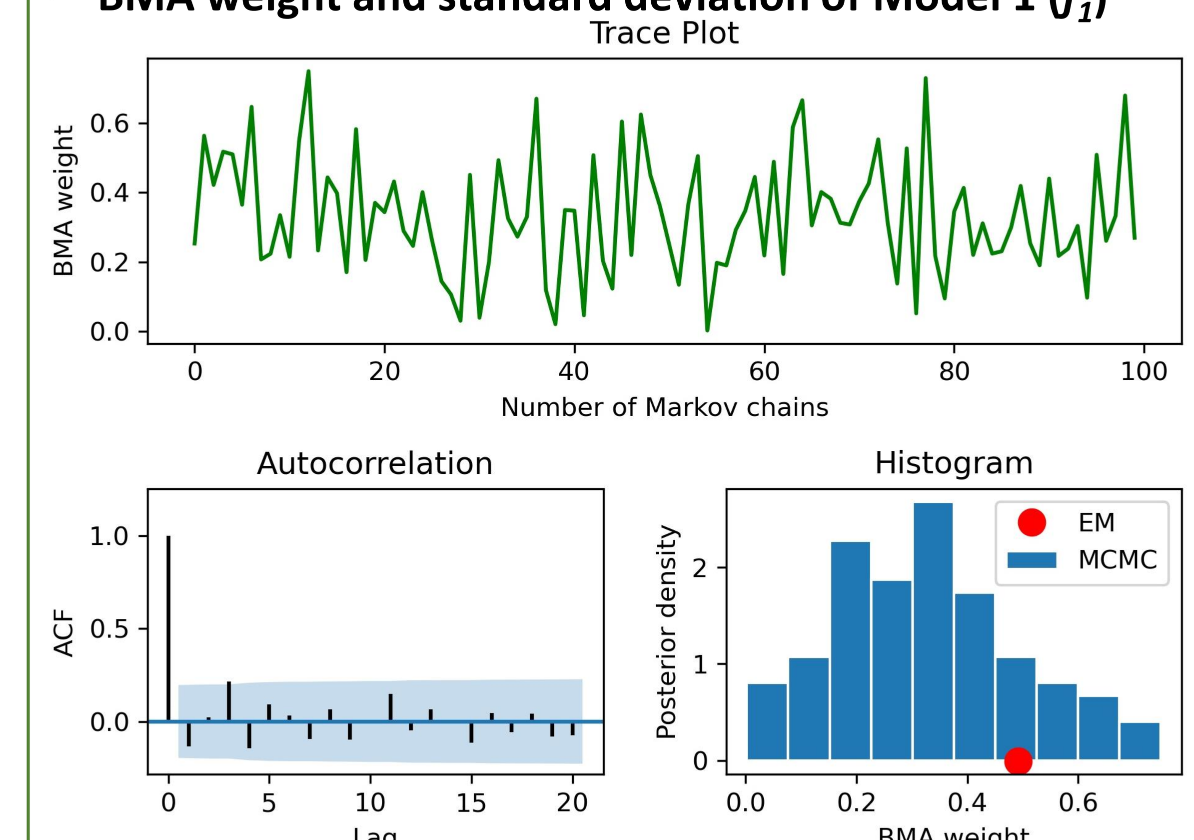
- **Effect of sample sizes in M-H MCMC algorithm**





Results and Discussion (cont.)

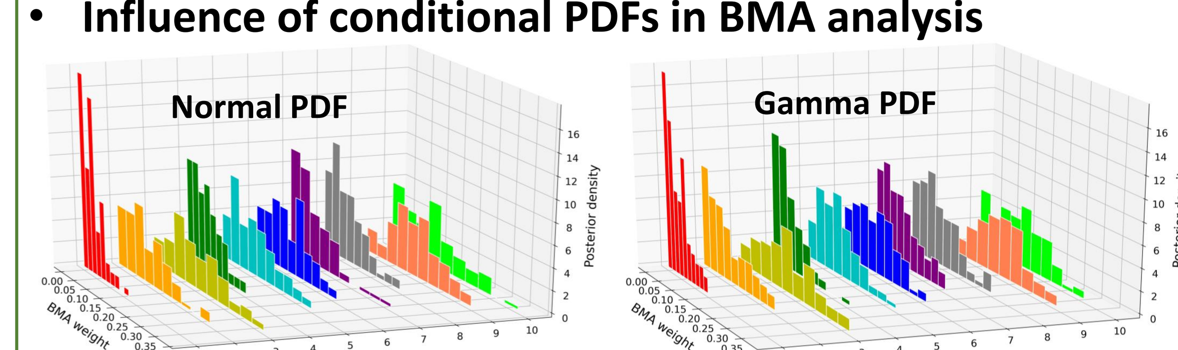
- **BMA weight and standard deviation of Model 1 (f_1)**



No.	EM weight	EM σ (m)	MCMC weight (Uniform & Normal)	MCMC σ (m)	Given σ (m)
1	0.492	0.04	0.332 & 0.354	0.05	0.06
2	0.497	0.04	0.320 & 0.35	0.04	0.06
3	0.0002	0.02	0.068 & 0.062	0.08	0.12
4	0	0.05	0.059 & 0.053	0.09	0.12
5	0	0.08	0.040 & 0.034	0.14	0.18
6	0	0.02	0.044 & 0.033	0.12	0.18
7	0	0.11	0.035 & 0.029	0.16	0.24
8	0	0.05	0.032 & 0.028	0.20	0.24
9	0.01	0.001	0.037 & 0.028	0.20	0.30
10	0	0.09	0.033 & 0.028	0.21	0.30

Algorithm	RMSE (m)	Average 90% prediction interval (m)	UC1 (%)	UC2 (%)	UC3 (%)	UC4 (%)
EM	0.041	0.19	5.00	1.61	1.06	2.00
MCMC (Uniform)	0.037	0.47	0.00	1.33	1.04	1.66
MCMC (Normal)	0.037	0.41	0.00	1.33	0.96	1.64

- **Influence of conditional PDFs in BMA analysis**



Algorithm	RMSE (m)	Average 90% prediction interval (m)	UC1 (%)	UC2 (%)	UC3 (%)	UC4 (%)
Normal PDF	0.421	1.32	13.00	7.54	5.69	7.76
Gamma PDF	0.419	1.34	11.00	7.48	5.67	7.71

Conclusions

- MCMC method with multiple independent chains is applicable in the BMA analysis and can demonstrate a full view of the uncertainty of BMA weights and variances.

References

1. Data source: https://www.weather.gov/hazstat/
2. Raftery, A. E., Gneiting, T., Balabdaoui, F., Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. Monthly weather review, 133(5), 1155-1174.
3. Huang, T., & Merwade, V. (2023). Uncertainty Analysis and Quantification in Flood Insurance Rate Maps Using Bayesian Model Averaging and Hierarchical BMA. Journal of Hydrologic Engineering, 28(2), 04022038.