

RECLAMATION

Managing Water in the West

Developing tool to assess model uncertainty in sediment simulation

**Research and Development Office
Science and Technology Program
Final Report ST-2017-8680-1**



**U.S. Department of the Interior
Bureau of Reclamation
Research and Development Office**

February 27, 2018

Mission Statements

Protecting America's Great Outdoors and Powering Our Future

The Department of the Interior protects and manages the Nation's natural resources and cultural heritage; provides scientific and other information about those resources; and honors its trust responsibilities or special commitments to American Indians, Alaska Natives, and affiliated island communities.

Disclaimer:

This document has been reviewed under the Research and Development Office Discretionary peer review process https://www.usbr.gov/research/peer_review.pdf consistent with Reclamation's Peer Review Policy CMP P14. It does not represent and should not be construed to represent Reclamation's determination, concurrence, or policy.

REPORT DOCUMENTATION PAGE			<i>Form Approved</i> <i>OMB No. 0704-0188</i>		
T1. REPORT DATE: 02/2018		T2. REPORT TYPE: RESEARCH		T3. DATES COVERED 10-01-2017 to 2-28-2018	
T4. TITLE AND SUBTITLE Developing tool to assess model uncertainty in sediment simulation			5a. CONTRACT NUMBER		
			5b. GRANT NUMBER R12AC80251		
			5c. PROGRAM ELEMENT NUMBER 8680 (S&T)		
6. AUTHOR(S) Youngjai Jung, Jeff Niemann, and Blair Greimann			5d. PROJECT NUMBER ST-2017-8680-01		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER 86-68240		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Colorado State University, Department of Civil Engineering Campus Delivery 1372, Fort Collins, CO 80523-1372			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Research and Development Office U.S. Department of the Interior, Bureau of Reclamation, PO Box 25007, Denver CO 80225-0007			10. SPONSOR/MONITOR'S ACRONYM(S): R&D: Research and Development Office, BOR/USBR: Bureau of Reclamation, DOI: Department of the Interior		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) ST-2017-8680-01		
12. DISTRIBUTION / AVAILABILITY STATEMENT Final report can be downloaded from Reclamation's website: https://www.usbr.gov/research/					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT (Maximum 200 words) The uncertainty in numerical model predictions can result from simplifications in the model's representations of the physical systems (model structure uncertainty), errors in the values assigned to model parameters (parameter uncertainty), and errors in the model inputs (forcing or input uncertainty). The long-term objective of this research is to develop a formal and efficient framework to evaluate uncertainty in predictions from hydrologic, hydraulic, and sediment-transport models. This project aims to assess the uncertainty associated with parameter, forcing, and model structure using Bayesian uncertainty methods, and reduce the computational cost of the Bayesian method substantially while still providing reliable uncertainty estimates. The new approach for uncertainty reduces the required number of simulations to be applied to complex model applications, and retains enough formality to reliably evaluate data collection and model calibration strategies. To constrain the scope, this research focuses on applying the framework to a sediment transport model called Sedimentation and River Hydraulics – One Dimension (SRH-1D).					
15. SUBJECT TERMS Uncertainty Analysis, Sediment Transport Modeling, Bayesian					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT U	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER 303-445-2673

BUREAU OF RECLAMATION

Research and Development Office
Science and Technology Program

Sedimentation and River Hydraulics, Technical Service Center,
86-68240

Final Report ST-2017-8680-01

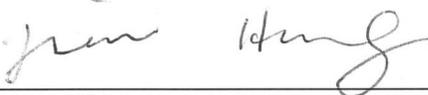
Developing tool to assess model uncertainty in sediment simulation



2-28-2018

Prepared by: Blair Greimann, P.E., Ph.D.

Hydraulic Engineer, Sedimentation and River Hydraulics Group, Technical Service Center, 86-68240



2/28/2018

Peer Review: Jianchun Victor Huang, P.E., Ph.D.

Hydraulic Engineer, Sedimentation and River Hydraulics Group, Technical Service Center, 86-68240

For Reclamation disseminated reports, a disclaimer is required for final reports and other research products, this language can be found in the peer review policy: *This document has been reviewed under the Research and Development Office Discretionary peer review process https://www.usbr.gov/research/peer_review.pdf consistent with Reclamation's Peer Review Policy CMP P14. It does not represent and should not be construed to represent Reclamation's determination, concurrence, or policy.*

Executive Summary

Colorado State University has been working with Reclamation to develop a framework to evaluate the uncertainty associated with numerical models developed at the U.S. Bureau of Reclamation (Reclamation). A variety of hydrologic and hydraulic numerical models has been developed and widely used by the Reclamation to predict impacts of potential river restoration activities. Recently, assessing the uncertainty in predictions from such models has been underscored in the field of hydraulic and sediment transport modeling. Those uncertainties can result from the simplifications and assumptions used in the model's mathematical structure, errors in the model parameter values, and errors in the data used for the model inputs. Several methods using a Bayesian uncertainty framework have been suggested to quantify the uncertainty in the predictions specifically from the Sedimentation and River Hydraulics – One Dimension (SRH-1D) model. GSA-GLUE (Global sensitivity analysis – Generalized likelihood uncertainty estimation) method was developed to assess uncertainty of individual model parameters. The combination of the MSU (Multivariate shuffled complex evolution Metropolis) and BMA (Bayesian model averaging) methods was suggested to evaluate the uncertainty of model parameters and the uncertainty associated with the selection of a transport equation in SRH-1D simulations.

The purpose of this research project was to develop and test a formal and efficient framework to assess the uncertainty in the predictions from hydrologic, hydraulic, and sediment transport models. Through this research, we aimed to develop improved methodologies that: (1) require few model simulations and (2) retain enough formality so that data collection and model calibration strategies can be simplified to reduce the uncertainty in the model predictions. To constrain the immediate scope, the new methods focused on quantifying how the uncertainties originating from model parameter values, input data, and the model's mathematical structure affect the predictions from the SRH-1D model, but those methods are transferrable to other types of models. The chapters in the final report address three specific objectives that have been accomplished to achieve the primary purpose of this research. Those chapters are summarized below:

1. Simple error models are developed for the input data of a sediment transport model and integrated into an existing Bayesian method in order to determine whether uncertain inputs contribute substantially to the overall uncertainty in the predictions. Input errors are modeled using Gaussian distributions separately for various input data such as discharges, sediment rating curves, and cross section elevations. The means and standard deviations of those distributions are treated as uncertain parameters, and they are estimated within the Bayesian framework for parameter uncertainty. This approach enables a modeler to identify the contribution of each uncertain input to the overall uncertainty, which can suggest strategies to reduce the uncertainty and improve reliability in the model predictions.
2. A new algorithm is developed to improve the efficiency of the uncertainty estimation process for sediment transport model parameters. In order to reduce the computational cost, the new method is designed to use repeated parameter sets in the sample when specifying the probability distributions of parameters instead of generating new but similar parameter sets that require new model simulations, which is the typical approach of existing Markov chain Monte Carlo methods. This new approach can save large numbers of model simulations when evaluating the uncertainty in model predictions due to uncertainty in the parameter values.

3. A multivariate version of the BMA method is developed to assess the uncertainty associated with the selection and application of a transport equation in sediment transport models. The existing BMA method is modified to enable consideration of multiple model output variables and allow the uncertainty associated with each equation to vary with the magnitude of the variables if needed. This methodology can reduce the effects of imperfections in a single model prediction and provide a forecast along with its credible interval to characterize the uncertainty through combining the predictions from a set of competing transport equations.

The Final Report for this project is submitted in the form of a project report from Colorado State University and is found in Appendix A.

Contents

Executive Summary iii
Appendix A – Final Report From Colorado State UniversityA-1

Appendix A – Final Report From Colorado State University

Final Report for USBR Grant

Agreement Number: R12AC80251

**DESIGN AND TESTING OF COMPUTATIONALLY-EFFICIENT METHODS TO
EVALUATE PARAMETER AND MODEL UNCERTAINTIES**

Submitted by

Jeffrey Y. Jung and Jeffrey D. Niemann

Department of Civil and Environmental Engineering

Colorado State University

Fort Collins, Colorado

Fall 2017

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the U.S. Bureau of Reclamation Science and Technology Program under Project 1596 (Agreement Number: R12AC80251) for their financial support.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	ii
LIST OF TABLES	v
LIST OF FIGURES	vi
CHAPTER 1 INTRODUCTION	8
1.1 Project History.....	8
1.2 Objectives.....	9
References	10
CHAPTER 2 MODELING INPUT ERRORS TO IMPROVE UNCERTAINTY ESTIMATES FOR THE PREDICTIONS FROM A ONE-DIMENSIONAL SEDIMENT TRANSPORT MODEL	11
Abstract	11
2.1 Introduction	11
2.2 Methodology	15
2.2.1 Estimating Parameter Uncertainty.....	15
2.2.2 Error Modeling for Hydraulic Input Data.....	17
2.3 Applications	21
2.3.1 Sediment Transport Model	21
2.3.2 Modeling Tachia River.....	25
2.4 Results and Analysis	29
2.4.1 Required Number of Simulations.....	29
2.4.2 Uncertainty in Model Parameters.....	30
2.4.3 Uncertainty in Input Variables	33
2.4.4 Uncertainty in Predictions	35
2.5 Conclusions.....	39
References	42
CHAPTER 3 REDUCED COMPUTATIONAL COST FOR ESTIMATING PREDICTION UNCERTAINTY DUE TO SEDIMENT TRANSPORT MODEL PARAMETERS	46
Abstract	46
3.1 Introduction.....	46
3.2 Methodology for Parameter Uncertainty Assessment.....	52
3.2.1 Existing GLUE and MCMC Methods.....	52

3.2.2 Evolving Latin Hypercube Method	54
3.3 Case Studies for Synthetic Distributions.....	61
3.3.1 A Single Bimodal Distribution	61
3.3.2 Multi-dimensional Gaussian Distributions.....	64
3.4 Case Study for Sediment Transport Model Parameters	68
3.4.1 Application to Tachia River Simulation.....	68
3.4.2 Computational Cost for Uncertainty Estimates	70
3.4.3 Uncertainty in Model Parameters	73
3.4.4 Model Prediction Uncertainty	77
3.5 Conclusions	84
References	87
CHAPTER 4 COMBINING PREDICTIONS AND ASSESSING UNCERTAINTY FROM SEDIMENT TRANSPORT EQUATIONS USING MULTIVARIATE BAYESIAN MODEL AVERAGING.....	
Abstract	93
4.1 Introduction	93
4.2 Methodology	98
4.2.1 Existing BMA Method	98
4.2.2 Multivariate BMA Model.....	100
4.3 Application.....	102
4.3.1 Flume Experiments.....	102
4.4 Results and Analysis	108
4.4.1 Model Weights and Standard Deviations	108
4.4.2 BMA Predictions and Uncertainty	110
4.4.3 Applying Coefficient of Variation.....	118
4.5 Conclusions	120
References	123

LIST OF TABLES

Table 1 Uncertain model parameters of SRH-1D and their sampling ranges.....	24
Table 2 Error models for uncertain input variables and sampling ranges for the input error parameters.....	27
Table 3 Average of root mean squared error values for 2,000 simulations for the calibration and forecast periods using parameter sets sampled after convergence. The lowest values in each period are shown in bold face.	39
Table 4 Median values and IQR ratios of the estimated posterior distributions for SRH-1D model parameters obtained using the three uncertainty methods for the Tachia River case.....	75
Table 5 Correlation coefficients between well-specified four model parameters of SRH-1D estimated by ELH and SCEM-UA in Tachia River case.....	77
Table 6 Available Observations from Two Experiments for the Calibration and Forecast Periods of BMA Modeling.....	104
Table 7 Optimized Eight Parameters of SRH-1D Used for Each Model’s Simulation of Two Experiments.	106
Table 8 Standard Deviations of Each Model Prediction Obtained from BMA Calibration: Multivariate BMA Model Uses Bed Elevation and D50 Data for the Seal et al. (1997) Case and Bed Elevation and Sediment Transport Rate for the Pender et al. (2001) Case.	111
Table 9 Nash-Sutcliffe Coefficient of Efficiency (NSCE) Values for Individual Models and BMA Models from Calibration (Calib.) and forecast (Fore.) Periods of Two Experiments.	116
Table 10 Percentage of Observations Covered by 90% Credible Intervals of BMA Models from Calibration and Forecast Periods of Seal et al. (1997) and Pender et al. (2001) Experiments... ..	117
Table 11 Multivariate BMA Calibration Results from Applying Standard Deviations (STDEV) for Bed Elevation and Coefficients of Variation (CV) for Sediment Transport Rate in the Pender et al. (2001) Case.	118
Table 12 NSCE and Percentage of Observations Covered by the 90% Credible Intervals of the Multivariate BMA Model Applying CV from Calibration and Forecast Periods of the Pender et al. (2001) Case.	119

LIST OF FIGURES

Fig. 1 Conceptual diagram for modeling input data error using a Gaussian distribution.	18
Fig. 2 Input data used as forcing variables in the SRH-1D simulations of the Tachia River where the black markers represent the measured values.	28
Fig. 3 Number of simulations to reach convergence for model and input error parameters.	30
Fig. 4 Histograms for Manning's n using 2,000 samples generated after convergence. Also shown are the median (black point) and the interquartile range (horizontal line) of the sampled values for each case.	31
Fig. 5 Comparison of the medians and interquartile ranges for input error parameter m from Cases 1~7.	34
Fig. 6 Observations and model predictions for sediment deposition volume during the forecast period for the Tachia River using 2,000 parameter sets generated after convergence. Also shown are the points (the vertical lines with asterisk markers) where the internal and downstream boundary conditions are applied.	36
Fig. 7 Percentage of observations covered by the predictive intervals (PI) and the average width ratios of the PI for each case to the PI for Case 0 from the simulations for both the calibration and forecast periods using 2,000 parameter sets sampled after convergence.	38
Fig. 8 Flowchart of the Evolving Latin Hypercube (ELH) algorithm.	55
Fig. 9 (a) Bimodal target distribution and histogram of the generated 5,000 parameter values and (b) the changes in the posterior sample properties during 5,000 computations using ELH and SCEM-UA.	63
Fig. 10 (a) Posterior sample properties during 10,000 computations for Case 2, and (b) sample variances and (c) correlation coefficients estimated during 100,000 computations for Case 3 using ELH and SCEM-UA.	66
Fig. 11 SD changes for SRH-1D model parameters using (a) ELH and (b) SCEM-UA, and (c) the total number of model simulations required to estimate the prediction uncertainty using the three uncertainty methods for the Tachia River case.	71
Fig. 12 Cumulative marginal posterior distributions for SRH-1D model parameters generated using the three uncertainty methods for the Tachia River case.	74
Fig. 13 Observations and mean predictions for sediment net deposition volume in the forecast period of the Tachia River case. The vertical width of the gray region is the 99% credible interval of the predictions.	79
Fig. 14 Observation, mean prediction, and uncertainty distributions for net deposition volume at the cross section located 21 km upstream from the ocean for the forecast period in the Tachia River case.	80

Fig. 15 (a) Nash-Sutcliffe Coefficient of Efficiency (NSCE) values for the mean prediction from each method, (b) standard deviation (SD) of the sample of simulation outputs, (c) Continuous Rank Probability Score (CRPS) values for distributions of the same simulations, and (d) percentage of observations covered by the 99% credible intervals of the same simulations.	82
Fig. 16 Illustration of BMA distribution generated by a weighted-average of normal distributions from four competing models.....	95
Fig. 17 Observations (crosses) and individual model outputs (lines) of (a) bed elevation profile and (b) D50 distribution at 20 hour of the Seal et al. (1997) experiment; (c) bed elevation profile at 32.1 hour and (d) sediment transport rate during the calibration period of the Pender et al. (2001) experiment.....	107
Fig. 18 Weights of individual models determined by BMA using each type of variable from the two experiments.....	109
Fig. 19 Observation (dashed vertical line), and predictions from the individual models (white markers) and the BMA model using sediment transport rate (black marker), predictive distributions of BMA using bed elevation (solid black line with filled circle marker) and multivariate BMA model (solid grey line with filled triangle marker) for bed elevation at 10.5 m upstream at 79.6 hour in the Pender et al. (2001) experiment.....	111
Fig. 20 Observations, predictions and 90% credible intervals from each type of BMA model for bed elevations at 50 hr and D50 at 53 hour in the forecast period of the Seal et al. (1997) experiment.....	113
Fig. 21 Observations, predictions and 90% credible intervals from each type of BMA model for bed elevations at 84.6 hr and sediment transport rate during the forecast period of the Pender et al. (2001) experiment.....	114
Fig. 22 Observations, predictions and 90% credible intervals for sediment transport rate from the multivariate BMA models applying standard deviations (STDEV) and coefficient of variations (CV) during both the calibration and forecast periods of the Pender et al. (2001) experiment..	119

CHAPTER 1

INTRODUCTION

1.1 Project History

The U.S. Bureau of Reclamation (USBR) protects and manages water resources in the western United States. A variety of hydrologic and hydraulic numerical models has been developed and widely used by the USBR to predict impacts of potential river restoration activities. Recently, assessing the uncertainty in predictions from such models has been underscored in the field of hydraulic and sediment transport modeling. Those uncertainties can result from the simplifications and assumptions used in the model's mathematical structure, errors in the model parameter values, and errors in the data used for the model inputs. Colorado State University has been working with USBR to develop a framework to evaluate the uncertainty associated with USBR models. Through this collaboration, several methods using a Bayesian uncertainty framework have been suggested to quantify the uncertainty in the predictions specifically from the Sedimentation and River Hydraulics – One Dimension (SRH-1D) (Huang and Greimann 2013) model. GSA-GLUE (Global sensitivity analysis – Generalized likelihood uncertainty estimation) method (Ruark et al. 2011) was developed to assess uncertainty of individual model parameters. The combination of the MSU (Multivariate shuffled complex evolution Metropolis – uncertainty analysis) and BMA (Bayesian model averaging) methods (Sabatine et al. 2015) was suggested to evaluate the uncertainty of model parameters and the uncertainty associated with the selection of a transport equation in SRH-1D simulations.

1.2 Objectives

The purpose of this research project was to develop and test a formal and efficient framework to assess the uncertainty in the predictions from hydrologic, hydraulic, and sediment transport models. Through this research, we aimed to develop improved methodologies that: (1) require few model simulations and (2) retain enough formality so that data collection and model calibration strategies can be modified to reduce the uncertainty in the model predictions. To constrain the immediate scope, the new methods focused on quantifying how the uncertainties originating from model parameter values, input data, and the model's mathematical structure affect the predictions from the SRH-1D model, but those methods are transferrable to other types of models. The following chapters address three specific objectives that have been accomplished to achieve the primary purpose of this research. Those chapters are summarized below:

1. Simple error models are developed for the input data of a sediment transport model and integrated into an existing Bayesian method in order to determine whether uncertain inputs contribute substantially to the overall uncertainty in the predictions. Input errors are modeled using Gaussian distributions separately for various input data such as discharges, rating curves, cross section coordinates, and benchmark elevations. The means and standard deviations of those distributions are treated as uncertain parameters, and they are estimated within the Bayesian framework for parameter uncertainty. This approach enables a modeler to identify the contribution of each uncertain input to the overall uncertainty, which can suggest strategies to reduce the uncertainty and improve reliability in the model predictions.

2. A new algorithm is developed to improve the efficiency of the uncertainty estimation process for sediment transport model parameters. In order to reduce the computational cost, the new method is designed to use repeated parameter sets in the sample when specifying the probability distributions of parameters instead of generating new but similar parameter sets that require new model simulations, which is the typical approach of existing Markov chain Monte Carlo methods. This new approach can save large numbers of model simulations when evaluating the uncertainty in model predictions due to uncertainty in the parameter values.
3. A multivariate version of the BMA method is developed to assess the uncertainty associated with the selection and application of a transport equation in sediment transport models. The existing BMA method is modified to enable consideration of multiple model output variables and allow the uncertainty associated with each equation to vary with the magnitude of the variables if needed. This methodology can reduce the effects of imperfections in a single model prediction and provide a forecast along with its credible interval to characterize the uncertainty through combining the predictions from a set of competing transport equations.

References

- Huang, J.V., and Greimann, B.P. (2013). *User's Manual for SRH-1D 3.0*. Bureau of Reclamation, U.S. Department of the Interior, Denver.
- Ruark, M.D., Niemann, J.D., Greimann, B.P., and Arabi, M. (2011). "Method for assessing impacts of parameter uncertainty in sediment transport modeling applications." *Journal of Hydraulic Engineering*, 137(6), 623-636.
- Sabatine, S.M., Niemann, J.D., and Greimann, B.P. (2015). "Evaluation of parameter and model uncertainty in simple applications of a 1D sediment transport model." *Journal of Hydraulic Engineering*, 141(5), DOI: 10.1061/(ASCE)HY.1943-7900.0000992.

CHAPTER 2

MODELING INPUT ERRORS TO IMPROVE UNCERTAINTY ESTIMATES FOR THE PREDICTIONS FROM A ONE-DIMENSIONAL SEDIMENT TRANSPORT MODEL

Abstract

Bayesian methods using Markov chain Monte Carlo algorithms have recently been applied to sediment transport models to assess the uncertainty in the model predictions due to the parameter values. Unfortunately, the existing approaches can only attribute overall uncertainty to the parameters, and this limitation is critical because no model can produce accurate forecasts if forced with inaccurate input data. In this research, error models are developed to address the uncertainty in hydraulic input data and integrated into an existing Bayesian method to improve the uncertainty estimates. The input error is modeled using a Gaussian distribution, and the mean and standard deviation are treated as uncertain parameters. This approach is tested by coupling it to the Sedimentation and River Hydraulics - One Dimension (SRH-1D) model and simulating a 23-km reach of the Tachia River in Taiwan. Six types of input data are considered as uncertain: input discharge, two rating curves, vertical and horizontal distance of cross sections, and benchmark elevations. The results indicate that the errors in benchmark elevations have the largest impact on the model prediction uncertainty among those considered. In addition, the predictive intervals from the new method can cover 1.5~2 times more observations and improve the performance up to 6 % than the existing method.

2.1 Introduction

Numerical hydraulic and sediment transport models have been used widely to make predictions about river morphological changes that result from natural or human influences. Those

predictions are often used for decision making in water resources management, which involves the issues of ecological impacts, potential economic loss, and/or risks to human health.

Therefore, a full understanding of the modeling strategy and the uncertainty in those predictions is essential. In the field of sediment transport modeling, the uncertainty has typically been assessed by examining how model predictions spread according to variability in inputted physical properties such as discharge, flow depth, and channel slope (McLean 1985; Cui and Parker 1998; Bunte and Abt, 2005; Pinto et al. 2006; Gaeuman et al., 2009), model parameters like roughness coefficient or critical shear stress (Chang et al. 1993; Yeh et al. 2004; Lai and Greimann 2010), and mathematical equations that are used to address transport mechanism (Wilcock 2001; Davies et al. 2002; Camenen and Larroudé 2003; Bertin et al. 2008). Those traditional approaches are relatively simple to implement, but they do not guarantee the reliability of the predictions because they are often restricted to sensitivity analysis without any calibration process.

Bayesian inference provides a formal way to assess the uncertainty in model predictions by considering the likelihoods, which calculate the model's ability to reproduce the data from the calibration period. Bayesian methods have recently been applied to sediment transport modeling cases to predict erosion within sewer systems (Kanso et al. 2005), sediment entrainment (Wu and Chen 2009), bed elevation and material changes (Ruark et al. 2011; Sabatine et al. 2015), and cohesive sediment behavior (Cho et al. 2016). The Bayesian methods treat a model parameter as a random variable having a posterior probability density function (pdf), which describes the uncertainty in the parameter values given a dataset. Numerous Markov chain Monte Carlo (MCMC) algorithms, which generate parameter samples and use them in model simulations, have been developed to obtain a numerical approximation for that pdf. The algorithm is intended

to sample the parameter values with higher likelihoods more frequently over the simulations by considering the likelihood information from the previous simulations. Once the MCMC algorithm has run sufficient simulations, the parameter sets are sampled from a stationary distribution, which is an approximation of the posterior pdf. The parameter samples generated from that pdf are then used in the simulations for the forecast scenario in order to quantify how the uncertainty due to the parameter values affects the predictions. The key advantages of using the MCMC algorithm are that (1) its sampling processes are potentially efficient in computation time (Vrugt et al. 2003), (2) the likelihood function is a formal way to calculate model accuracy (van Griensven and Meixner 2007), and (3) it can consider the correlations between model parameters, which has shown significant impacts on sediment transport model predictions (Sabatine et al. 2015), because the algorithm develops the joint probability distribution of the multiple parameters when assessing the uncertainty. Readers are referred to Green (2001) to learn the details of the MCMC algorithm.

A distinct limitation in applying Bayesian methods to sediment transport modeling is that the past research usually assigned overall uncertainty on model parameters by ignoring other potential uncertainty sources. Specifically, the forcing variables and boundary conditions used in numerical sediment transport models such as discharge, input sediment flow rate, and channel geometry always include uncertainty due to their inherent unsteadiness, heterogeneity, and difficulty being measured in the real world (Wilcock 2001; Bunte and Abt 2005; Gaeuman et al. 2009). For example, the discharge data generally contain 5% - 15% errors (USGS 1996), and the flow depth can have errors with a standard deviations of 10 % of the measured value in natural rivers (USGS 1992). Adequate characterization of those variables is fundamental to success of modeling because no model can produce accurate predictions if forced with inaccurate input data,

even if the model is well founded in physical theory or empirically justified by past performance (Clement and Piegay 2005). Moreover, in hydrologic applications, it has been demonstrated that not accounting for errors in the input data leads to corrupted parameter uncertainty estimates as well as unreliable predictions (Ajami et al. 2007).

A number of methods have been proposed to consider the errors in forcing variables when estimating the uncertainty, but those studies only dealt with the measured rainfall data particularly for hydrologic modeling (Kavetski 2002; Carpenter and Georgakakos 2006; Huard and Mailhot 2006; Ajami et al. 2007; Vrugt et al. 2008). To predict the sediment budget in fluvial beds, Schmelter et al. (2011, 2012, and 2015) and Schmelter and Stevens (2013) applied a variance to the predictions in order to address the uncertainty due to mathematical misspecification, measurement error, and random variations. The uncertain variance parameter was calibrated using an MCMC algorithm, and it was found that such errors increase the overall uncertainty. However, this approach is not able to evaluate how much uncertainty in model predictions comes from the uncertainty in input data or identify the errors in each of various forcing variables because the different error sources are lumped together in a single variance parameter. In addition, they used a single transport rate equation, rather than a numerical model, as a sediment transport model so that neither the number of model simulations nor the computational times required for the MCMC algorithm were investigated.

The goal of this research is to develop error models to address the uncertainty in various hydraulic input data and evaluate how the uncertainty estimates are improved when those error models are integrated in an existing Bayesian method using an MCMC algorithm. The potential input errors are modeled using Gaussian distributions, and the means and standard deviations are treated as uncertain parameters, which can be estimated jointly with model parameters. To test

the advanced method, deposition volumes along the 23-km reach of the Tachia River in Taiwan are simulated using the Sedimentation and River Hydraulics - One Dimension (SRH-1D) model. Nine model parameters and six different input data used as forcing variables in SRH-1D simulations are considered. The benefits of modeling the input errors in the uncertainty analysis are evaluated by comparing (1) the number of simulations required for the MCMC algorithm, (2) the posterior pdfs of the model and input error parameters, (3) and the coverage of the observed data by the prediction intervals to those of the existing method.

2.2 Methodology

2.2.1 Estimating Parameter Uncertainty

Multiple parameters in a sediment transport model are uncertain and need to be determined by the modeler. In the Bayesian inference paradigm, those uncertain parameters can be treated as random variables having a joint probability distribution, which can be written as:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto L(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \quad (1)$$

where $p(\boldsymbol{\theta}|\mathbf{y})$ is the joint posterior pdf, which describes the uncertainty in a set of parameters $\boldsymbol{\theta}$ given a calibration dataset \mathbf{y} . The likelihood $L(\mathbf{y}|\boldsymbol{\theta})$ represents the model's ability to reproduce the dataset \mathbf{y} when parameters $\boldsymbol{\theta}$ are used, and the prior pdf $p(\boldsymbol{\theta})$ summarizes the information about the parameters $\boldsymbol{\theta}$ before considering any calibration data (Christensen et al. 2011).

A Shuffled Complex Evolution Metropolis – Uncertainty Analysis (SCEM-UA) (Vrugt et al. 2003), which is one of the most advanced MCMC algorithms, is employed to approximate the joint posterior pdf $p(\boldsymbol{\theta}|\mathbf{y})$ in this study. Assuming no prior information about the parameters, the SCEM-UA algorithm starts by generating an initial population (≈ 250) of parameter sets from a

uniform prior joint distribution, which is defined within parameter ranges pre-specified by a user. Then, the posterior density of each parameter set is evaluated using the model simulation and observations from the calibration period as:

$$P(\boldsymbol{\theta}|\mathbf{y}) \propto \left\{ \sum_{i=1}^N [e_i(\boldsymbol{\theta})]^2 \right\}^{-N/2} \quad (2)$$

where e_i represents the model error given the parameter set at a time or space i , and N is the number of available measurements in the calibration dataset \mathbf{y} . The equation contains a primary assumption that the model errors are mutually independent, Gaussian distributed, with constant variance (Box and Tiao 1973). The value from the right side of Eq. (2) is proportional to the posterior density (and is often called the posterior density for simplicity). Hence, the posterior density of different parameter sets can be compared using Eq. (2). After that, the parameter samples are partitioned into a number of complexes (≈ 10 groups), and the population of each complex is updated in parallel using a Metropolis algorithm (Metropolis et al. 1953). The highest posterior density parameter set in each complex is treated as a starting point of each updating sequence. For each sequence, a new parameter set $\boldsymbol{\theta}^*$ is generated from a multivariate normal distribution, which is centered on either the current parameter set of the sequence or the mean of the parameters in the complex with the covariance structure inferred from the parameters in the complex, and its posterior density is assessed using the model simulation. If the posterior density of the new parameter set $\boldsymbol{\theta}^*$ is larger than the current one $\boldsymbol{\theta}$, then the new parameter set is retained and the highest posterior density member of complex is replaced by the new set. Otherwise, the new parameter set can replace the current parameter set with a randomly specified probability. Such sampling behavior that can accept a trial with lower posterior density helps to escape from locally optimal areas. This updating continues iteratively based on the last

retained value for each complex sequence. After a few iterations (≈ 5) of the updating, the members of all complexes are recombined, shuffled, and re-divided into complexes. Once the algorithm has repeated sufficient updating and shuffling, all complexes have the same properties such as the average and variance among the contained parameter sets, which implies algorithm convergence.

To diagnose the convergence of SCEM-UA algorithm, Scale Reduction Score (SRS) (Gelman and Rubin 1992) is used. The SRS is the ratio of the variance of the average parameter values from each complex to the average of the variances of parameter values within each complex. If the SRS for all parameters is below 1.2, then adequate convergence is indicated (Vrugt et al. 2003, Sabatine et al. 2015) and it is reasonable to believe that the samples are generated from a stationary distribution, which can be considered as being consistent with joint posterior pdf of the uncertain parameters (Wu and Chen 2009).

2.2.2 Error Modeling for Hydraulic Input Data

In this research, potential errors in an input variable are modeled using a Gaussian distribution where the mean represents the measurement bias and the standard deviation reflects the independent errors at each measurement (Fig. 1). Both the mean and standard deviation are treated as uncertain parameters in order to account for input data uncertainty in the Bayesian method. This approach follows the idea, used for rainfall error in hydrologic modeling cases, that the observed data (what we have) are assumed to be from the true values (what we want to know) corrupted by random numbers at each measurement (Ajami et al. 2007). This assumption can be written as:

$$\varepsilon_{ij} \sim N(m_i, \sigma_i) \quad (3)$$

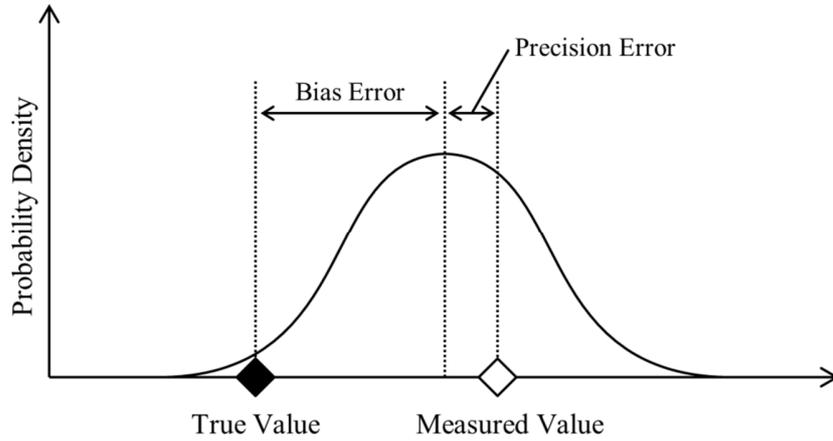


Fig. 1 Conceptual diagram for modeling input data error using a Gaussian distribution.

where e_{ij} is the error at measurement i for the input variable j , and a single error value is generated from an identical normal distribution with mean m_j and standard deviation σ_j for each measurement.

Input error models, which relate the errors and the observed data, can be developed separately for several input variables based on the error characteristics. The discharge data is usually used for an upstream boundary condition (BC) in a model simulation, and a “fair” measurement performance shows that 95% of the measured values are within 10% of the “true” discharge value (USGS 1996). It implies that the measurement errors have a standard deviation of 5% of the “true” value so that the errors are expected to depend on the variable’s magnitude. Such non-homogeneous characteristics of the errors can be addressed using a multiplicative error model (Ajami et al. 2007) as:

$$\hat{x}_{ij} = x_{ij}\epsilon_{ij} \quad (4)$$

where \hat{x}_{ij} is the observed value and x_{ij} is the estimate of true value at the measurement i for the input variable j . Unlike the discharge, the errors in water surface elevation data are independent of the scale of its measured value because those elevations are often defined relative to an arbitrary datum (sea level). By treating those errors as being homogeneous, an additive error model is available:

$$\hat{x}_{ij} = x_{ij} + \varepsilon_{ij} \quad (5)$$

This model would be helpful to reflect the errors in a rating curve, which controls the water surface elevation given discharge at internal or downstream BCs. Specifically, the rating curve inherently possesses large uncertainty because the relationship between discharge and water surface elevation would vary in time as channel geometry changes during the simulation.

The geometry data are obtained by measuring horizontal and vertical distances from a benchmark point of known elevation at each cross section. The horizontal distance errors are usually treated as negligible because their magnitude is less than 1 % of the measured value (USGS 1992). However, the horizontal distances might be measured shorter than the actual length if the cross section is not orthogonal to the channel direction. The multiplicative error model can be modified to account for this error as:

$$\hat{x}_{ij} = x_{ij} \frac{1}{\cos(\varepsilon_{ij})} \quad (6)$$

where ε_{ij} is the angle from the orthogonal line of the channel direction, \hat{x}_{ij} and x_{ij} represent the measured and true distances from the benchmark point, respectively. For vertical distances, the measurements can have errors with a standard deviation of 10 % of flow depth in natural rivers

(USGS 1992). Those errors might originate from both the existence of dunes (or antidunes) and the uneven distributions of large bed materials (cobbles, rocks, and boulders) in active mobile channel bed. The vertical distance can apply to the multiplicative error model (Eq. (4)) as the error scales are proportional to its magnitude.

A set of input error parameters $\boldsymbol{\eta} = \{m_1, \sigma_1, \dots, m_J, \sigma_J\}$, where J is the total number of input error models considered, can be integrated into the governing equation of Bayesian uncertainty method (Eq. (1)) as:

$$p(\boldsymbol{\theta}, \boldsymbol{\eta} | \mathbf{y}) \propto L(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\eta}) p(\boldsymbol{\theta}, \boldsymbol{\eta}) \quad (7)$$

where $p(\boldsymbol{\theta}, \boldsymbol{\eta} | \mathbf{y})$ is the joint posterior pdf of model parameters $\boldsymbol{\theta}$ and input error parameters $\boldsymbol{\eta}$, which describes the uncertainty associated with model parameters and input data simultaneously given calibration dataset \mathbf{y} . To specify the posterior density $P(\boldsymbol{\theta}, \boldsymbol{\eta} | \mathbf{y})$ in the SCEM-UA algorithm, the model is simulated using the sampled model parameters $\boldsymbol{\theta}$ and the input variables x_{ij} , which are computed from Eq. (4), (5), or (6) using the sampled input error parameters $\boldsymbol{\eta}$. If the measured input data are correct, it is expected that (1) the posterior distributions of the model parameters would be the same as if the uncertainty analysis did not consider the input errors (same as the results from using Eq. (1)), (2) the mean would concentrate around one for the multiplicative error models and zero for the additive error models (it means the data are unbiased), and (3) the standard deviation would approach zero. This approach allows a user to avoid excessive computational costs in using the MCMC algorithm because it only requires identifying two additional uncertain parameters for each type of input data instead of calibrating the all measurement points considered.

2.3 Applications

2.3.1 Sediment Transport Model

The proposed method is tested by coupling it to SRH-1D model, which was developed and is currently used by the U.S. Bureau of Reclamation to simulate flows and sediment transport in open channels and river networks. The section summarizes the mathematical structure of the SRH-1D model including the role of its key model parameters, but readers are referred to Huang and Greimann (2013) for details.

SRH-1D computes flow hydraulics by solving the energy equation for steady, gradually varied flow using the standard step method, and a time series of input flow rate is used as an upstream BC. The energy equation between downstream cross section i and upstream cross section $i + 1$ is expressed as:

$$z_{i+1} + \alpha_{i+1} \frac{U_{i+1}^2}{2g} = z_i + \alpha_i \frac{U_i^2}{2g} + h_c + h_f \quad (8)$$

where z represents the water surface elevation, α is the kinematic coefficient, U is the cross-sectional averaged velocity, g is the gravitational acceleration. The contraction or expansion loss h_c is computed based on the relationship between the velocity heads at the two cross sections, and the friction loss h_f is calculated from the friction slopes at the cross sections, which can be determined using Manning's roughness coefficient n .

For sediment transport computations, Exner equation routing is used to calculate the changes of the sediment volume in the bed. By ignoring changes in suspended sediment concentration over time, the Exner equation expresses mass conservation as:

$$\frac{\partial Q_s}{\partial x} + (1-\phi) \frac{\partial A_d}{\partial t} - q_{sl} = 0 \quad (9)$$

where x is the longitudinal direction, t is time, Q_s is volumetric sediment discharge, ϕ is porosity, A_d is volume of bed sediment per unit length, and q_{sl} is lateral sediment input rate per unit length.

The volumetric sediment discharge Q_s is computed by calculating the transport capacity separately for each predefined grain size class.

Among several equations available in SRH-1D, the equation suggested by Wu et al. (2000) is used here to compute the transport capacity. The equation computes total bed material load for grain size class k by combining the bed load q_{bk} and suspended load q_{sk} , which are calculated separately as:

$$q_{bk} = 0.0053 p_k \sqrt{g [(\rho_s/\rho) - 1]} d_k^3 \left[\left(\frac{0.05 d_{50}^{1/6}}{n} \right)^{1.5} \frac{\tau_b}{\tau_{ck}} - 1 \right]^{2.2} \quad (10)$$

$$q_{sk} = 0.0000262 p_k \sqrt{g [(\rho_s/\rho) - 1]} d_k^3 \left[\frac{U}{w_{fk}} \left(\frac{\tau_b}{\tau_{ck}} - 1 \right) \right]^{1.74} \quad (11)$$

where ρ is the density of water, ρ_s is the density of the sediment, d_{50} is the median grain diameter, τ_b is bed shear stress, p_k is the fraction of material in class k , d_k is the median diameter of class k , w_{fk} is the fall velocity of particles in class k , and τ_{ck} is a dimensional critical shear stress for particle in class k , which is computed as:

$$\tau_{ck} = \theta_r [(\rho_s/\rho) - 1] d_k \xi_k \quad (12)$$

where θ_r is the non-dimensional reference shear stress. ξ_k is the hiding and exposure function, which accounts for the reduction in the reference shear stress for relatively large particles and the increase in the reference shear stress for relatively small particles. It is determined:

$$\xi_k = \left(\frac{p_{hk}}{p_{ek}} \right)^\lambda \quad (13)$$

where λ is a hiding and exposure coefficient, p_{hk} and p_{ek} are the hidden and exposed probabilities of particles in class k .

SRH-1D assumes that the computed transport capacity is reached over some length controlled by the total adaptation length L_{total} , which can be calculated as follows:

$$L_{total} = (1 - f_s)L_b + \frac{\zeta f_s}{w_{fk}} U h \quad (14)$$

where f_s is the fraction of suspended load as computed in Greimann et al. (2008), h is the hydraulic depth, and the bed load adaptation length L_b is calculated as:

$$L_b = b_L h \quad (15)$$

where the b_L is the bed load adaptation length multiplier. The suspended sediment recovery factor ζ possesses different values where deposition (ζ_d) or scour (ζ_s) occurs.

SRH-1D models the bed material mixing by dividing the bed into one active layer, which is a thin upper zone containing the bed materials available for transport, and several inactive layers below the active layer. All particles in the active layer are assumed to be equally exposed to the flow, and the thickness of the active layer can be calculated by multiplying the geometric mean of the largest grain size class by the user-specified constant, which is the active layer thickness

multiplier parameter n_{alt} . In addition, the weight of bed load fraction parameter χ , which controls the weighting of the grain size distribution of bed load when materials transfer from the active layer to the inactive layers during deposition, should also be specified by a user.

Table 1 summarizes the nine model parameters and their feasible ranges for natural river simulations. The range for Manning's roughness n spans the flow resistances in natural rivers with smooth meanders, pools, and riffles (Limeneros 1970). The reference shear stress θ_r spans the observed θ_r values from 45 study sites of gravel-bed streams (Mueller et al. 2005), and the porosity ϕ range considers mixtures of sand, gravel, and cobbles (Frings et al. 2011). The remaining six parameters (λ , n_{alt} , ζ_d , ζ_s , b_L , χ) have ranges based on the suggestions of the SRH-1D developers (Lai and Greimann 2010).

Table 1 Uncertain model parameters of SRH-1D and their sampling ranges.

Parameters	Ranges
Manning's roughness coefficient n	0.035 ~ 0.045
Reference shear stress θ_r	0.01 ~ 0.10
Hiding and exposure coefficient λ	0 ~ 1
Active layer thickness multiplier n_{alt}	0.5 ~ 5
Deposition recovery factor ζ_d	0.05 ~ 1
Scour recovery factor ζ_s	0.05 ~ 1
Bed load adaptation length multiplier b_L	0 ~ 25
Weight of bed load fraction χ	0 ~ 1
Porosity ϕ	0.25 ~ 0.40

2.3.2 Modeling Tachia River

Simulations for sediment erosion and deposition volumes along the 23-km reach of the Tachia River in Taiwan (Lai and Greimann, 2010) are considered as a case study to evaluate the new method. A natural river was chosen because it includes various uncertainty sources such as the varied channel geometry, input discharge, and grade controls, which can be examined by the suggested method. The case is also able to consider interactions between scour and sedimentation processes within a single simulation.

Severe erosion has occurred in the Tachia River due to the lack of sediment supply downstream of Shih-gang Dam, which is located 23 km upstream from the ocean on the river. The reach has an average slope of 0.0113. The dominant substrate is cobbles and gravels, and the bed material sizes measured in 2007 ranged from 0.125 mm to 512 mm (fine sand to small boulder) with a median size D_{50} of 108 mm. Deposited sediment volumes during two periods (2001 to 2005 and 2006 to 2009) were measured at cross sections along the reach from the ocean to the dam. The historical data indicates that the erosion might continue to progress downstream with large flow events in the future as the erosion occurred primarily from the dam to approximately 5 km downstream during 2001 to 2005 and to about 8 km downstream during 2006 to 2009. In this research, the deposition volume at each cross section is considered as the variable of interest for uncertainty analysis, and the datasets are separately treated as a calibration period (2001 to 2005) and a forecast period (2006 to 2009). The preliminary test showed that the residuals for the deposition volumes, calculated using the calibrated model outputs, satisfy the assumptions of SCEM-UA (that the residuals should be mutually independent, Gaussian distributed, with constant variance).

Input error models are developed for six input variables used in the simulation of the Tachia River, and the input error parameters are specified to capture the feasible ranges defined by the measurement characteristics (Table 2). Prior to investigating the input uncertainty, Case 0 implemented the uncertainty analysis only for the nine model parameters. Through Cases 1~6, two input error parameters (m and σ) were added when estimating the uncertainty to consider the uncertainty in each of the input variables separately. Case 1 used a multiplicative error model for discharge data at Shih-gang Dam from 2001 to 2009 (Fig. 2a), which is used for the upstream BC. Case 2 and Case 3 applied additive error models to the rating curves used for the internal and downstream BCs, respectively. The internal BC was set to the water surface elevations at the middle of the river using discharge computed from HEC-RAS simulations (WRA, 2005) (Fig. 2b), and the downstream BC was set to a critical depth at the mouth of the river from the same simulations (Fig. 2c). Case 4 modeled the errors in the angles of measuring cross sections (Fig. 2d). The ranges for both error parameters were constrained to prevent the angle is over 45° with 95% confidence. Case 5 considered the vertical distance errors in cross section geometry using a multiplicative error model. Case 6 treated the elevations of benchmark points of the cross sections as uncertain. Typically, the benchmark elevation is calculated relative to the elevations of nearby benchmarks in a network extending from a datum (mean sea level), and the measurements usually have errors less than 1 mm, which are negligible. However, in the Tachia River case, no information is available about how they were determined or where the benchmarks are located. For this reason, this approach assumed that the benchmark elevations are the same as the bankfull elevations at each cross section (Fig. 2e), and an additive error model was applied. At last, Case 7 used a total of 21 uncertain parameters to include the uncertainty from both the nine model parameters and the six input variables listed above

simultaneously. This case is helpful to explore the impact of the correlations between the errors in different input variables on the uncertainty estimates.

Table 2 Error models for uncertain input variables and sampling ranges for the input error parameters.

Case	Uncertain Input Variables	Error Models	Parameters	Ranges
0	None			
1	Discharge at Upstream BC	$x = \tilde{x}\varepsilon_1$	m_1 σ_1	0.95 ~ 1.05 0 ~ 0.1
2	Rating Curve at Internal BC	$x = \tilde{x} + \varepsilon_2$	m_2 σ_2	-2 ~ 2 0 ~ 0.05
3	Rating Curve at Downstream BC	$x = \tilde{x} + \varepsilon_3$	m_3 σ_3	-2 ~ 2 0 ~ 0.05
4	Cross Section Angles where x is horizontal distance from benchmark	$x = \tilde{x}/\cos \varepsilon_4$	m_4 σ_4	0 ~ 15 0 ~ 15
5	Vertical Distance at Cross Sections	$x = \tilde{x}\varepsilon_5$	m_5 σ_5	0.9 ~ 1.1 0 ~ 0.15
6	Benchmark Elevations	$x = \tilde{x} + \varepsilon_6$	m_6 σ_6	-2 ~ 2 0 ~ 0.5
7	All Data Used in Cases 1~6			

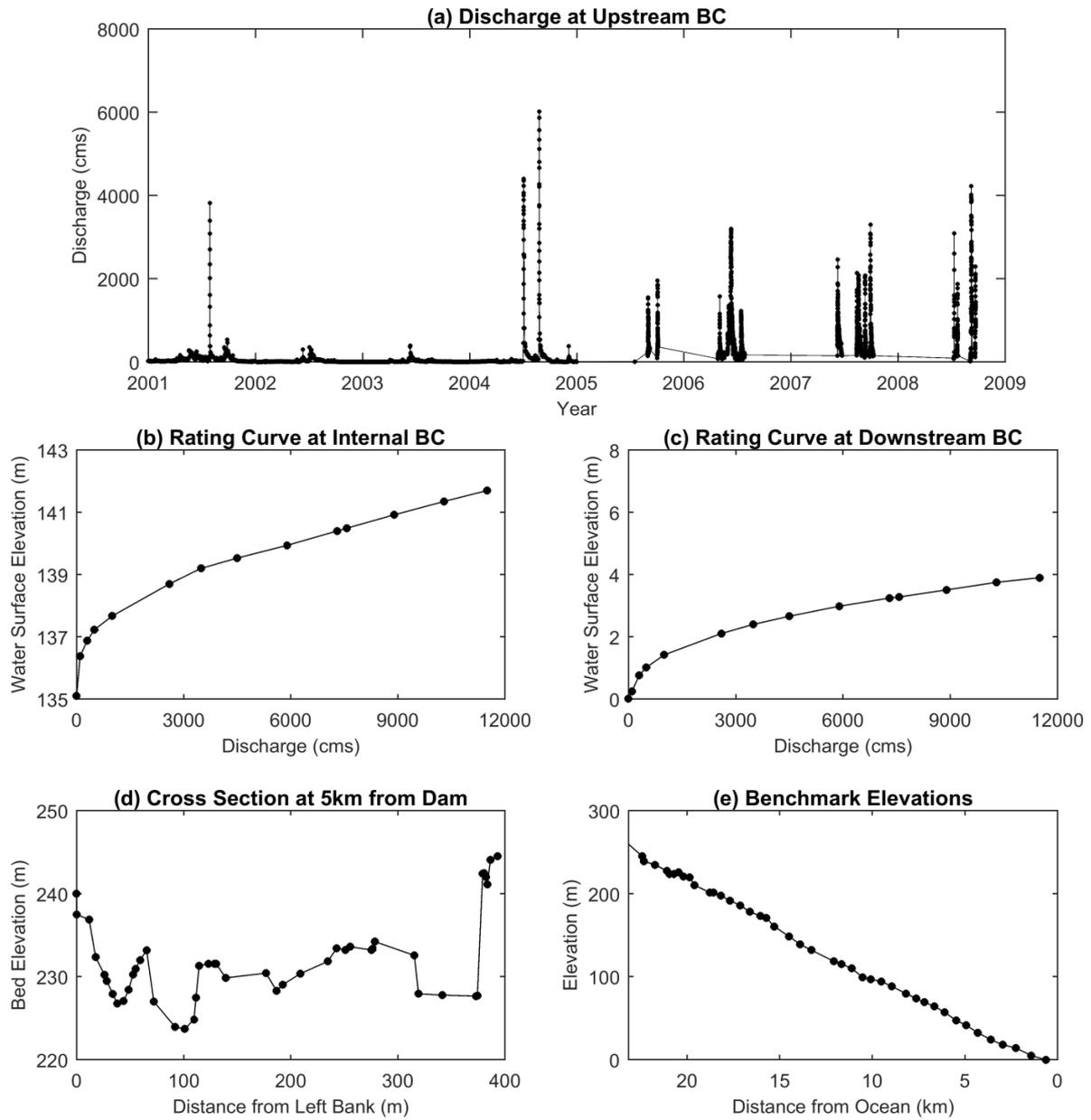


Fig. 2 Input data used as forcing variables in the SRH-1D simulations of the Tachia River where the black markers represent the measured values.

2.4 Results and Analysis

2.4.1 Required Number of Simulations

The number of simulations required to reach convergence of the SCEM-UA algorithm for the Tachia River cases were determined based on the SRS criteria (Gelman and Rubin 1992) (Fig. 3). Each bar represents the simulation numbers where the SRS value for each parameter goes below 1.2. The gray bars for model parameters are placed following the order in Table 1, the black bars are for input error parameters m and σ for each case, and the numbers above the bars indicate the convergence of each case.

Compared to Case 0, which considered only model parameter uncertainty, Cases 1, 2, 3, 4 and 6 ran about 1.5~2.5 times more simulations, but Cases 5 required five times more simulations. This difference should be expected because the latter case treats many more elements as uncertain compared to the other cases. Specifically, Case 1 varies the time series of discharge data but the same discharge values are applied over the entire reach during the simulation. Cases 2 and 3 cause only local changes near the BCs. Case 4 relaxes the channel widths, but the horizontal coordinates at each cross section change proportional to the distance from the benchmark point. In Case 6, the elevation of each cross section can vary independently, but the shape of the cross sections remains the same. On the other hand, the channel geometry varied tremendously every simulation in Case 5 because all points of the cross section data are allowed to move independently in the vertical direction. In addition, the standard deviation error parameter, which controls the variability of vertical point errors, is one of the parameters that needed many simulations to converge in this case. This makes sense because a parameter related

to a large number of uncertain factors would be primarily responsible for an increase in the computational cost of estimating uncertainty.

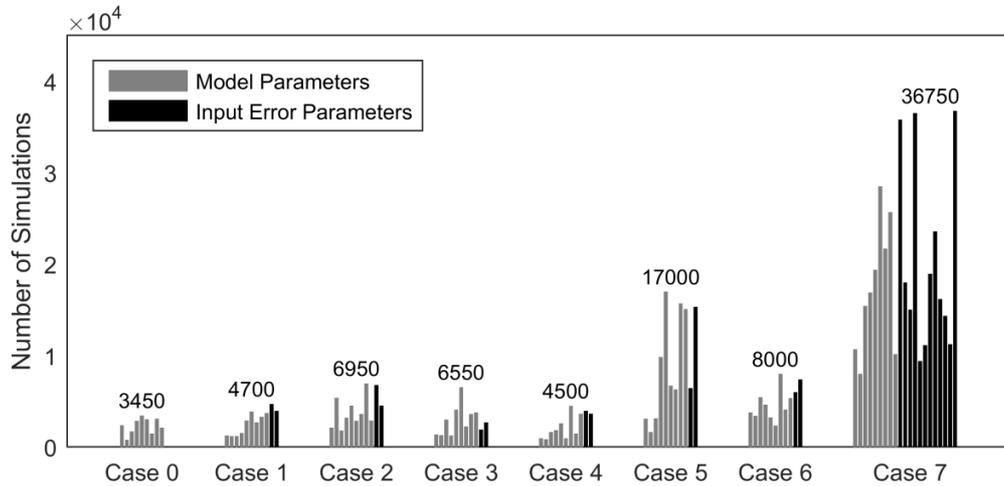


Fig. 3 Number of simulations to reach convergence for model and input error parameters.

Case 7 required much more simulations to estimate the uncertainty including all those uncertain input variables than the other cases. For example, the mean error for discharge m_1 , standard deviation errors for internal BC rating curve σ_2 and benchmark elevations σ_6 achieved convergence after more than 35,000 simulations, which are remarkably later than the model parameters. However the required simulations for those input error parameters were similar to those for the model parameters in Cases 1, 2, and 6. This result might imply that the joint posterior distribution is more complex and harder to specify when it considers the interactions between forcing variables compared to when it only deals with a single uncertain input.

2.4.2 Uncertainty in Model Parameters

The marginal posterior pdf for a single parameter can be approximated using the histogram of the parameter samples generated after convergence, and the distribution describes the uncertainty in the parameter values after the analysis. Fig. 4 presents the marginal posterior pdfs for

Manning's n . For each case, the horizontal line shows the interquartile range (IQR), which indicates the difference between the 25 % and 75 % quantiles of the sampled values, and the median point is also shown with its value. The percentage value is the ratio of the IQR to the sampling range, which is specified at the beginning of the analysis (Table 1). The IQR would be placed near the middle of the sampling range with the ratio value of 50 % if the parameter is sampled independently without considering any calibration data. In this research, the IQR is expected to decrease, which indicates the reduction in the parameter uncertainty, as the SCEM-UA algorithm calibrates the parameters. Specifically, the IQR ratio value would approach 0 % if the algorithm converges to a single parameter value.

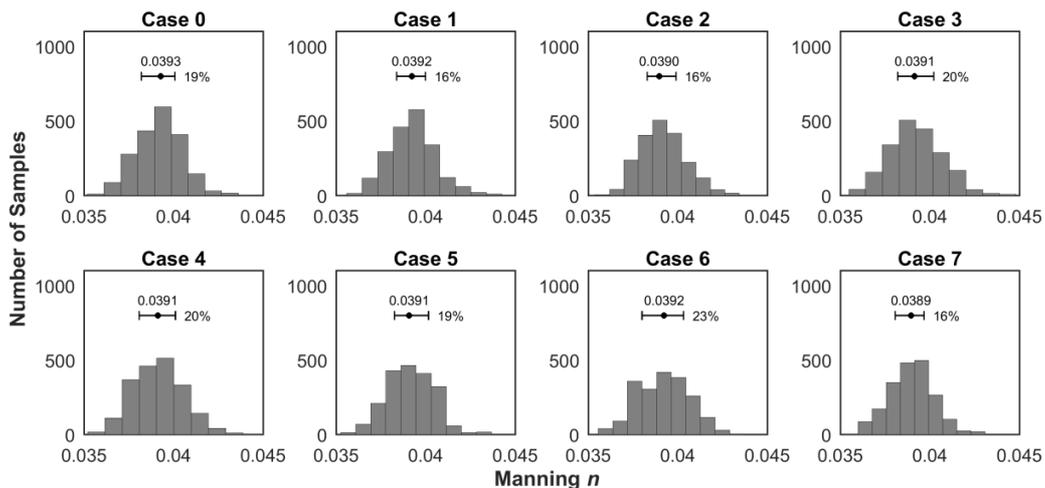


Fig. 4 Histograms for Manning's n using 2,000 samples generated after convergence. Also shown are the median (black point) and the interquartile range (horizontal line) of the sampled values for each case.

Considering input uncertainty leads to only small changes in the posterior pdfs for Manning's n compared to Case 0 (Fig. 4). The medians are around 0.039, and the IQR ratios are 16~23 % among the cases, which is acceptable for natural rivers including the smooth meander, pool, and riffles. In addition to Manning's n , the reference shear stress θ_r and hiding and exposure coefficient λ are well calibrated and have the IQRs about 5 % and 15 %, respectively, but their posterior pdfs are rarely changed over the cases. These three parameters have shown a high impact (sensitivity) on the SRH-1D simulations for bed profile, which is highly related to the deposition volume considered here, in both the erosion and deposition cases (Ruark et al. 2011). For these three parameters, the changes in posterior pdfs are too small to address the differences in sediment transport behavior, and it might imply that all the considered input data are correct. Another reason for the small changes might be that these parameters usually depend on the composition of bed material sizes whereas all the input uncertainties considered here are related to the errors in flow rate or channel geometry. Specifically, Manning's n can reflect the particle roughness in flow modeling, and the reference shear stress θ_r and hiding and exposure coefficient λ control the motion of particles separately for each grain size class as described in Eqs. (12) and (13). As the bed material composition is assumed to be correct in this research and applied identically for all simulations, the parameter uncertainty estimates do not differ over the cases. Unlike those three parameters, the other six parameters (n_{alt} , ζ_d , ζ_s , b_L , χ , ϕ) have IQRs that are usually larger than 40 %, which means that they are poorly calibrated by the algorithm and still remain highly uncertain. These six parameters have little impact on the SRH-1D model simulations of the Tachia River so the parameter values with higher posterior densities cannot be specified well. This also produces similar posterior pdfs for those parameters among the given cases.

Form roughness reflects the flow resistance caused by the channel geometry, and it might be able to explain the differences in the IQRs for Manning's n . The posterior pdf from Case 6 has an IQR 4% wider than that from Case 0, which means the calibrated Manning's n contains more uncertainty when the benchmark elevations are treated as uncertain during the estimation process. The local channel slopes between the adjacent cross sections vary as the benchmark elevations are relaxed in Case 6, so Manning's n is also allowed to have more variability because the slope changes affect the solution of Manning's equation. On the other hand, Case 7 (which is one of the lowest IQRs) shows 3% narrower IQR than Case 0, whereas a wider IQR is expected because it includes six uncertain input variables. This IQR reduction can originate from the correlations between the uncertainties of different input variables. For example, the errors in rating curves can interact with both the errors in the vertical measurements and the benchmark elevations because they are all related to the channel geometry and strongly affect the relationship between water surface elevation and discharge.

2.4.3 Uncertainty in Input Variables

Fig. 5 compares the medians and IQRs for input error parameters $m_1 \sim m_6$ obtained from Cases 1~6 separately and from Case 7. Based on the error model types, m_1 and m_5 are expected to be centered on one, and $m_2, m_3, m_4,$ and m_6 would be on zero when the input data are unbiased. The parameter values calibrated from each of Cases 1~6 indicate that the bias in the discharge and the rating curves for the internal and downstream BCs are negligible even though they still contain uncertainty with IQRs of 27~40 %. The median for the cross section measurement angles is about 9 degrees, and it suggests the true channel widths might be about 90% of the measured widths (Eq. (6)). Vertical distance measurements for the cross sections are positively biased by about 3 % of the measured value. Benchmark elevations are also measured with positive bias of

0.11 m, which means the actual elevations should be 0.11 m lower than the measured points.

Overall, the IQRs indicate that the uncertainty in the rating curve at the downstream BC (m_3), the vertical measurements (m_5), and benchmark elevations (m_6) are relatively small.

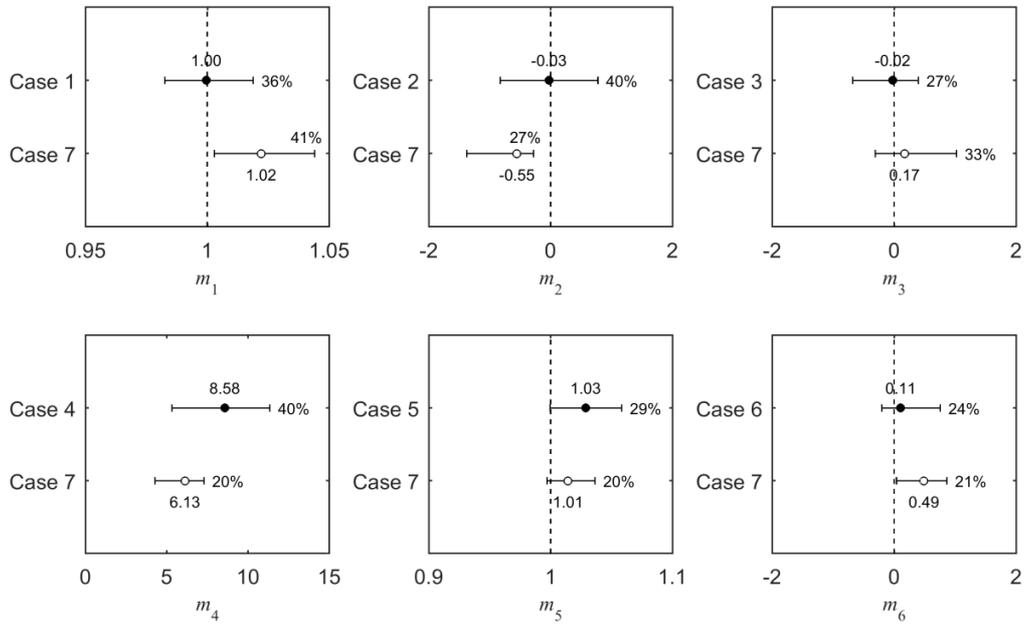


Fig. 5 Comparison of the medians and interquartile ranges for input error parameter m from Cases 1~7.

When the mean input error parameters are jointly specified by a single calibration process in Case 7, both the medians and IQRs of the posterior pdfs are notably changed from the above cases (Fig. 5). For example, the results suggest that the discharge data possess 2 % bias, and the rating curves have biases of -0.55 m and 0.17 m for the internal and downstream BCs, respectively. Both the cross section angles and the vertical measurements for the cross sections show less bias than those identified from the separate cases. For the benchmark elevations, the bias in the measurement data shows a 0.38 cm increase, but it has a 3 % narrower IQR, which means the calibrated value is less uncertain. Moreover, Case 7 presents noticeable reductions in the IQRs (9~20 %) for the three parameters (m_2, m_4, m_5) that are strongly related to both channel geometry and flow conditions.

The standard deviation parameters $\sigma_1 \sim \sigma_6$ are expected to have posterior pdfs that approach zero when the forcing variables include no errors. Overall, both the medians and IQRs from all cases are usually located near the middle of the pre-specified ranges. Only little differences in those values are observed between the cases like the median for σ_6 is 0.25 m from Case 6 and 0.23 m from Case 7.

2.4.4 Uncertainty in Predictions

Predictive intervals (PI) for sediment deposition volumes in the Tachia River are produced by the model simulations for the forest period using 2,000 parameter samples generated after convergence (Fig. 6) in order to evaluate how the uncertainty in the model parameters and input variables is propagated to the model forecasts. The input error models used in the calibration for each case were also applied to the forecast simulations. The spread of the predictions shows the forecast uncertainty originating from both the model parameter and input data uncertainties.

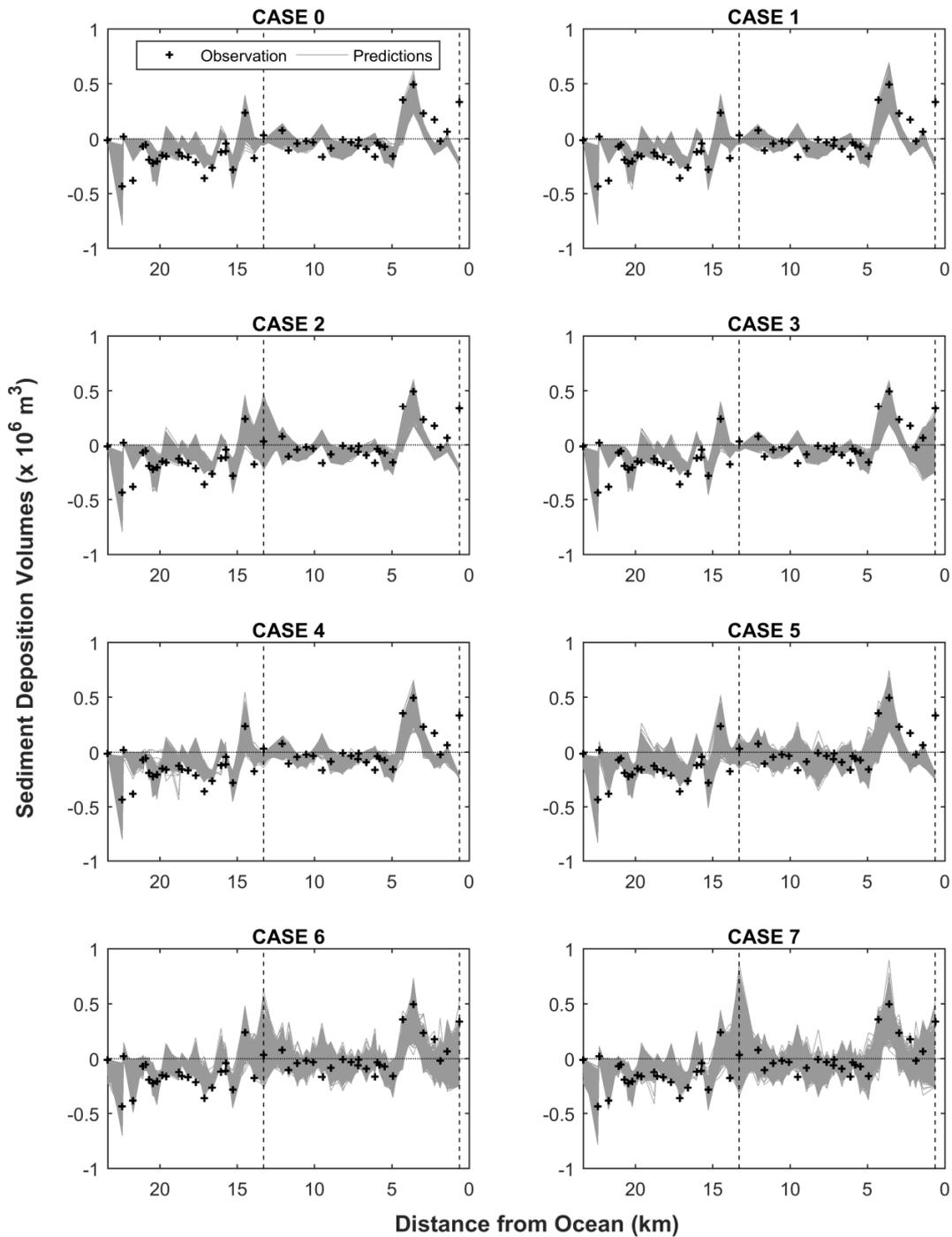


Fig. 6 Observations and model predictions for sediment deposition volume during the forecast period for the Tachia River using 2,000 parameter sets generated after convergence. Also shown are the points (the vertical lines with asterisk markers) where the internal and downstream boundary conditions are applied.

Cases 1~4 show similar PIs to Case 0 along the river reach except that Cases 2 and 3 have wider intervals near the points where the uncertain rating curves are applied as BCs. Case 5 shows the PI widths increase evenly along the reach compared to Case 0, and uneven and large increases in the uncertainty bounds are found in Case 6. The large forecast uncertainty in Case 6 might originate from the spatial correlation of the uncertain variables because the uncertainty in a single benchmark elevation affects not only the associated cross section but also the slopes to both the adjacent upstream and downstream cross sections. In Case 7, the PIs have similar profiles with those from Case 6, but they have slightly narrower intervals over the reach except near the internal BC. The upper bound of the PI at that point is almost same as the sum of the upper bounds from Case 2 and Case 6, which implies the predictive uncertainties from the different sources are aggregated.

The percentages of observations covered by the PIs from each case are shown in Fig. 7a. Overall, the estimated uncertainty bounds do not cover 100 % of the data for both the calibration and forecast periods, which indicates an underestimation of the uncertainty. This underestimate might be caused by neglecting the other sources of uncertainty (such as errors in the observations or structural deficiencies in the sediment transport model) or by deviations from the assumptions used in the SCEM-UA algorithm or the input error models. However, they reveal that the estimated uncertainty bounds generated by considering more uncertain input variables as well as the model parameters contain more observations. Specifically, the PI from Case 7 covers more than 1.5~2 times the observations that are covered by the PI from Case 0 in the calibration and forecast periods. Besides the observation coverage, a comparison of the PI widths averaged along the river reach from each case to those from Case 0 (Fig. 7b) indicates that Case 7 covers more observations in the calibration period even though it has narrower PIs than Case 6.

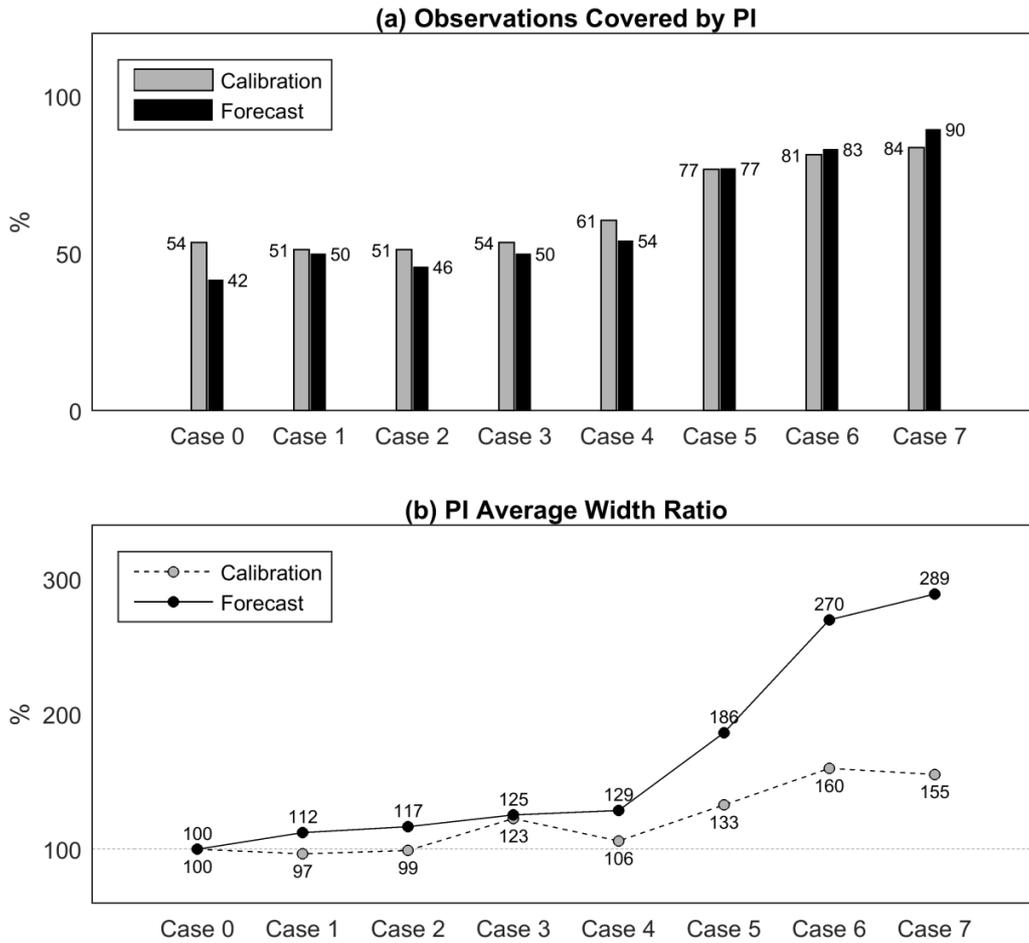


Fig. 7 Percentage of observations covered by the predictive intervals (PI) and the average width ratios of the PI for each case to the PI for Case 0 from the simulations for both the calibration and forecast periods using 2,000 parameter sets sampled after convergence.

Moreover, the average of the root mean squared errors (RMSEs) over the simulations shows that Case 7 provides the best performance in calibration period (Table 3) even though its intervals are second widest among the cases. It suggests that when all input uncertainties are jointly considered the uncertainty estimates not only widen their bounds but also improve their accuracy.

The averaged PI width ratios when including input uncertainty are larger in the forecast period than the calibration period (Fig. 7b). This difference might originate from the discharge data

(Fig. 2a), which indicate that three or four large flood events occurred every year during the forecast period (2006 to 2009) whereas the peak flow appeared only five times during the calibration period (2001 to 2005).

Table 3 Average of root mean squared error values for 2,000 simulations for the calibration and forecast periods using parameter sets sampled after convergence. The lowest values in each period are shown in bold face.

Case	Calibration	Forecast
0	112,931	140,036
1	112,326	140,595
2	112,070	142,722
3	113,907	137,013
4	112,260	141,385
5	108,426	139,375
6	113,384	144,024
7	106,293	146,934

2.5 Conclusions

This research proposes a way to address the uncertainty due to hydraulic input variables as well as model parameters in order to improve uncertainty estimates for sediment transport model predictions. The proposed error models allow the forcing variables in a model simulation to vary by applying error values generated from a Gaussian distribution. The mean and standard deviation of the Gaussian distribution are then treated as additional uncertain parameters in the Bayesian uncertainty evaluation process. Based on the tests performed in this study, we can draw the following conclusions:

1. Considering input uncertainty requires more model simulations to estimate uncertainty.

The increase in the computational cost is more notable when the error model deals with

an input variable containing a larger number of uncertain elements. In such a case, the SCEM-UA algorithm used here for the Bayesian analysis converges slower. Compared to the 3,450 simulations required for only parameter uncertainty, the algorithm ran about 5,000~8,000 simulations when the error models are applied to the variables that only affect the flow conditions. In contrast, about 17,000 simulations were required for using the error model that changes the channel geometry in every simulation. In addition, the joint posterior pdf could be approximated after 35,000 simulations when the interactions between input variables are included.

2. For the cases considered, considering input uncertainty leads to only small changes in the uncertainty estimates for the SRH-1D model parameters. Specifically, the IQR of the posterior pdf for Manning's n was reduced 3 % by considering the uncertainties in all the input variables considered in this study. Only a slight change occurs because well-calibrated parameters such as Manning's n , reference shear stress, and the hiding and exposure coefficient usually depend on the composition of bed material sizes, but the suggested error models were applied to the forcing variables related to only flow rate or channel geometry.
3. Estimated uncertainty for a single input variable can vary if the associated error parameters (m and σ) are calibrated jointly with the error parameters for other input variables. From the cases considered, the posterior pdfs for the input error parameters indicated that the discharge data and the two rating curves do not contain bias when each variable was considered separately. When the joint pdfs were developed including the error parameters for all uncertain input variables, the IQRs for some input error

parameters were reduced by half, which implies a significant decrease in the estimated uncertainty in the associated input variables.

4. The PIs produced by considering input uncertainty are more accurate than the PIs when only considering parameter uncertainty. The most notable improvement was observed when the predictions are generated by using all input error models suggested here for a single calibration process. The simulations from this approach show better performance in calibration period as they have 6 % less averaged RMSE compared to the existing method, and their PIs also cover 1.5 and 2 times more observations in the calibration and forecast periods, respectively.

The research described in this study should be expanded to several avenues for future work.

First, the proposed method can be applied to river network cases, where several reaches are linked and interact or to reservoir sedimentation cases. One could also consider other forcing variables like bed material size composition, which might have meaningful impacts on the model parameters including Manning's n , reference shear stress, and hiding and exposure coefficient. Second, the model parameters and input data are not the only source of uncertainty in sediment transport modeling. The uncertainty might originate from the selection of the transport equation, which is part of the model's mathematical structure, and the observations used for calibration (Ruark et al. 2011). Third, the Bayesian method requires too many model simulations. A single simulation of the Tachia River took only 20 seconds to run, but 7 days of continuous computation time were required to achieve convergence for the case that considered all uncertain input variables. The computational cost would increase tremendously when the proposed method is applied to more complex and higher dimensional modeling cases unless high

performance computing resources are available. Therefore, methods should be developed to improve the efficiency.

References

- Ajami NK, Duan Q, Sorooshian S (2007) An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resources Research* 43:W01403. doi:10.1029/2005WR004745
- Bertin X, Fortunato AB, Oliveira A (2007) Sensitivity analysis of a morphodynamic modeling system applied to a Portuguese tidal inlet. *Proceedings of 5th IAHR Symposium on River, Coastal, and Estuarine Morphodynamics*: 11-17
- Box GEP, Tiao GC (1973) *Bayesian Inference in Statistical Analyses*. Addison-Wesley-Longman, Reading MA
- Bunte K, Abt SR (2005) Effect of sampling time on measured gravel bed load transport rates in a coarse-bedded stream. *Water Resources Research* 41:W11405. doi:10.1029/2004WR003880
- Camenen B, Larroudé P (2003) Comparison of sediment transport formulae for the coastal environment. *Coastal Engineering* 48(2):111-132
- Carpenter TM, Georgakakos KP (2006) Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales. *Journal of Hydrology* 329(1): 174-185
- Chang CH, Yang JC, Tung YK (1993) Sensitivity and uncertainty analysis of a sediment transport model: A global approach. *Stochastic Hydrology and Hydraulics* 7(4): 299–314
- Cho E, Arhonditsis GB, Khim J, Chung S, Heo T (2016) Modeling metal-sediment interaction processes: Parameter sensitivity assessment and uncertainty analysis. *Environmental Modeling & Software* 80:159-174
- Christensen R, Johnson W, Branscum A, Hanson TE (2011) *Bayesian ideas and data analysis: An introduction for scientists and statisticians*. CRC Press, Boca Raton
- Clément P, Piégay H (2003) Statistics and fluvial geomorphology. In: Kondolf GM, Piégay H (ed) *Tools in Fluvial Geomorphology*. John Wiley & Sons, Chichester UK, pp 597-630
- Cui Y, Parker G (1998) The arrested gravel front: stable gravel-sand transitions in rivers Part 2: General numerical solution. *Journal of Hydraulic Research* 36(2): 159-182
- Davies AG, van Rijn LC, Damgaard JS, van de Graaff J, Ribberink JS (2002) Intercomparison of research and practical sand transport models. *Coastal Engineering* 46(1): 1-23

- Frings RM, Schüttrumpf H, Vollmer S (2011) Verification of porosity predictors for fluvial sand-gravel deposits. *Water Resources Research* 47:W07525. doi:10.1029/2010WR009690
- Gaeuman D, Andrews ED, Krause A, Smith W (2009) Predicting fractional bed load transport rates: Application of the Wilcock-Crowe equations to a regulated gravel bed river. *Water Resources Research* 45:W06409. doi:10.1029/2008WR007320
- Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Statistical Science* 7:457-472
- Green PJ (2001) A primer on Markov chain Monte Carlo. In: Barndorff-Nielsen OE, Cox DR, Klüppelberg C (ed) *Complex Stochastic Systems*. Chapman & Hall/CRC, Boca Raton, pp 1-62
- Greimann BP, Lai Y, Huang JV (2008) Two-dimensional total sediment load model equations. *Journal of Hydraulic Engineering* 134(8):1142-1146
- Huang JV, Greimann BP (2013) User's manual for SRH-1D 3.0. Bureau of Reclamation, U.S. Department of the Interior, Denver
- Huard D, Mailhot A (2006) A Bayesian perspective on input uncertainty in model calibration: Application to hydrological model "abc." *Water Resources Research* 42:W07416. doi:10.1029/2005WR004661
- Kanso A, Chebbo G, Tassin B (2005) Bayesian analysis for erosion modelling of sediments in combined sewer systems. *Water Science and Technology* 52(5):135-142
- Kavetski D, Franks SW, Kuczera G (2002) Confronting input uncertainty in environmental modelling. In: Duan Q, Gupta HV, Sorooshian S, Rousseau AN, Turcotte R (ed) *Calibration of Watershed Models*. American Geophysical Union, Washington DC. doi: 10.1029/WS006p0049
- Lai Y, Greimann BP (2010) SRH model applications and progress report on bank erosion and turbidity current models. Bureau of Reclamation, U.S. Department of the Interior, Denver
- Laloy E, Vrugt JA (2012) High-dimensional posterior exploration of hydrologic models using multiple-try DREAM (ZS) and high-performance computing. *Water Resources Research* 48:W01526. doi:10.1029/2011WR010608
- Limneros JT (1970) Determination of the Manning coefficient from measured bed roughness in natural channels. Geological Survey, U.S. Department of the Interior, Washington DC
- McLean DG (1985) Sensitivity analysis of bedload equations. *Proceedings of Canadian Society for Civil Engineering Annual Conference 1985*:1-15
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. *Journal of Chemical Physics* 21(6):1087-1092

- Mueller ER, Pitlick J, Nelson JM (2005) Variation in the reference Shields stress for bed load transport in gravel bed streams and rivers. *Water Resources Research* 41:W04006. doi:10.1029/2004WR003692
- Nelson WB (1982) *Applied Life Data Analysis*. John Wiley & Sons, New York
- Owen-Joyce SJ, Raymond LH (1996) An accounting system for water and consumptive use along the Colorado River, Hoover Dam to Mexico (No. 2407). Geological Survey, U.S. Department of the Interior, Denver
- Pinto L, Fortunato AB, Freire P (2006) Sensitivity analysis of non-cohesive sediment transport formulae. *Continental Shelf Research* 26(15):1826-1839
- Ruark MD, Niemann JD, Greimann BP, Arabi M (2011) Method for assessing impacts of parameter uncertainty in sediment transport modeling applications. *Journal of Hydraulic Engineering* 137(6). doi:10.1061/(ASCE)HY.1943-7900.0000343.
- Sabatine SM, Niemann JD, Greimann BP (2015) Evaluation of parameter and model uncertainty in simple applications of a 1D sediment transport model. *Journal of Hydraulic Engineering* 141(5). doi:10.1061/(ASCE)HY.1943-7900.0000992.
- Sauer VB, Meyer RW (1992) Determination of error in individual discharge measurements. Geological Survey, U.S. Department of the Interior, Denver
- Schmelter ML, Erwin SO, Wilcock PR (2012) Accounting for uncertainty in cumulative sediment transport using Bayesian statistics. *Geomorphology* 175:1-13
- Schmelter ML, Hooten MB, Stevens DK (2011) Bayesian sediment transport model for unisize bed load. *Water Resources Research* 47:W11514. doi:10.1029/2011WR010754
- Schmelter ML, Stevens DK (2013) Traditional and Bayesian statistical models in fluvial sediment transport. *Journal of Hydraulic Engineering* 139(3). doi:10.1061/(ASCE)HY.1943-7900.0000672
- Schmelter ML, Wilcock PR, Hooten MB, Stevens DK (2015) Multi-fraction Bayesian sediment transport model. *Journal of Marine Science and Engineering* 3(3):1066-1092
- van Griensven A, Meixner T (2007) A global and efficient multi-objective auto-calibration and uncertainty estimation method for water quality catchment models. *Journal of Hydroinformatics* 9(4):277-291
- Vrugt JA, Gupta HV, Bouten W, Sorooshian S (2003) A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research* 39(8):1201. doi:10.1029/2002WR001642
- Vrugt JA, Ter Braak CJ, Clark MP, Hyman JM, Robinson BA (2008) Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research* 44:W00B09. doi:10.1029/2007WR006720

- Wilcock PR (2001) Toward a practical method for estimating sediment-transport rates in gravel-bed rivers. *Earth Surface Processes and Landforms* 26(13): 1395-1408
- Wu FC, Chen CC (2009) Bayesian updating of parameters for a sediment entrainment model via Markov chain Monte Carlo. *Journal of Hydraulic Engineering* 135(1):22-37
- Wu W, Wang SS, Jia Y (2000) Nonuniform sediment transport in alluvial rivers. *Journal of Hydraulic Research* 38(6):427-434
- Ye M, Neuman SP, Meyer PD (2004) Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff. *Water Resources Research* 40:W05113. doi:10.1029/2003WR002557.

CHAPTER 3

REDUCED COMPUTATIONAL COST FOR ESTIMATING PREDICTION UNCERTAINTY DUE TO SEDIMENT TRANSPORT MODEL PARAMETERS

Abstract

Bayesian methods have recently been applied to sediment transport models to assess the uncertainty in the model predictions due to uncertainty in the parameter values. However, these approaches require too many model runs, so they are not feasible for models of complex fluvial systems, which might take significant computation time for each simulation. This research suggests a method to estimate the sediment transport model parameters and quantify the uncertainty in the model predictions with reduced numbers of model simulations. To accomplish this goal, a new algorithm is developed to enable the Generalized Likelihood Uncertainty Estimation (GLUE) method to implement conditional sampling. In addition, the algorithm is also intended to identify the correlation between parameters. To test the method, four case studies including three numerical experiments and a real river simulation are conducted. The results indicate that the new approach significantly enhances the efficiency of uncertainty analysis as it requires only about 15~25% of the simulations required by existing Bayesian methods while still providing similar estimates for both parameter values and model predictions.

3.1 Introduction

Numerical hydraulic and sediment transport models have been widely used to predict river morphological changes that result from natural and/or human influences. The results from such models are often used for decision making in water resources management (e.g., dredging, river restoration, and other environmental remediation plans). The predictions from those models

always possess uncertainty. One major source of uncertainty is the model parameter values that need to be determined by the modeler. Sediment transport models contain various physical and/or conceptual (numerical) model parameters that are either difficult or impossible to directly measure. Such parameters are calibrated by adjusting their values so that the model results successfully reproduce the observed response from the fluvial system, but no single parameter value is expected to perfectly represent the system behavior. Thus, it is essential to understand the uncertainty in the parameter values and account for the associated uncertainty in the predictions when identifying water management strategies based on the model results.

Bayesian inference offers a formal way to assess the uncertainties in the model parameter values by comparing to calibration data and to quantify how those uncertainties affect the model predictions. In the past 10 years, Bayesian methods have been applied to sediment transport modeling cases that predict erosion within sewer systems (Kanso et al. 2005), sediment entrainment in a gravel-bed flume (Wu and Chen 2009), bed elevation and material changes in flume experiments (Ruark et al. 2011; Sabatine et al. 2015), cumulative sediment load in a fluvial channel (Schmelter et al. 2011, 2012, and 2015; Schmelter and Stevens 2013), and cohesive sediment behavior (Cho et al. 2016). The Bayesian paradigm derives the posterior pdf by evaluating the ability of parameter sets that are sampled from the prior pdf to reproduce available calibration data. It is impossible to derive the large dimensional posterior pdf analytically because the numerical simulation models are non-linear and often complex. In practice, the posterior distributions are approximated by integrating the likelihoods or posterior density values, which are acquired from a large number of model simulations, over the parameter space. Such uncertainty analysis can be implemented using Generalized Likelihood Uncertainty

Estimation (GLUE) method (Beven and Binley 1992) or Markov Chain Monte Carlo (MCMC) algorithm-based methods.

While the traditional GLUE method (Beven and Binley 1992) and its various transformations (Blasone et al. 2008; Stedinger et al. 2008) have been widely used in hydrologic applications due to its algorithmic simplicity and flexibility, the large number of model simulations required for reliable estimates is the key limitation for sediment transport model cases. Specifically, studies using distributed watershed models use 50,000 ~ 100,000 model evaluations depending on the number of uncertain parameters considered (Beven and Freer 2001; Blazkova et al. 2002; Jia and Culver 2006). However, such simulation numbers would not be feasible for sediment transport models where each simulation can be time consuming. For example, a three dimensional sediment transport model took 30 minutes for a steady simulation of flow and sediment transport near an intake facility where the area was about 7,000 m² (1.73 acre) using 9,000 coarse cells (Ruether et al. 2005). Although this case is simple, 10,000 samples would require a continuous computation time of 200 days, which would not be feasible in practice. Moreover, both the number of simulations and the time for each simulation would increase substantially when assessing the uncertainty in the predictions for more complex and higher dimensional modeling cases. Such expensive computational costs originate from inefficient sampling method used in GLUE. Specifically, GLUE typically generates a large population of independent parameter sets using simple random sampling within specified parameter ranges, which is known as Monte Carlo (MC) sampling. As a result, many simulations are performed using parameter sets that have low likelihoods of being correct (van Griensven and Meixner 2007; Blasone et al. 2008). Some strategies are available to solve the inefficiency in GLUE, but the limitations still exist in each of those strategies as discussed following. One of the easiest ways to reduce the massive

computation time is to parallelize the simulations using a parallel computing processor (or network). Continuous computation time is expected to decrease as the number of available processors increases, but it does not mean that the algorithm efficiency is improved. Second, sensitivity tests can be used to reduce the number of uncertain parameters prior to conducting the uncertainty analysis (Tolson and Shoemaker 2008). The reduced dimension of the parameter space would certainly require a smaller sample size to cover that space. However, the sensitivity test for a high-dimensional case can also be computationally expensive. For example, Zak and Beven (1999) used 60,000 model runs for the sensitivity test before conducting GLUE with another 60,000 simulations, and Athira and Sudheer (2015) used 28,000 simulations to reduce the number of considered parameters in SWAT model from 13 to 4. Third, Latin Hypercube sampling (LHS) can reduce the number of necessary model runs to cover the parameter space in GLUE compared to Monte Carlo sampling (Uhlenbrook and Sieber 2005). While its efficiency in sampling within a multi-dimensional space has been widely addressed (Jones and Johnson 2009; Loepky et al. 2009), the question of whether LHS is able to produce enough samples to represent the high likelihood region in detail still remains.

Another criticism of GLUE is that the method does not provide an estimate of the joint posterior pdf because it is not able to identify correlation between the parameters. Specifically, GLUE produces the cumulative marginal posterior distribution for each parameter using the likelihood values from the calibration period, and the parameter sets generated from those marginal distributions are used for the forecast period to determine the prediction uncertainty. Thus, the forecast uncertainty estimated by this method does not include any impact of correlation between the model parameters. However, multiple parameters in sediment transport models are usually highly correlated. In addition, Wu and Chen (2009) and Sabatine et al. (2015) demonstrated that

the estimated uncertainty in the predictions is substantially different depending on whether the correlations between the sediment transport model parameters are included or not.

In context of MCMC methods, the parameter sets are generated and used for simulations one-by-one to iteratively obtain a numerical approximation for the posterior pdf (Green 2001). Once an MCMC algorithm has performed sufficient model runs, the parameter sets are sampled from a stationary distribution, which implies that the algorithm has converged. The sample generated after convergence is then used to estimate the joint posterior pdf. In order to achieve convergence with fewer computations, various MCMC algorithms have been suggested such as adaptive Metropolis (Haario et al. 2001), delayed rejection adaptive Metropolis (Haario et al. 2006), Shuffled Complex Evolution (Duan et al. 1992), Shuffled Complex Evolution Metropolis - Uncertainty Analysis (SCEM-UA) (Vrugt et al. 2003), Differential Evolution-Markov Chain (Ter Braak 2006), and the family of Differential Evolution Adaptive Metropolis (DREAM) methods (Vrugt et al. 2008) including $DREAM_{(D)}$ (Vrugt and Ter Braak 2011), $DREAM_{(ZS)}$ (Laloy and Vrugt 2012), and $DREAM_{(ABC)}$ (Sadegh and Vrugt 2014). Many of those methods have been widely used for hydrologic model applications because their sampling processes are more efficient (Vrugt et al. 2003) and they can infer correlations between model parameters, which can significantly impact the uncertainty estimates (Sabatine et al. 2015).

Although the MCMC algorithms provide the results based on a rigorous statistical foundation and overcome the weaknesses in GLUE, they also possess distinct limitations. First, using the MCMC methods is still computationally expensive because they require many simulations not only for the algorithm convergence but also for collecting the posterior sample after convergence. Specifically, Vrugt et al. (2009) used at least 40,000 simulations of the conceptual watershed model to achieve convergence, and Ajami et al. (2007) generated 20,000 parameter sets after

convergence to provide well-specified histogram. In addition, Sabatine et al. (2015) found that an MCMC method did not substantially improve the computational cost compared to GLUE for sediment transport model simulations of flume experiments. Unfortunately, parallel computing is not feasible because the MCMC sampling is a sequential and dependent process such that the updated parameter set from one step becomes the current point for the next step (Foglia et al. 2009). Second, stringent assumptions and complex structure used in those algorithms might cause difficulty in conducting uncertainty analysis (Tolson and Shoemaker 2008) for the modeler who is not an expert on statistics. In addition, a number of algorithmic parameters that might affect the results need to be determined by the modeler to use those methods.

The main goal of this study is to develop and test an efficient way to assess the uncertainties in sediment transport model parameters and evaluate their contributions to the uncertainty in the model predictions. A new algorithm is built on modifications of GLUE to use its merit of simplicity. In order to reduce the computational expense, the algorithm is designed to consecutively generate a sample of parameter sets by considering the information about high posterior density regions identified during the analysis. The posterior sample collected by the new method also enables forecasts to include the correlations between the model parameters. To test the method, three published numerical experiments using synthetic posterior distributions (Vrugt et al. 2003; Vrugt et al. 2009) and a natural river (Lai and Greimann 2010) are considered as case studies. For the real river case, nine sediment transport parameters included in the Sedimentation and River Hydraulics – One Dimension (SRH-1D) (Huang and Greimann 2013) model are treated as uncertain parameters to simulate the net deposition volumes along the 23-km reach of the Tachia River in Taiwan. The proposed method is evaluated by comparing to both GLUE and an MCMC method based on (1) the number of simulations required for

estimating both parameter and prediction uncertainties, (2) the estimated posterior distributions for uncertain parameters, and (3) the performance of the model forecasts.

3.2 Methodology for Parameter Uncertainty Assessment

3.2.1 Existing GLUE and MCMC Methods

Before introducing the new algorithm, brief summaries of the traditional GLUE (Beven and Binley 1992) and SCEM-UA (Vrugt et al. 2003) methods are provided. These are used as the reference methods to evaluate results from the new algorithm in this study.

In typical GLUE applications, model parameters are initially assumed to conform to uniform distributions within specified ranges under the assumption that no prior information is available about the parameter values aside from their feasible limits. A population of parameter sets is then generated using MC sampling and used in the model to simulate the calibration period. The likelihood of each parameter set is computed, most commonly using Nash-Sutcliffe Coefficient of Efficiency (Nash and Sutcliffe 1970), and the behavioral parameter sets that produce reasonable predictions of the calibration data are determined. Then, the likelihood value of each of the behavioral parameter sets is divided by the sum of their likelihoods. For each parameter, the resulting likelihood values are accumulated along the parameter values, and the resulting distribution can be treated as an estimate for a cumulative marginal posterior distribution for that parameter. Parameter sets generated from these posterior distributions are then used to simulate the forecast period, and the associated distributions of the model outputs are used to assess the forecast uncertainty.

SCEM-UA is one of the most efficient MCMC methods and has been used for various hydrologic applications (Laloy and Vrugt 2012), and it also has a relatively simple algorithmic

structure compared to the other advanced MCMC methods. Readers are referred to Vrugt et al. (2003) for a detailed discussion. SCEM-UA starts by generating an initial sample of parameter sets from a uniform prior joint pdf, and the parameter sets are used in model simulations of the calibration period. After specifying the posterior density of each parameter set, the sample is partitioned into a number of complexes. The parameter sets in the complexes are updated in parallel using the Metropolis algorithm (Metropolis et al. 1953) by treating the highest posterior density parameter set in each complex as the starting point of an updating sequence. For each sequence, a new parameter set is generated from a multivariate normal distribution, which is centered on the current parameter set of the sequence with a covariance structure that is inferred from the parameters in the complex, and the parameter set is used in the model to evaluate its posterior density. If the posterior density of the new parameter set is larger than the current parameter set, the current one is replaced by the new one. Otherwise, the new parameter set can replace the current parameter set with a randomly specified probability. Such sampling behavior that can accept a trial with lower posterior density helps to escape from locally optimal areas. After a few iterations of the updating procedure, the members of all complexes are recombined, shuffled, and re-divided into complexes in order to use the most likely parameter sets as the starting point for the updating sequences more frequently. While such a process can improve the changes of sampling the high posterior density region, it violates the detailed balance principle of MCMC algorithm that the update should always be performed from their last updated point in a sequence (Vrugt et al. 2008; Laloy and Vrugt 2012).

As the iterative procedure continues, the parameters are being calibrated because the parameter values with higher posterior density are sampled more frequently, and the convergence of SCEM-UA can be determined using the Scale Reduction Score (SRS) (Gelman and Rubin 1992),

which is the ratio of the variance of the average parameter values from each complex to the average of the variances of parameter values within each complex. When the SRS for all parameters is below 1.2, the samples are generated from an adequately stationary distribution, and the joint posterior pdf of the uncertain parameters can be estimated using a large sample collected after that convergence point. In addition, the prediction uncertainty can be investigated by using those collected parameter sets for the forecast simulations.

3.2.2 Evolving Latin Hypercube Method

The new algorithm entitled the Evolving Latin Hypercube (ELH) method is suggested in this study by resolving the limitations of the GLUE method. The method is developed by focusing on the following strategies: (1) selecting an efficient sampling method to reduce the number of parameter sets that require model runs, (2) constructing the posterior sample by collecting parameter sets that include the correlations between the parameters, (3) estimating the posterior distributions every few simulations by considering the information about the high posterior density region obtained during the analysis, and (4) determining the point to stop the process to avoid unnecessary additional model simulations. The details about the ELH algorithm are below and illustrated in Fig. 8.

a. Generate Sample (Step 1)

The ELH algorithm starts by generating a prior sample of parameter sets from a uniform prior joint pdf under the same assumption used in GLUE (no prior information). In order to explore the large dimensional parameter space efficiently, LHS (McKay et al. 1979), which has been shown to more efficiency sample than the MC method from previous studies (van Griensven et al. 2006; Matala 2008), is used. For generating a sample size S given a joint uniform prior of the

P parameters $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_P]$, LHS first divides each parameter range into S non-overlapping equal sized intervals. For each parameter, only one value is selected randomly from each interval, and the S values of θ_1 are paired in a random manner with the values of θ_2 . These pairs are then associated similarly with the values of θ_3 and so on.

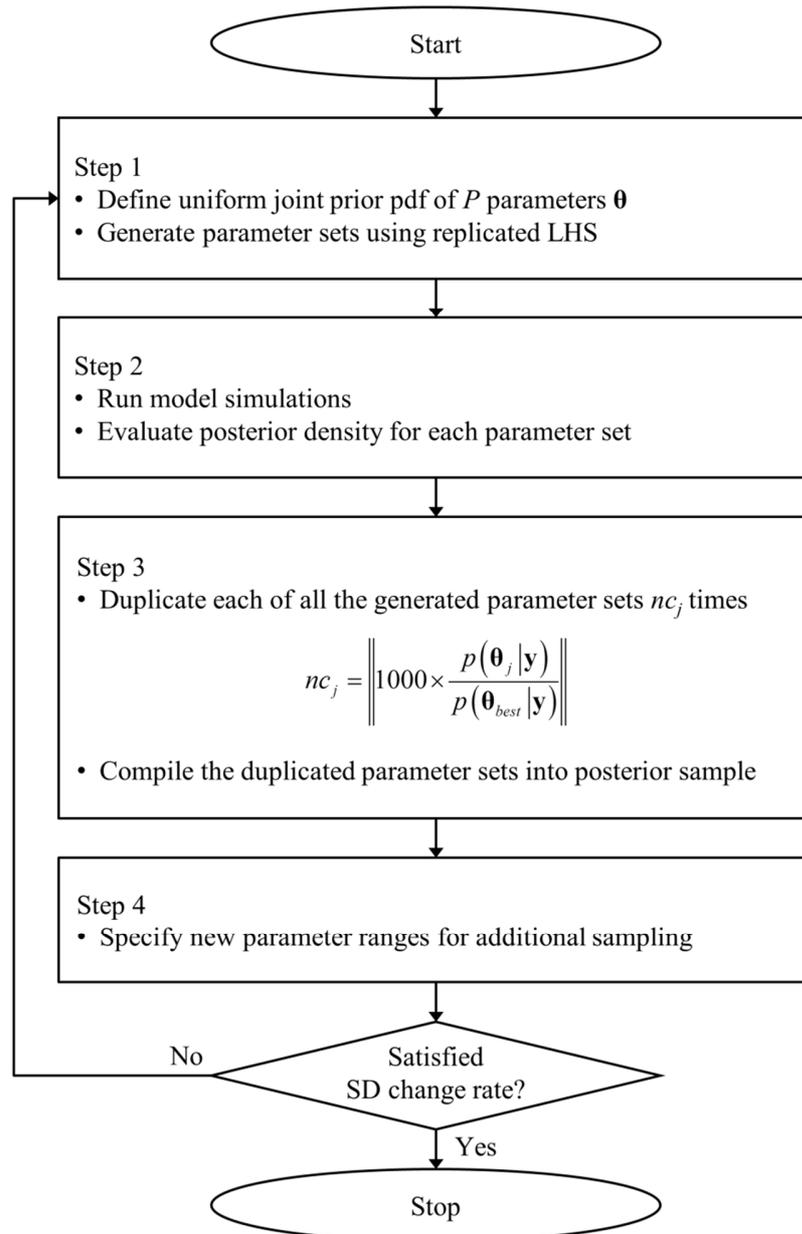


Fig. 8 Flowchart of the Evolving Latin Hypercube (ELH) algorithm.

Although no clear recommendation for the appropriate sample size S for LHS has been proposed in previous applications, Stein (1987) found that the estimated sample variances are stable with smaller sample size if several independent LHS samples are generated. The method is called replicated LHS. Specifically, Stein (1987) demonstrated that increasing the number of replications while each LHS sample size S is fixed increases the estimated sample variance, but the stable sample variance can be acquired as long as ratio S/P is large. This approach helps to reach stable properties of sampled parameter values as well as efficient parameter space filling properties with a smaller sample size. Preliminary tests indicate that the replicated LHS method is much better than original LHS, MC sampling, and stratified sampling methods. The recommendations from Stein (1987) that the number of replication should be at least 5 and the independent LHS sample size S should be larger than $10 * P$ work well for the cases considered here, so the replicated LHS sample size of $50 * P$ will be used for the ELH algorithm.

b. Run Models and Compute Posterior Densities (Step 2)

The parameter sets in the sample are used in model simulations of the calibration period, and the posterior density of each parameter set is assessed. The likelihood can be computed in a formal way by assuming that the model errors are mutually independent and Gaussian with constant variance. This formal function (Box and Tiao 1973) can be written as:

$$L(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma_{res}^2}} \exp\left[-\frac{(O_i(\boldsymbol{\theta}) - y_i)^2}{2\sigma_{res}^2}\right] \quad (16)$$

where $O_i(\boldsymbol{\theta})$ represents the simulated output variable when the parameter set $\boldsymbol{\theta}$ is used and y_i is the observed value at time or space i , and σ_{res}^2 is the variance of the model errors. Under the assumption of no prior parameter information, the prior pdf is:

$$p(\boldsymbol{\theta}) \propto \frac{1}{\sigma_{res}} \quad (17)$$

Box and Tiao (1973) demonstrated that the influence of σ_{res} can be integrated out, and suggested a simplified form to evaluate the posterior density of each parameter:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto \left[\sum_{i=1}^N (O_i(\boldsymbol{\theta}) - y_i)^2 \right]^{-\frac{1}{2}N} \quad (18)$$

This simplified posterior density equation has been used in various MCMC methods (Tiemann et al. 2001; Vrugt et al. 2003; Vrugt et al. 2008; Vrugt et al. 2009) because it can reflect the likelihood information of each parameter set in a formal way. Thus, this equation is used in this algorithm when evaluating the posterior density of a parameter set.

c. Estimate Posterior Distribution (Step 3)

Once the posterior densities of the parameter sets are identified, the posterior sample is constructed in order to estimate the joint posterior pdf. To do this, the parameter set $\boldsymbol{\theta}_{best}$ that has the highest posterior density among the prior sample is duplicated 1000 times and all of those copies are collected into the posterior sample. Next, the other parameter sets $\boldsymbol{\theta}_l$ are also duplicated and compiled into the posterior sample, but the number of copies for a parameter set nc_l is determined according to the ratio of its posterior density to the most likely parameter set's posterior density as:

$$nc_l = \left\| 1000 \times \frac{p(\boldsymbol{\theta}_l|\mathbf{y})}{p(\boldsymbol{\theta}_{best}|\mathbf{y})} \right\| \quad (19)$$

where $\lfloor \cdot \rfloor$ represents a rounding function. The duplication process is applied for all the parameter sets included in the prior sample. The parameters are expected to be calibrated because the parameter values with higher posterior densities will be more included more frequently in the posterior sample. As a result, the marginal posterior distribution for each parameter can be approximated using the histogram of the parameter values included in the posterior sample.

The posterior distributions from ELH are expected to be similar to the results from GLUE because both methodologies consider the ratio of each parameter set's likelihood to the other likelihoods when approximating the posterior distributions. Specifically, GLUE divides the likelihood value of each behavioral parameter set by the sum of the likelihoods of all behavioral sets to generate the cumulative posterior distributions, while ELH duplicates the parameter sets according the ratio of posterior densities. However, ELH can estimate the joint posterior pdf because the covariance structure can be identified using the sample of parameter sets compiled in the posterior sample.

The specification for nc_l can vary according to the modeler's decision. However, preliminary tests revealed that when the most likely parameter set is duplicated only 100 times the ELH method leads to a posterior distribution with lower variance than expected, which might be caused by discarding too many parameter sets in the low posterior density region (due to the rounding function). On the other hand, the results do not vary significantly when that parameter set is duplicated more than 1,000 times.

d. Specify New Parameter Ranges (Step 4)

The acquired posterior sample might not be able to provide an accurate distribution for high posterior density regions if the initially specified parameter ranges are too broad. Specifically,

the histogram from the posterior sample would not be able to describe the exact posterior distribution in detail because the prior sample ($50 \cdot P$ parameter sets) from the LHS method is supposed to cover the parameter space efficiently (coarsely). To overcome this limitation, ELH specifies new parameter ranges by excluding the outlier ranges that are found from the previous estimation and generates additional samples to improve the detail in the high posterior density region. For each parameter, the interquartile range (IQR), which is the difference between the 25% and 75% quantiles in the posterior sample, is computed, and the new bounds for the next sampling are determined as:

$$[25\% \text{ quantile point} - 2.5 \cdot \text{IQR} \quad 75\% \text{ quantile point} + 2.5 \cdot \text{IQR}] \quad (20)$$

The new bounds are not allowed to exceed the initial bounds. The reduction in the parameter ranges enables the next sampling to concentrate on the higher posterior density region.

e. Repeat Sampling and Estimating (Steps 1-4)

The process continues by generating $50 \cdot P$ new parameter sets from a uniform prior joint pdf with the updated parameter ranges (Step 1). The model simulations of calibration period are run using the new parameter sets, and the posterior densities are also computed using Eq. (18) (Step 2). Before moving to Step 3, the replications of the previous sample are removed. After that, the new sample is combined with the previous sample, and the duplicating process is applied to all the parameter sets generated so far (Step 3). Thus, the total number of parameter sets considered at Step 3 in the second estimation would be $2 \cdot 50 \cdot P$. The most likely parameter set θ_{best} can be either from the previous or the new sample. After updating the sample (with replications), the new parameter ranges are specified for the next sampling, and the iterative process continues until the results are stable.

f. Check Stability of Estimation

In order to check the stability of the estimation results, the standard deviation (SD) of the posterior sample (with replications) is calculated for each parameter after every updating loop. After the second loop, the relative SD change can be computed as:

$$\epsilon_{p,r} = \frac{|\text{SD}_{p,r} - \text{SD}_{p,r-1}|}{(\text{SD}_{p,r} + \text{SD}_{p,r-1})/2} \quad (21)$$

where $\text{SD}_{p,r}$ is the SD of the posterior sample for parameter p at the r th loop. A stationary posterior sample is then diagnosed if the relative SD change for all P parameters is less than 1 % during at least two updating loops. At that point, the available posterior sample is considered to represent the joint posterior pdf of the uncertain P parameters.

The SD is considered because it can quantify how the posterior distribution is spread, which can be interpreted as the uncertainty in the parameter values after calibration. In addition, using the relative change can allow more variation for parameters that are poorly calibrated whereas well-constrained parameters are diagnosed more strictly. Other statistical properties (e.g., IQR or coefficient of variation) were also considered and found to produce similar results. However, preliminary tests revealed that the relative SD change is able to provide more consistent diagnosis for stability than the other metrics.

g. Simulate Forecast Period

The parameter sets contained in the final posterior sample obtained from the simulations of the calibration period are then used for the forecast scenario. Because ELH duplicated the parameter sets, there are many identical parameter sets in the posterior sample. Simulations are only

required for the unique parameter sets, which greatly reduces the computational cost compared to the total number of parameter sets. The model results are then duplicated according to the number of identical parameter sets. The resulting distribution of the model outputs can describe the uncertainty in the predicted variables due to the uncertainty of the parameter values.

3.3 Case Studies for Synthetic Distributions

Numerical experiments using three synthetic target distributions with increasing complexity, which have been used by Vrugt et al. (2003 and 2009), are considered as case studies to evaluate the ELH method. The cases cover a diverse set of problem features including multimodality, correlation, and high-dimensionality in target probability distributions. These synthetic cases are well-controlled systems that do not allow uncertainty from sources other than the parameter value itself. In addition, the posterior density value can be computed directly from a probability density function (pdf) for given parameter values, instead of running a model, so that the computational costs for testing the method can be reduced. No forecast simulation is considered here, so the accuracy of estimated posterior distributions is examined through these case studies. SCEM-UA is considered as a comparison method to evaluate the performance of the ELH for these cases.

3.3.1 A Single Bimodal Distribution

To investigate the performance of ELH in presence of multimodality, the first case study involves a single parameter θ with a bimodal probability distribution (Case 1) where the pdf is written as:

$$p(\theta) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\theta^2\right] + \frac{2}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(2\theta-8)^2\right] \quad (22)$$

This target distribution is the sum of two Gaussian probability distributions where each has an optimal point at $\theta = 0$ and 4, respectively. In addition, the distribution has a mean of 2 and a variance of 4.6, and its 25% and 75% quantiles are located at $\theta = 0$ and 4, respectively. For uncertainty analysis, the prior range of the parameter was limited to $[-10 \ 10]$. ELH updated the posterior distribution by generating 50 parameter values in every loop, and SCEM-UA used four parallel updating sequences with initial 100 parameter values.

Fig. 9a presents the histograms of the posterior sample generated by ELH and SCEM-UA from 5,000 computations to estimate the target distribution of Eq. (22), which is shown as a black solid line. SCEM-UA converged after 200 computations so that the parameter values of the initial 200 draws from the non-converged sampling process are discarded when constructing the histogram. ELH provides the histogram that is not only smoother but also closer to the target than SCEM-UA. The accuracy of the histogram is measured by averaging the distances between the midpoints of each bar in the histogram and their corresponding point on the target curve. The resulting distances indicate that the ELH histogram (0.005) is more accurate than the SCEM-UA histogram (0.016).

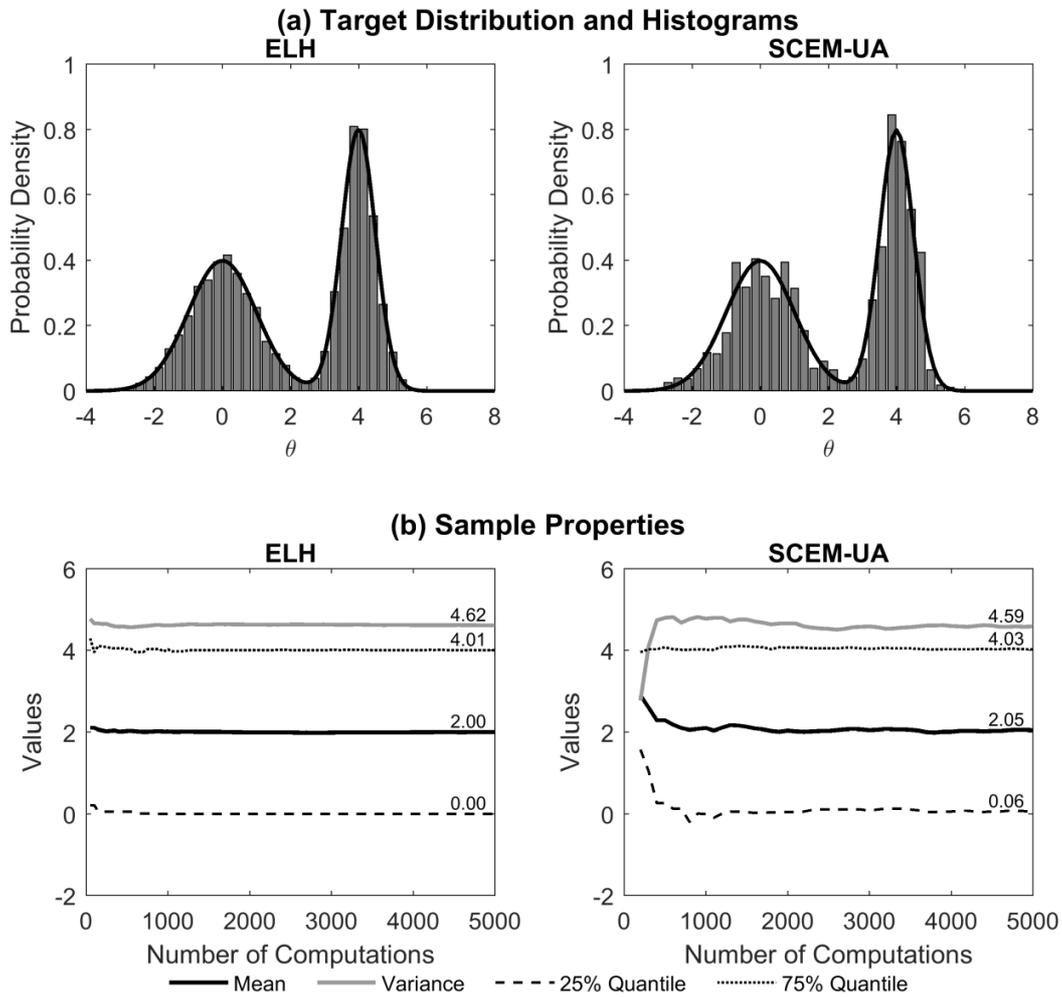


Fig. 9 (a) Bimodal target distribution and histogram of the generated 5,000 parameter values and (b) the changes in the posterior sample properties during 5,000 computations using ELH and SCEM-UA.

Fig. 9b compares the changes in the statistics of the posterior sample according to the number of computations during each analysis. Such properties should be stationary and match their target values if the algorithm approximates the target distribution well. The results from ELH become stationary much quicker than those from SCEM-UA. For example, the 25% and 75% quantiles from ELH change little and are fixed near the target values after about 100 computations. In addition, the final estimates at 5,000 computations (shown as a number above each line) also

show that ELH provides more accurate estimates for all the considered target values than SCEM-UA. Those results suggest that ELH is able to infer the single dimensional bimodal distribution defined in Eq. (22) with quicker and better performance than SCEM-UA.

Such difference might originate from the difference in estimating strategies between two methods. Specifically, ELH represents the posterior distribution over the considered parameter space every single updating loop by generating several parameter values at the same time and comparing their posterior densities, which allows exploration of the region with higher likelihood. On the other hand, SCEM-UA wanders the parameter space by comparing the posterior densities of the parameter values one-by-one every trial, which might require many comparisons to explore the entire space.

3.3.2 Multi-dimensional Gaussian Distributions

The second case study considers a two-dimensional Gaussian density function to evaluate ELH for application to two correlated parameters (Case 2). The target pdf for a set of two parameters $\theta = [\theta_1, \theta_2]$ can be written as:

$$p(\boldsymbol{\theta}) = \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} \exp\left[-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})\right] \quad (23)$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0.5\sigma_1\sigma_2 \\ 0.5\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} 1 & 0.71 \\ 0.71 & 2 \end{bmatrix} \quad (24)$$

where the distribution is centered on zero $\boldsymbol{\mu} = [0, 0]$ and the covariance matrix $\boldsymbol{\Sigma}$ includes the variances of p ($p = 1, 2$) for p th parameter with correlation coefficient of 0.5. The prior ranges for both parameters were limited to $[-10, 10]$, and ELH generated 100 combinations of two

parameter values every updating loop, while SCEM-UA used four parallel sequences with 100 initial parameter sets.

Fig. 10a compares the means, variances, and correlation coefficient of the posterior sample from each method during 10,000 computations (note that SCEM-UA collected the posterior sample after 1,000 computations, during which the algorithm converged). From the results, ELH reached the target values quicker and closer than SCEM-UA. In addition, the estimation accuracy is also calculated using an average normalized Euclidean distance D as:

$$D = \sqrt{\frac{1}{2} \sum_{p=1}^P \left(\left[\frac{\mu_p - \hat{\mu}_p}{\sigma_p} \right]^2 + \left[\frac{\sigma_p - \hat{\sigma}_p}{\sigma_p} \right]^2 \right)} \quad (25)$$

where μ_p and σ_p are the mean and standard deviations of target distribution and $\hat{\mu}_p$ and $\hat{\sigma}_p$ are the mean and standard deviations of the estimated distribution for parameter p , respectively.

According to the Euclidian distance D , ELH ($D = 0.003$) is more accurate than SCEM-UA ($D = 0.022$).

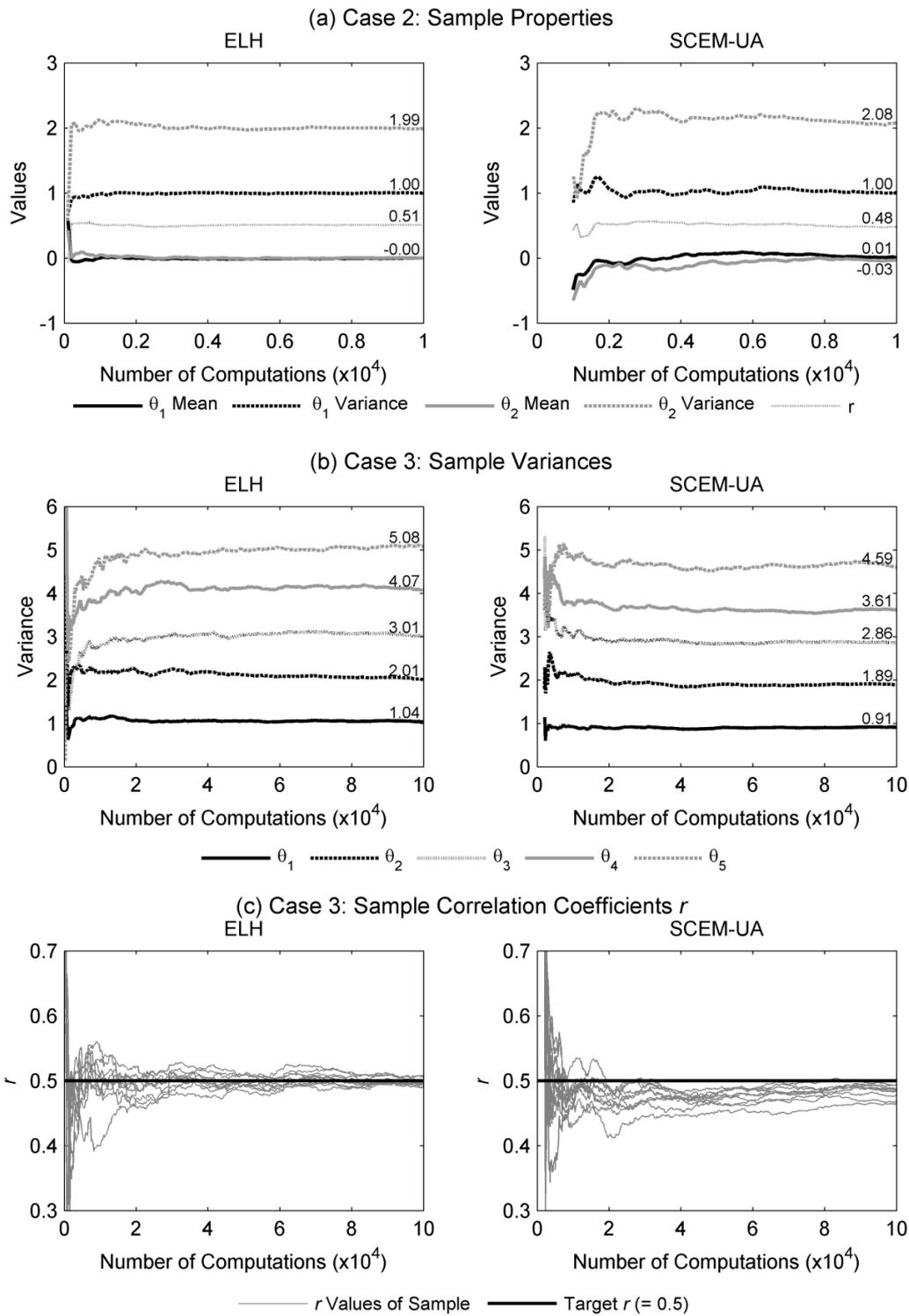


Fig. 10 (a) Posterior sample properties during 10,000 computations for Case 2, and (b) sample variances and (c) correlation coefficients estimated during 100,000 computations for Case 3 using ELH and SCEM-UA.

To explore the higher dimensional problem, the dimensions of the target Gaussian pdf of Eq. (23) is extended to five (Case 3). The covariance matrix can be written as follows:

$$\Sigma = \begin{bmatrix} 1 & 0.5\sigma_1\sigma_2 & 0.5\sigma_1\sigma_3 & 0.5\sigma_1\sigma_4 & 0.5\sigma_1\sigma_5 \\ 0.5\sigma_1\sigma_2 & 2 & 0.5\sigma_2\sigma_3 & 0.5\sigma_2\sigma_4 & 0.5\sigma_2\sigma_5 \\ 0.5\sigma_1\sigma_3 & 0.5\sigma_2\sigma_3 & 3 & 0.5\sigma_3\sigma_4 & 0.5\sigma_3\sigma_5 \\ 0.5\sigma_1\sigma_4 & 0.5\sigma_2\sigma_4 & 0.5\sigma_3\sigma_4 & 4 & 0.5\sigma_4\sigma_5 \\ 0.5\sigma_1\sigma_5 & 0.5\sigma_2\sigma_5 & 0.5\sigma_3\sigma_5 & 0.5\sigma_4\sigma_5 & 5 \end{bmatrix} \quad (26)$$

The means are well estimated (not shown), but notable differences between two algorithms are observed in the estimated variances (Fig. 10b). Specifically, SCEM-UA underestimates the variances for the 3rd, 4th, and 5th parameters, which have relatively high target variances, whereas ELH does not miss the target values as much. The Euclidian distance of ELH ($D = 0.027$) is also better than that of SCEM-UA ($D = 0.066$). In addition, the correlation coefficients are a little underestimated by SCEM-UA as all of them range from 0.45 to 0.49, which is lower than the target value of 0.5 (Fig. 10c).

The underestimated variance from SCEM-UA might be due to its sampling strategy, which violates the detailed balance principle. Specifically, the detailed balance principal forces continuity of the sampling process, but SCEM-UA disrupts the continuous sampling when the algorithm recombines, shuffles, and re-divides the parameter sets of all complexes in order to place the highest posterior density parameter sets as the starting point of the updating procedure more frequently. It might improve the efficiency to sample the parameter values more from the higher likelihood region, but the estimated posterior distribution can also be misrepresented. Unlike SCEM-UA, ELH removes only the outlier low likely regions when updating the parameter ranges. In addition, even a parameter set with a posterior density value that is lower

by a factor of 0.001 than most likely parameter set can be included in posterior sample with the impact exactly corresponding to its posterior density value according to Eq. (19).

3.4 Case Study for Sediment Transport Model Parameters

The last case study applies the ELH methodology to the Sedimentation and River Hydraulics - One Dimension (SRH-1D) (Huang and Greimann 2013) model for simulating a 23-km reach of the Tachia River in Taiwan (Lai and Greimann 2010). The posterior distributions for nine model parameters used in SRH-1D are estimated, and the credible ranges of the model predictions, which reflect the impact of the identified parameter uncertainty, are evaluated. In order to assess ELH comparatively, GLUE is implemented for this case in addition to SCEM-UA. The following section describes the simulation of the Tachia River. After that, the computational cost for uncertainty analysis, the posterior distributions of the model parameters, and the prediction accuracy are discussed. The detailed mathematics of the sediment transport model SRH-1D and its nine model parameters considered in this research have already been addressed in detail in the previous chapter (Section 2.3.1).

3.4.1 Application to Tachia River Simulation

Scour and sedimentation behavior along the 23-km reach of the Tachia River in Taiwan is considered. The selected reach stretches between Shih-Gang Dam and the ocean with bankfull widths of 300~1200 m and an average slope of 0.011. In 2007, the dominant substrate was cobbles and gravels, and the bed material sizes ranged from 0.125 mm to 512 mm (sand to boulders) with a median grain size (D50) of 108 mm. Net deposited/eroded sediment volumes were measured during two periods (2001 to 2005 and 2005 to 2009) at cross-sections along the reach from the dam to the ocean. During those periods, severe erosion occurred in part due to

the lack of sediment supply below the dam. The erosion occurred primarily from the dam to approximately 5 km downstream during 2001 to 2005 and to about 8 km downstream during 2005 to 2009. This case study uses the net deposition volume at each cross-section as the variable of interest for the uncertainty analysis and considers the first period (2001 to 2005) as the calibration period and the second (2005 to 2009) as the forecast period.

For the SRH-1D simulations, a time series of measured discharges at Shih-Gang Dam provides the upstream BC, but flow and sediment inputs from tributaries downstream of Shih-Gang Dam are not considered here due to lack of data. The downstream BC is specified using the critical depth obtained from the same HEC-RAS simulations and the 2005 geometry (Lai and Greimann 2010). Channel geometry data were collected in both 2001 and 2005, and these datasets provide the initial bed geometry for each period. Sediment size gradation data from 2007 were averaged along the channel and used as the initial bed material distribution for both the calibration and forecast periods. No levees, ineffective flow, or blocked obstructions are considered in the simulations.

All three uncertainty methods (ELH, SCEM-UA, and GLUE) use Eq. (18) to compute the posterior density of parameter set, which is the combination of the nine SRH-1D model parameters, when estimating the posterior distributions for those parameters. Then, the calibrated distributions are applied to the forecast period, and the uncertainty estimates are compared to the actual spread of the observations. The residuals for the net deposition volumes, which were calculated using the calibrated model, were examined to see whether they satisfy the assumptions of the posterior density function of Eq. (18). The residual lag plot shows a random pattern, so the residuals can be treated as independent, and the Kolmogorov-Smirnov test indicates the residuals are approximately Gaussian with constant variance.

3.4.2 Computational Cost for Uncertainty Estimates

In the ELH implementation, 500 parameter sets were generated for the simulations of the calibration period in every updating loop. Fig. 11a illustrates the SD change that is calculated every 500 simulations, and the stability of ELH is diagnosed at 6,000 simulations. The resulting posterior sample contains 2,378 unique parameter sets, and its total size is 116,981. For the forecast period, those 2,378 unique sets were used in the simulations, and the resulting predictions were also duplicated according to the number of corresponding identical sets.

SCEM-UA was conducted using 10 parallel updating sequences with 250 initial parameter sets following the suggestions from Vrugt et al. (2003) for the case where numerous uncertain parameters are expected to be highly correlated. Based on the SRS (Gelman and Rubin 1992), SCEM-UA required 8,000 model runs to reach algorithm convergence prior to collecting the posterior sample. The SD change is also applied to check the stability of the posterior sample from SCEM-UA by calculating it every 500 simulations after the convergence point, and the posterior sample is stationary at 13,000 model runs (Fig. 11b). As 13,000 parameter sets are acquired as part of the posterior sample, the forecast simulations use all of those parameter sets to obtain the sample of the predictions.

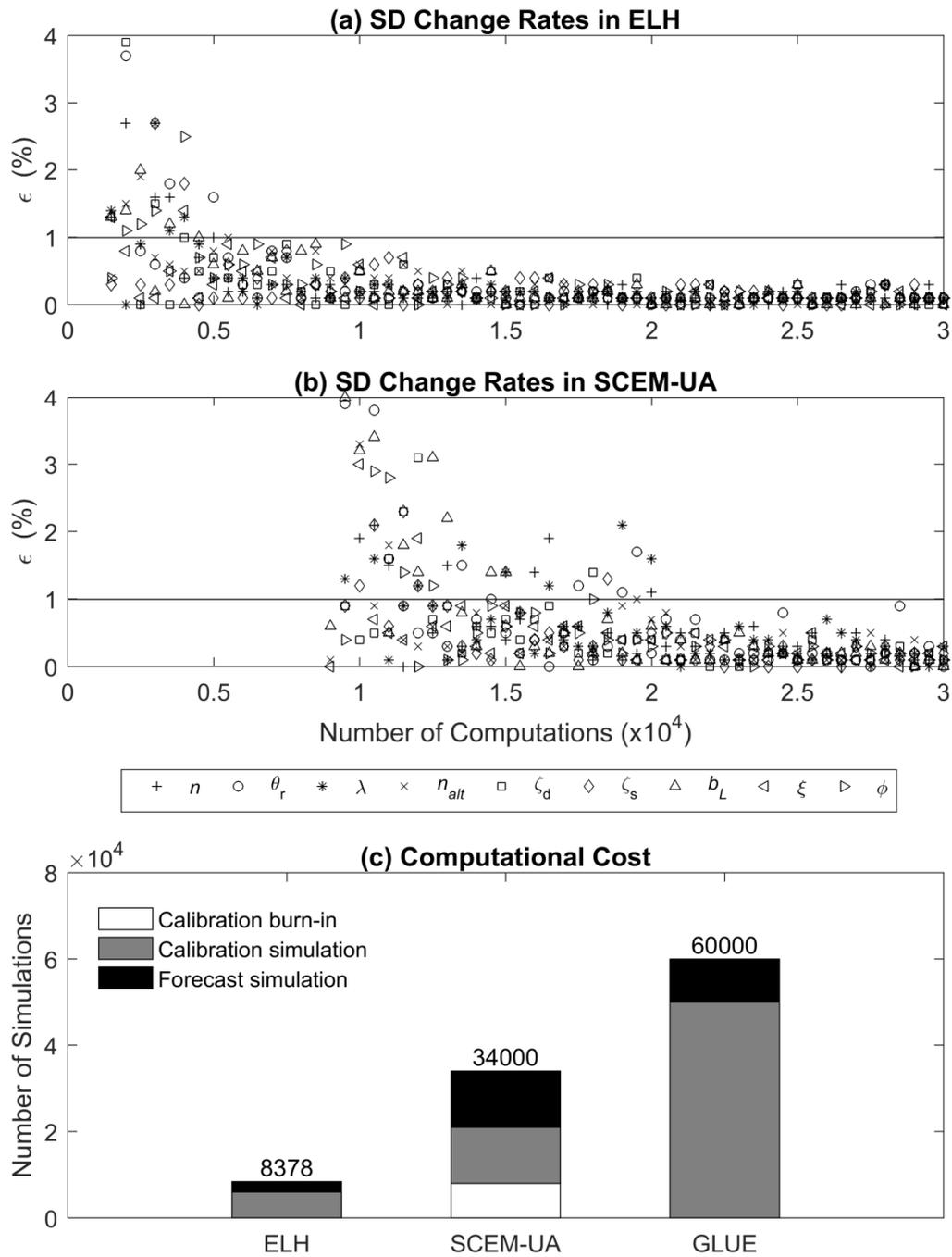


Fig. 11 SD changes for SRH-1D model parameters using (a) ELH and (b) SCEM-UA, and (c) the total number of model simulations required to estimate the prediction uncertainty using the three uncertainty methods for the Tachia River case.

Because the GLUE methodology is not a consecutive process, it was implemented several times by increasing the sample size from 5,000 up to 150,000 in order to find the size that produces stable estimation. The behavioral sets were determined if the posterior density of a parameter set is larger than 0.1 % of the most likely parameter set's posterior density, which is similar with the threshold used in ELH. The SD change cannot be applied because GLUE does not collect the posterior sample. By considering the median, 25% and 75% quantiles of the cumulative posterior distributions, GLUE provided nearly stable results when the simulation number is larger than 50,000. Using LHS, 10,000 parameter sets were generated from the marginal posterior distributions for each parameter and were used for the forecast simulations.

Fig. 11c compares the total number of simulations used for each method to implement the calibration and simulate the forecast scenario. Overall, the ELH method reduces the required simulations about 75% and 86% compared to the SCEM-UA and GLUE methods, respectively. For the calibration, ELH could estimate the joint posterior pdf before SCEM-UA achieved algorithm convergence. In addition, the sample duplication process in ELH enables reduced computational cost when simulating the forecast period.

A single simulation of the Tachia River took 20 and 28 seconds to run for calibration and forecast periods, respectively. In this study, four parallel computing processors of Intel i5-4690 CPU@ 3.50GHz 8GB RAM were used, and the continuous computation times are about 13 hours for ELH, 55 hours for SCEM-UA, and 90 hours for GLUE. The ratios of the computation times among the methods are almost the same as the ratio of the number of the required model runs. However, it is expected that high performance computing systems with larger number of multiple processors would greatly reduce the computation time for ELH. For example, ELH is expected to need only 0.05 hour for all simulations when 500 parallel processors are available

because it will be able to run 500 simulations at the same time for the 500 parameter sets that are independently generated from a single updating loop. On the other hand, such a reduction cannot be anticipated for SCEM-UA because its sequential sampling and updating processes strongly depend on the parameter sets updated from the previous step. Specifically, as 10 parallel updating sequences were used for this case, only 10 simulations can apply at the same time although much more processors are available.

3.4.3 Uncertainty in Model Parameters

The marginal posterior distribution for each parameter describes the uncertainty in the parameter value that remains after model calibration is complete. Fig. 12 compares the cumulative posterior distributions that are produced by the three uncertainty methods using the calibration data. In addition, Table 4 presents the median values and the ratio of the posterior IQRs to prior IQRs for the corresponding posterior distributions. The percentage values for IQR ratio can be interpreted as the fraction of the initial uncertainty that remains in the parameter after calibration.

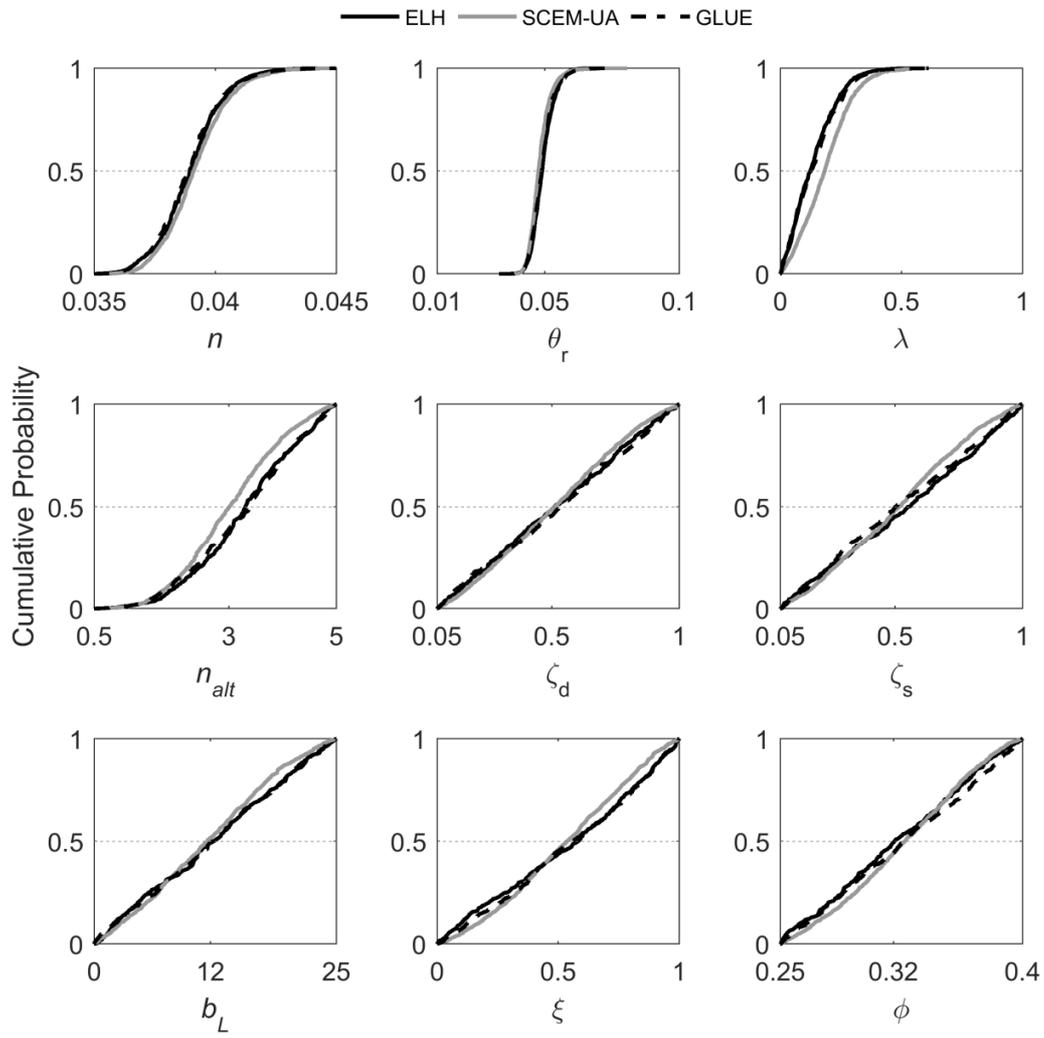


Fig. 12 Cumulative marginal posterior distributions for SRH-1D model parameters generated using the three uncertainty methods for the Tachia River case.

Table 4 Median values and IQR ratios of the estimated posterior distributions for SRH-1D model parameters obtained using the three uncertainty methods for the Tachia River case.

Parameters	n	θ_r	λ	n_{alt}	ζ_d	ζ_s	b_L	χ	ϕ
<i>Median</i>									
ELH	0.039	0.049	0.13	3.31	0.51	0.55	12.28	0.58	0.32
SCEM-UA	0.039	0.048	0.18	3.01	0.51	0.52	11.67	0.54	0.33
GLUE	0.039	0.049	0.13	3.39	0.53	0.50	12.26	0.56	0.33
<i>IQR Ratio</i>									
ELH	35%	11%	28%	66%	100%	104%	96%	99%	102%
SCEM-UA	35%	11%	30%	59%	85%	86%	81%	83%	80%
GLUE	33%	12%	28%	63%	96%	105%	101%	104%	96%

Overall, the reference shear stress is the best identified parameter as it has the steepest cumulative distribution (Fig. 12). Its estimated median values are around 0.048~0.049 (Table 4), which are acceptable for the bed material sizes found in the Tachia River (Lai and Greimann 2010). The hiding and exposure coefficient λ , Manning's roughness n , and active layer thickness multiplier n_{alt} are also reasonably constrained by the data. These parameters were found to have the highest impact on SRH-1D simulations of bed profile elevation in erosional flume experiments (Ruark et al. 2011), which are related to the net deposition volumes considered here. As an example, θ_r relates the flow velocity to the bed's overall susceptibility to erosion. Thus, it plays a large role in the evolution of the bed profile and the net deposition volumes, and the calibration data are expected to constrain their values. Their IQR ratio values (10~66 %) also indicate that less uncertainty remains after calibration (Table 4). The other five parameters (ζ_d , ζ_s , b_L , χ , and ϕ) show nearly linear posterior distributions (Fig. 12), which means they are poorly constrained by the calibration data and large uncertainty remains in their values. These five parameters have little impact on the SRH-1D model simulations of the Tachia River, so preferred

values cannot be identified. In addition, their median values are determined near the middle of their feasible ranges and have the IQR ratios higher than 80% (Table 4). Both ELH and GLUE produce IQR ratios for these parameters near 100%, whereas SCEM-UA provides lower IQR ratios for them. On the other hand, the four well-specified parameters are given similar IQR ratios from all three methods. This tendency is consistent with the results observed for Case 3 where SCEM-UA underestimated the target variances for the parameters with larger dispersion.

From both Fig. 12 and Table 4, notable differences between the uncertainty methods are found in the estimates for hiding and exposure coefficient λ and active layer thickness multiplier n_{alt} , which are well-calibrated numerical parameters. ELH and GLUE provide similar posterior distributions for the both parameters, but SCEM-UA produces different results. Specifically, the estimated median for hiding and exposure coefficient from both ELH and GLUE are centered on 0.13, which is lower than the value of 0.18 from SCEM-UA. However, those changes might be attenuated by slight increases in the estimated median for reference shear stress from 0.048 for SCEM-UA to 0.049 for ELH and GLUE. The reference shear stress controls the initiation of particle movement and its increase of only 0.001 would decrease the particle mobility sufficiently. Several studies have already reported its large impact on the SRH-1D model results in both flume experiments (Ruark et al. 2011) and real rivers (Lai and Greimann 2010). This interpretation can be supported by the fact that the reference shear stress was much better calibrated than the other parameters according to the cumulative posterior distributions and the IQR ratios obtained from this case.

Table 5 shows that the correlation coefficients between the four well-calibrated parameters. They are similar for ELH and SCEM-UA (note that GLUE neglects parameter correlations). The five poorly-calibrated parameters have negligible coefficient values that are all near zero (not

shown at the table). From Table 5, the reference shear stress can interact with three other parameters. Specifically, its strong relationship with hiding and exposure coefficient can be addressed by the mathematical structure of the Wu equation in Eqs. (12) and (13). In addition, its negative correlation with λ and the positive correlation with n_{alt} can support the interpretation for the different estimated medians in Table 4.

Table 5 Correlation coefficients between well-specified four model parameters of SRH-1D estimated by ELH and SCEM-UA in Tachia River case.

Parameters	ELH				SCEM-UA			
	n	θ_r	λ	n_{alt}	n	θ_r	λ	n_{alt}
n	1.00	0.28	-0.03	-0.13	1.00	0.31	0.02	-0.12
θ_r	0.28	1.00	-0.38	0.47	0.31	1.00	-0.35	0.45
λ	-0.03	-0.38	1.00	0.02	0.02	-0.35	1.00	0.00
n_{alt}	-0.13	0.47	0.02	1.00	-0.12	0.45	0.00	1.00

3.4.4 Model Prediction Uncertainty

The net sediment deposition volumes for the forecast period (2005 to 2009) were simulated using a sample of posterior parameter sets generated from each method. The forecast simulations use the cross-section data from 2005 as the initial geometry and the dam discharge data from 2005 to 2009 as the upstream BC, but the other conditions remain the same as the calibration period.

For each method in Fig. 13, the mean prediction is shown with a black bold line, and the 99% credible interval is shown by the grey region. It reflects the uncertainty due to the parameter values. Overall, no remarkable difference is observed between the predictions from ELH and SCEM-UA, but GLUE provides much wider ranges along the reach than the other two methods. Both ELH and GLUE consistently underestimate the eroded volumes in the upstream portion of

the reach (15-23 km) and underestimates deposition at the downstream end of the reach (0-4 km). Those predictions also possess similar uncertainty throughout the reach (i.e. the width of the prediction range remains relatively constant). On the other hand, the predictions from GLUE present much larger uncertainty for the eroded sediment volume, where the lower bound extends to the value that is more than 300% of the observed erosion volume at 16~21 km from the downstream end. However, the mean predictions from all three methods show similar results along the channel.

Fig. 14 illustrates the distribution of the predictions for net deposition volume at a cross section located 21 km upstream from the ocean. The numbers of predictions in each of the equal-sized bins was counted and those numbers were divided by the entire prediction sample size, which is 116,981 for ELH, 13,000 for SCEM-UA, and 10,000 for GLUE. ELH and SCEM-UA show similar distributions that have a single mode near $-2.2 \times 10^5 \text{ m}^3$ and range from -4×10^5 to $-0.5 \times 10^5 \text{ m}^3$. On the other hand, GLUE has a much wider prediction range and has modes near -2.2×10^5 and $-0.9 \times 10^5 \text{ m}^3$.

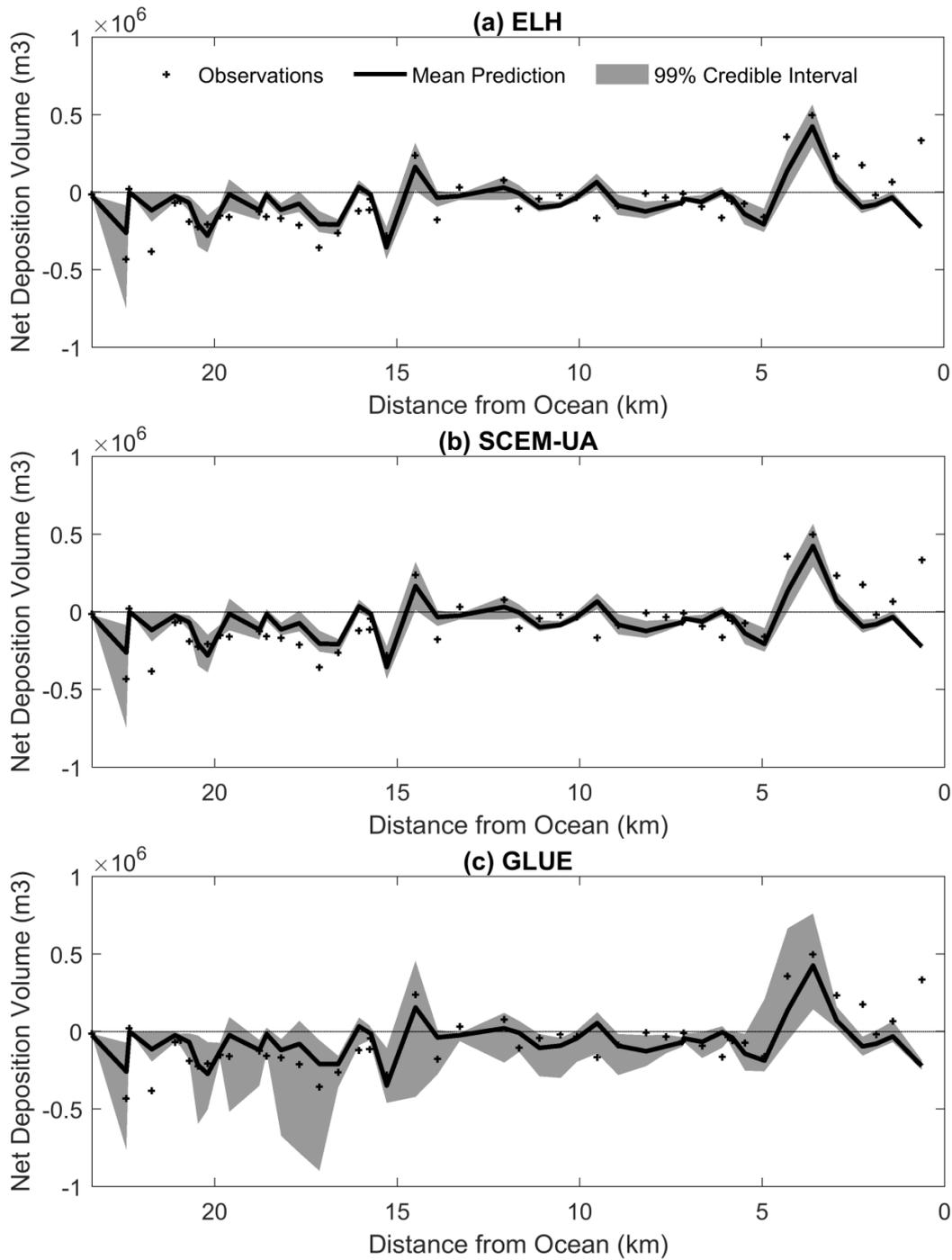


Fig. 13 Observations and mean predictions for sediment net deposition volume in the forecast period of the Tachia River case. The vertical width of the gray region is the 99% credible interval of the predictions.

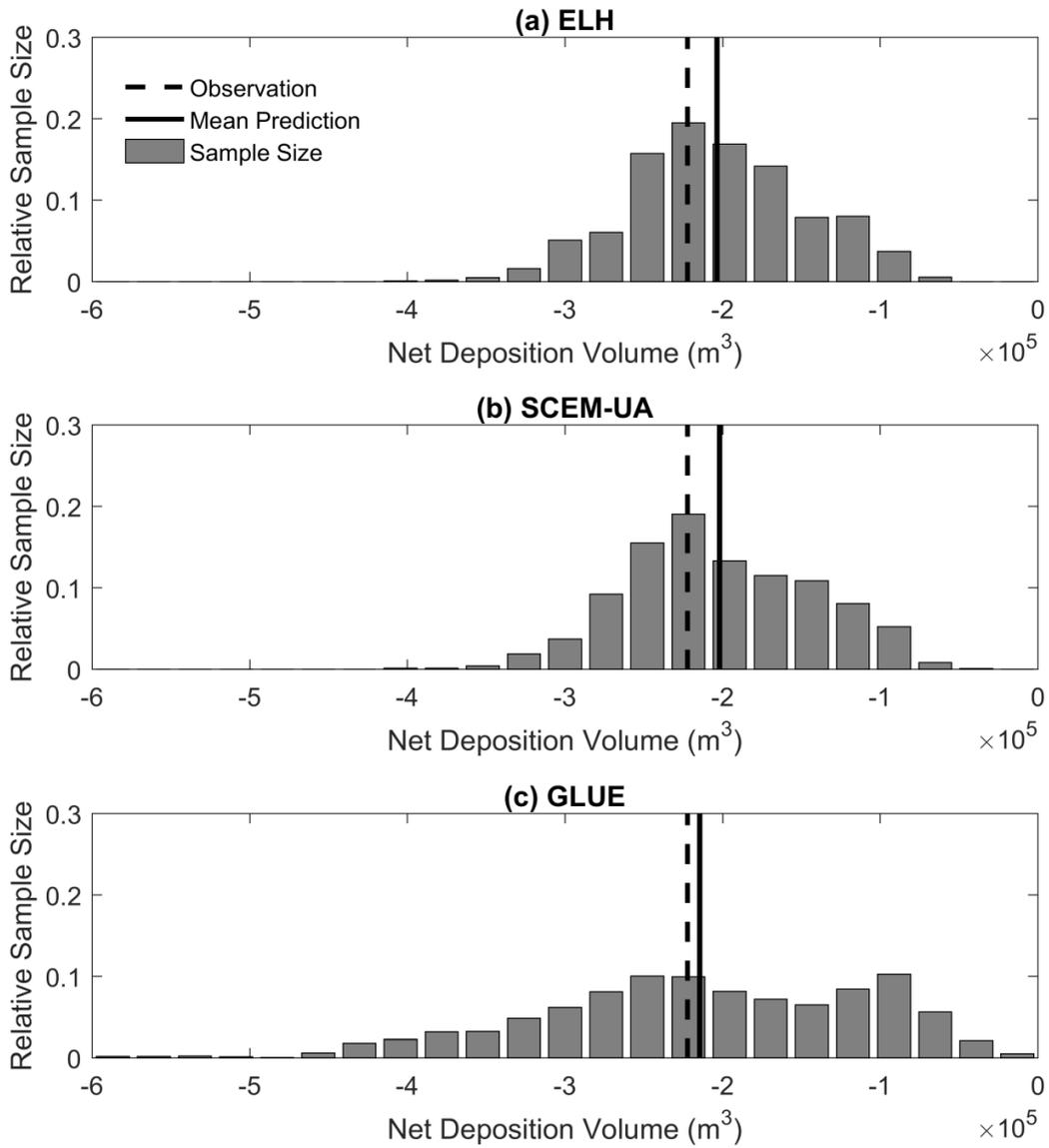


Fig. 14 Observation, mean prediction, and uncertainty distributions for net deposition volume at the cross section located 21 km upstream from the ocean for the forecast period in the Tachia River case.

In order to evaluate the performance and the prediction from each method, three metrics are computed and presented in Fig. 15. First, the NSCE is calculated for the mean prediction from each method during the calibration and forecast periods (Fig. 15a). NSCE can range from $-\infty$ to 1, where 1 means the model prediction perfectly matches the observations, but note that the methods do not use NSCE in their evaluation of the model performance. Overall, all three methods reproduce the observations for the calibration period better than the forecast period, but there is no meaningful difference in the performance of the mean predictions among the methods for both periods.

Second, Fig. 15b compares the prediction uncertainty estimated from each method by calculating the SD of the simulated net deposition volumes (averaged among all cross-sections). The resulting SD values indicate that GLUE has the largest uncertainty in the predictions for both the calibration and forecast periods. Specifically, ELH has a SD that is about 3 % less than SCEM-UA for both periods and 4 % and 63 % less than GLUE for calibration and forecast periods, respectively. Such large forecast uncertainty estimated by GLUE can be explained by the neglect of parameter correlation in GLUE as Sabatine et al. (2015) has previously demonstrated.

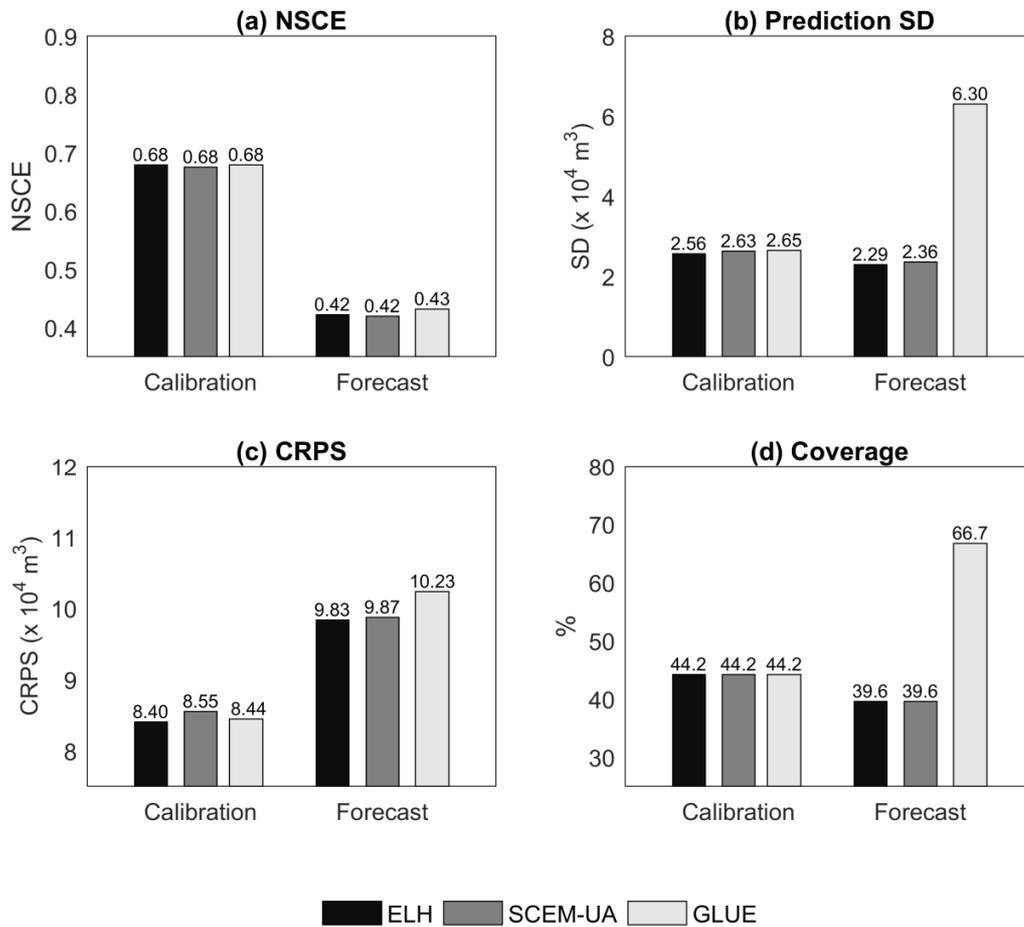


Fig. 15 (a) Nash-Sutcliffe Coefficient of Efficiency (NSCE) values for the mean prediction from each method, (b) standard deviation (SD) of the sample of simulation outputs, (c) Continuous Rank Probability Score (CRPS) values for distributions of the same simulations, and (d) percentage of observations covered by the 99% credible intervals of the same simulations.

Third, the continuous ranked probability score (CRPS) (Brown 1974; Matheson and Winkler 1976; Hersbach 2000) is calculated in order to evaluate the performance considering the predictive distribution (Fig. 15c). The CRPS measures the area between the cumulative distribution of the predictions and the observation as:

$$\text{CRPS} = \frac{1}{N} \sum_{i=1}^N \int [G_i(O) - 1\{O \geq y_i\}]^2 dO \quad (27)$$

where $G_i(O)$ denotes the cumulative distribution of the model outputs O at measurement i , and $1\{O \geq y_i\}$ is the Heaviside function that attains the value 1 if $O \geq y_i$ and the value 0 otherwise (Hersbach 2000). The CRPS has the units of the considered variable, and its minimal value of zero is only achieved when a sample of predictions perfectly matches the observations with no spread. Smaller CRPS values are preferred and indicate better performance of the predictions. It is based on the principle that probabilistic forecasting methods should be designed to maximize the sharpness to make the prediction intervals as narrow as possible subject to covering all the observations (Gneiting et al. 2003). Overall, ELH produces more accurate predictive distributions than the other methods. Such better performance is expected because the ELH shows similar mean prediction accuracy (Fig. 15a) with less spread in its predictions (Fig. 15b).

The reason of remaining uncovered observations should be interpreted as that the methods are not able to fully reflect other uncertainty sources in sediment transport modeling. Errors in the measurements of net deposition volume, which are not considered in this research, can be one of the most important uncertainty factors. Other reasons could be deviations from the assumptions used in each uncertainty analysis algorithm or flaws in the hydraulic and sediment transport model itself (e.g., the use of the Wu (2000) equation for sediment transport capacity). In addition, the predictions are expected to contain more uncertainty from errors in input data or

improper simulation setup as it considers modeling a natural river, rather than the previous numerical experiment cases.

3.5 Conclusions

This study suggested the ELH method to perform uncertainty analysis associated with sediment transport model parameters. Through the case studies conducted in this study, the benefits of the new method have been specified as follows:

1. From the synthetic cases where the posterior distributions are known, ELH reproduces the target values quicker and more accurately than SCEM-UA. Specifically, ELH constructs a histogram of generated parameter values that is a smoother and closer approximation for the target bimodal probability distribution. The improved performance is more notable when considering a number of parameters that are highly correlated. In such a case, ELH is able to estimate the variances and correlation coefficients between the parameters accurately whereas SCEM-UA underestimates those variances due to its lack of detailed balance.
2. In the application to the simulation of a natural river, ELH implements parameter calibration and evaluates prediction uncertainty much quicker than the existing uncertainty methods. When ELH is used, the number of model simulations required for obtaining the reliable estimates decreases about 75% and 86% compared to SCEM-UA and GLUE, respectively. In addition, high performance computing systems with larger numbers of processors is expected to reduce the computation time for ELH much more.
3. From the natural river case, ELH is able to estimate the uncertainty in the parameter values including the correlations between the model parameters. The marginal posterior

parameter distributions estimated by ELH are almost the same as the results from GLUE, but those are slightly different from the estimations of SCEM-UA, which might be due to the different sampling strategies. The correlation coefficients show that both ELH and SCEM-UA can identify the strong interactions between reference shear stress, hiding and exposure coefficient, and active layer thickness multiplier in the SRH-1D model simulation of the Tachia River.

4. For the natural river case, ELH can produce similar estimates of the prediction uncertainty compared to the results from SCEM-UA, whereas GLUE provides larger uncertainties in the model predictions. Specifically, the predictive distributions produced by ELH and SCEM-UA present similar accuracy based on the CRPS values for both calibration and forecast periods. In addition, the performance of both ELH and SCEM-UA are better than GLUE as they consider the parameter correlations when sampling the parameter sets for the forecast simulations (GLUE does not).

Overall, the results support using ELH for evaluating model prediction uncertainty as it can reduce the computational cost remarkably while still providing similar estimates compared to the existing methods. Furthermore, it includes parameter correlations when assessing the prediction uncertainty for sediment transport models. The methodology suggested in this study should be expanded along several avenues in the future. First, the likelihood function used in ELH might restrict the applicability of the method in the field of hydraulic and sediment transport modeling. Such formal function has been criticized for relying too strongly on residual error assumptions because the model errors are correlated, nonstationary, and non-Gaussian in many cases (Beven et al. 2008). Several approaches have been suggested to loosen the stringent assumptions for the residuals (Schoups and Vrugt 2010; Wöhling and Vrugt 2011; Sadegh and Vrugt 2013; Nourali

et al. 2016), but they often make the estimation process more complex or include more parameters to calibrate for specifying the form of the likelihood function. In addition, the function should be able to consider the likelihoods of multiple variables at the same time because sediment transport models generate multiple output variables of interest (e.g., bed elevation, median sediment diameter or D50, and water depth) and these variables are often considered together (Russel et al. 2010; Ahn and Yang 2015). Second, ELH requires a modeler to subjectively determine some algorithmic parameters such as the number of replications and the LHS size for the replicated LHS method, the number of duplications for the most likely parameter sets, and the SD change threshold. The numbers used in this study were specified based on related literature or preliminary tests, but they should be chosen in a more formal and robust way because such decisions might affect the uncertainty estimates or the number of required model simulations. Third, future research can additionally improve the efficiency of the uncertainty method. Although ELH significantly decreases the computational costs, it used about 8,500 simulations to obtain the final estimates for prediction uncertainty, which would be still expensive for more complex and higher dimensional modeling cases. Metamodeling approaches that mimic the likelihood response surface using a small sample of model outputs to approximate the posterior pdf (Khu and Werner 2003) might be an alternative. However, such an approach substantially increases the level of complexity of the uncertainty analysis relative to GLUE, and it would be inefficient as the problem dimensionality increases (Ong et al. 2004). Fourth, future work should also focus on other sources of uncertainty such as the model's mathematical structure (including the selection of the transport equation), model input data errors, and the observations used for calibration.

References

- Ahn, J. K., and C. T. Yang (2015), Determination of recovery factor for simulation of non-equilibrium sedimentation in reservoir, *International Journal of Sediment Research*, 30(1), 68-73.
- Ajami, N. K., Q. Duan, and S. Sorooshian (2007), An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, *Water Resources Research*, 43, W01403, doi:10.1029/2005WR004745.
- Athira, P., and K. P. Sudheer (2015), A method to reduce the computational requirement while assessing uncertainty of complex hydrological models, *Stochastic Environmental Research and Risk Assessment*, 29(3), 847-859.
- Beven, K. J., and A. Binley (1992), The future of distributed models: model calibration and uncertainty prediction, *Hydrological Processes*, 6(3), 279-298.
- Beven, K. J., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *Journal of Hydrology*, 249(1), 11-29.
- Beven, K. J., P. Smith, and J. Freer (2008), So just why would a modeler choose to be incoherent?, *Journal of Hydrology*, 354(1), 15-32.
- Blasone, R. S., H. Madsen, and D. Rosbjerg (2008), Uncertainty assessment of integrated distributed hydrological models using GLUE with Markov chain Monte Carlo sampling, *Journal of Hydrology*, 353(1), 18-32.
- Blazkova, S., K. J. Beven, and A. Kulasova (2002), On constraining TOPMODEL hydrograph simulations using partial saturated area information, *Hydrological Processes*, 16(2), 441- 458.
- Box, G. E. P., and G. C. Tiao (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Boston, Massachusetts.
- Brazier, R. E., K. J. Beven, J. Freer, and J. S. Rowan (2000), Equifinality and uncertainty in physically based soil erosion models: Application of the glue methodology to WEPP-The water erosion prediction project for sites in the UK and USA, *Earth Surface Processes and Landforms*, 25(8), 825- 845.
- Brown, T. A. (1974), *Admissible Scoring Systems for Continuous Distributions*, Technical Report P-5235, Rand Corporation, Santa Monica, California.
- Chib, S., and E. Greenberg (1995), Understanding the Metropolis-Hastings algorithm, *The American Statistician*, 49(4), 327-335.
- Cho E., G. B. Arhonditsis, J. Khim, S. Chung, and T. Heo (2016), Modeling metal-sediment interaction processes: Parameter sensitivity assessment and uncertainty analysis, *Environmental Modelling and Software*, 80, 159-174.

- Christensen R., W. Johnson, A. Branscum, and T.E. Hanson (2011), *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians*, CRC Press, Boca Raton.
- Duan, Q., S. Sorooshian, and V. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resources Research*, 28(4), 1015-1031.
- Foglia, L., M. C. Hill, S. W. Mehl, and P. Burlando (2009), Sensitivity analysis, calibration, and testing of a distributed hydrological model using error-based weighting and one objective function, *Water Resources Research*, 45, W06427, doi:10.1029/2008WR007255.
- Freer, J. E., H. McMillan, J. J. McDonnell, and K. J. Beven (2004), Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures, *Journal of Hydrology*, 291(3), 254- 277.
- Frings, R. M., H. Schüttrumpf, and S. Vollmer (2011), Verification of porosity predictors for fluvial sand-gravel deposits, *Water Resources Research*, 47, W07525, doi:10.1029/2010WR009690.
- Gelman, A., and D. B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Statistical Science*, 7, 457-472.
- Gneiting, T., A. E. Raftery, F. Balabdaoui, and A. Westveld (2004), Verifying probabilistic forecasts: Calibration and sharpness, *Proceedings of 17th Conference on Probability and Statistics in the Atmospheric Sciences*.
- Green, P. J. (2001), A primer on Markov chain Monte Carlo, in *Complex Stochastic Systems*, edited by O. E. Barndorff-Nielsen, D. R. Cox, and C. Klüppelberg, pp. 1-62, Chapman and Hall/CRC, Boca Raton.
- Greimann, B. P., Y. Lai, and J. V. Huang (2008), Two-dimensional total sediment load model equations, *Journal of Hydraulic Engineering*, 134(8), 1142-1146
- Haario, H., E. Saksman, and J. Tamminen (2001), An adaptive Metropolis algorithm, *Bernoulli*, 7(2), 223-242.
- Haario, H., M. Laine, A. Mira, and E. Saksman (2006), DRAM: efficient adaptive MCMC, *Statistics and Computing*, 16(4), 339-354.
- Hersbach, H. (2000), Decomposition of the continuous ranked probability score for ensemble prediction systems, *Weather and Forecasting*, 15(5), 559-570.
- Huang, J. V., and B. P. Greimann (2013), *User's Manual for SRH-1D 3.0*, Bureau of Reclamation, U.S. Department of the Interior, Denver.
- Jia, Y., and T. B. Culver (2006), Robust optimization for total maximum daily load allocations, *Water Resources Research*, 42, W02412, doi:10.1029/ 2005WR004079.
- Jones, B., and R. T. Johnson (2009), Design and analysis for the Gaussian process model, *Quality and Reliability Engineering International*, 25(5), 515-524.

- Kanso, A., G. Chebbo, and B. Tassin (2005), Bayesian analysis for erosion modelling of sediments in combined sewer systems, *Water Science and Technology*, 52(5), 135-142.
- Khu, S. T., and M. G. F. Werner (2003), Reduction of Monte-Carlo simulation runs for uncertainty estimation in hydrological modelling, *Hydrology and Earth System Sciences Discussions*, 7(5), 680-692.
- Lai, Y., and B. P. Greimann (2010), *SRH Model Applications and Progress Report on Bank Erosion and Turbidity Current Models*, Bureau of Reclamation, U.S. Department of the Interior, Denver.
- Laloy, E., and J. A. Vrugt (2012), High-dimensional posterior exploration of hydrologic models using multiple-try DREAM (ZS) and high-performance computing, *Water Resources Research*, 48, W01526, doi:10.1029/2011WR010608.
- Limeneros, J. T. (1970), *Determination of the Manning coefficient from Measured Bed Roughness in Natural Channels*, Geological Survey Water-Supply Paper 1898-B, U. S. Department of the Interior, Washington D.C.
- Loeppky, J. L., J. Sacks, and W. J. Welch (2009), Choosing the sample size of a computer experiment: A practical guide, *Technometrics*, 51(4), 366-376.
- Matala, A. (2008). *Sample Size Requirement for Monte Carlo Simulations using Latin Hypercube Sampling*, Technical Report, Systems Analysis Laboratory, Department of Engineering Physics and Mathematics, Helsinki University of Technology, Finland.
- Matheson, J. E., and R. L. Winkler (1976), Scoring rules for continuous probability distributions, *Management Science*, 22(10), 1087-1096.
- McKay, M. D., R. J. Beckman, and W. J. Conover (1979), Comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, 21(2), 239-245.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *The Journal of Chemical Physics*, 21(6), 1087-1092.
- Mueller, E. R., J. Pitlick, and J. M. Nelson (2005), Variation in the reference Shields stress for bed load transport in gravel bed streams and rivers, *Water Resources Research*, 41, W04006, doi:10.1029/2004WR003692.
- Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I-A discussion of principles, *Journal of Hydrology*, 10(3), 282-290.
- Nourali, M., B. Ghahraman, M. Pourreza-Bilondi, and K. Davary (2016), Effect of formal and informal likelihood functions on uncertainty assessment in a single event rainfall-runoff model, *Journal of Hydrology*, 540, 549-564.
- Ong, Y. S., P. B. Nair, A. J. Keane, and K. W. Wong (2004), Surrogate assisted evolutionary optimization frameworks for high-fidelity engineering design problems, in *Knowledge*

- Incorporation in Evolutionary Computation*, edited by Y. Jin, pp. 307- 332, Springer, New York.
- Ruark, M. D., J. D. Niemann, B. P. Greimann, and M. Arabi (2011), Method for assessing impacts of parameter uncertainty in sediment transport modeling applications, *Journal of Hydraulic Engineering*, 137(6), doi:10.1061/(ASCE)HY.1943-7900.0000343.
- Ruether, N., J. M. Singh, N. R. B. Olsen, and E. Atkinson (2005), 3-D computation of sediment transport at water intakes, *Proceedings of the Institution of Civil Engineers Water Management*, 158, 1-8.
- Russell, K., B. P. Greimann, B. Cluer, T. Hepler, D. King, S. O'Meara, A. Simon, J. Godaire, and D. Salas (2010), Klamath reservoir sediment characterization and drawdown impacts for the dam removal investigation, *Proceedings of 2010 AGU Fall Meeting*, 1053.
- Sabatine, S. M., J. D. Niemann, B. P. Greimann (2015), Evaluation of parameter and model uncertainty in simple applications of a 1D sediment transport model, *Journal of Hydraulic Engineering*, 141(5), doi:10.1061/(ASCE)HY.1943-7900.0000992.
- Sadegh, M., and J. A. Vrugt (2013), Bridging the gap between GLUE and formal statistical approaches: approximate Bayesian computation, *Hydrology and Earth System Sciences*, 17(12).
- Sadegh, M., and J. A. Vrugt (2014), Approximate Bayesian computation using Markov chain Monte Carlo simulation: Dream(abc), *Water Resources Research*, 50(8), doi:10.1002/2014WR015386.
- Schmelter, M. L., and D. K. Stevens (2013), Traditional and Bayesian statistical models in fluvial sediment transport, *Journal of Hydraulic Engineering*, 139(3), doi:10.1061/(ASCE)HY.1943-7900.0000672.
- Schmelter, M. L., M. B. Hooten, and D. K. Stevens (2011), Bayesian sediment transport model for unisize bed load, *Water Resources Research*, 47, W11514, doi:10.1029/2011WR010754.
- Schmelter, M. L., P. R. Wilcock, M. B. Hooten, and D. K. Stevens (2015), Multi-fraction Bayesian sediment transport model, *Journal of Marine Science and Engineering*, 3(3), 1066-1092.
- Schmelter, M. L., S. O. Erwin, and P. R. Wilcock (2012), Accounting for uncertainty in cumulative sediment transport using Bayesian statistics, *Geomorphology*, 175, 1-13.
- Schoups, G., and J. A. Vrugt (2010), A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors, *Water Resources Research*, 46, W10531, doi:10.1029/2009WR008933.
- Stedinger, J. R., R. M. Vogel, S. U. Lee, and R. Batchelder (2008), Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, *Water Resources Research*, 44, W00B06, doi:10.1029/2008WR006822.

- Stein, M. (1987), Large sample properties of simulations using Latin hypercube sampling, *Technometrics*, 29(2), 143-151.
- Ter Braak, C. J. (2006), A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces, *Statistics and Computing*, 16(3), 239-249.
- Thiemann, M., M. Trosset, H. Gupta, and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrologic models, *Water Resources Research*, 37(10), 2521- 2535.
- Tolson, B. A., and C. A. Shoemaker, (2008), Efficient prediction uncertainty approximation in the calibration of environmental simulation models, *Water Resources Research*, 44, W04411, doi:10.1029/2007WR005869.
- Uhlenbrook, S., and A. Sieber (2005), On the value of experimental data to reduce the prediction uncertainty of a process-oriented catchment model, *Environmental Modelling and Software*, 20(1), 19-32.
- van Griensven, A., and T. Meixner (2007), A global and efficient multi-objective auto-calibration and uncertainty estimation method for water quality catchment models, *Journal of Hydroinformatics*, 9(4), 277-291.
- van Griensven, A., T. Meixner, S. Grunwald, T. Bishop, M. Diluzio, and R. Srinivasan (2006), A global sensitivity analysis tool for the parameters of multi-variable catchment models, *Journal of Hydrology*, 324(1), 10-23.
- Vrugt, J. A., and C. J. Ter Braak (2011), DREAM (D): an adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems, *Hydrology and Earth System Sciences*, 15(12), 3701.
- Vrugt, J. A., C. J. Ter Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon (2009), Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, *International Journal of Nonlinear Sciences and Numerical Simulation*, 10(3), 273-290.
- Vrugt, J. A., C. J. Ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resources Research*, 44, W00B09, doi:10.1029/2007WR006720.
- Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003), A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resources Research*, 39(8), 1201, doi:10.1029/2002WR001642.
- Wöhling, T., and J. A. Vrugt (2011), Multiresponse multilayer vadose zone model calibration using Markov chain Monte Carlo simulation and field water retention data, *Water Resources Research*, 47, W04510, doi:10.1029/2010WR009265.
- Wu, F. C., and C. C. Chen (2009), Bayesian updating of parameters for a sediment entrainment model via Markov chain Monte Carlo, *Journal of Hydraulic Engineering*, 135(1), 22-37.

Wu, W., S. S. Wang, and Y. Jia (2000), Nonuniform sediment transport in alluvial rivers, *Journal of Hydraulic Research*, 38(6), 427-434.

Zak, S. K., and K. J. Beven (1999), Equifinality, sensitivity and predictive uncertainty in the estimation of critical loads, *Science of the Total Environment*, 236(1), 191-214.

CHAPTER 4

COMBINING PREDICTIONS AND ASSESSING UNCERTAINTY FROM SEDIMENT TRANSPORT EQUATIONS USING MULTIVARIATE BAYESIAN MODEL AVERAGING

Abstract

Bayesian model averaging (BMA) is a statistical hydrologic method that reduces the effects of imperfections in a single model prediction and characterizes the uncertainty due to the mathematical model structure. The authors apply BMA to quantify how the uncertainty originating from the selection of a transport equation affects the multivariate predictions from a sediment transport model. To overcome the limitation of the existing BMA that is only able to consider a single variable, the likelihood function of BMA is modified to calculate the likelihood of each equation with multiple variables. In addition, the coefficients of variation are used to describe the change in the uncertainty scale of model predictions. The proposed modification is applied to the transport equations included in Sedimentation and River Hydraulics—One Dimension (SRH-1D) program. For two flume experiment cases, the multivariate BMA models improve the predictions over the individual transport equations and the univariate BMA models and provide more realistic description for its predictive uncertainty.

4.1 Introduction

Sediment transport models are widely used to predict impacts of potential river restoration activities, but the predictions from these models always possess uncertainty. One major source of the uncertainty is the mathematical equation that is used to compute sediment transport capacity in these models. Several equations have been empirically developed for different fluvial conditions by applying various simplifications and assumptions. For example, the

equations are available for a certain range of sediment size such as sand (Yang 1973; Yang 1979), gravel (Yang 1984), or sand with high concentration of fine particles (Yang 1996). The equations are also designed for different transport types including bedload (Meyer-Peter and Müller 1948; Einstein 1950; Rottner 1959; Bijker 1968; Bagnold 1980; Parker 1990; Dibajnia and Watanabe 1992; Ribberink 1998; Wilcock and Crowe 2003) and total material load (Laursen 1958; Engelund and Hansen 1972; Ackers and White 1973; Bailard 1981; Brownlie 1981; van Rijn 1989; Wu et al. 2000). In addition, some equations revised existing equations to fit the data from additional flume experiments (Wallingford 1990; Wong and Parker 2006) and natural rivers (Madden 1993; Yang 1996; Geumann 2009). Thus, predictions based solely on a single transport formula are inherently uncertain because no single equation perfectly represents a physical system. Past research has focused on the fact that different equations provide different predictions even if the equations are calibrated using the same data (Wilcock 2001; Camenen and Larroudé 2003; Pinto et al. 2006; Bertin et al. 2008; Schmelter et al. 2011), but little consideration is given to quantifying the uncertainty originating from the selection of an sediment transport equation for forecast scenarios.

Multi-model averaging methods offer a formal way to reduce the effects of imperfections in a single model prediction and assess the uncertainty due to the model's mathematical structure. The methods combine the predictions from a set of competing models and provide averaged forecasts and their credible intervals. Several techniques of multi model averaging include: equal weights averaging (Anderson 1965), Bates-Granger averaging (Bates and Granger 1969), Granger-Ramanathan averaging (Granger and Ramanathan 1984), Akaike information criterion-based model averaging (Buckland et al. 1997), Bayesian information criterion-based model averaging (Burnham and Anderson 2002), Mallows model averaging (Hansen 2007), and

Bayesian model averaging (BMA) (Hoeting et al. 1999; Raftery et al. 2005). Among the available methods, BMA has been shown to produce the most accurate predictions and more realistic description of the predictive uncertainty in various case studies of hydrologic modeling (Clyde 1999, Viallefont et al. 2001, Ye et al. 2004, Raftery et al. 2005, Ajami et al. 2007, Slaughter et al. 2010). BMA represents the uncertainty associated with each competing model using a normal distribution that is centered on that model's predictions, and it combines the distributions of the models by weighted-averaging (Fig. 16). To generate the best forecasts, the weights and the standard deviations, which determine the normal distributions, of the competing models are estimated to maximize the likelihood of BMA for the calibration dataset. BMA assigns higher weights on better performing models, and it accounts for both the uncertainty due to the model selection (represented by the spread in the model predictions) and the uncertainty associated with each model (represented by the normal distributions) (Ellison 2004, Vrugt and Robinson 2007, Sabatine et al. 2015).

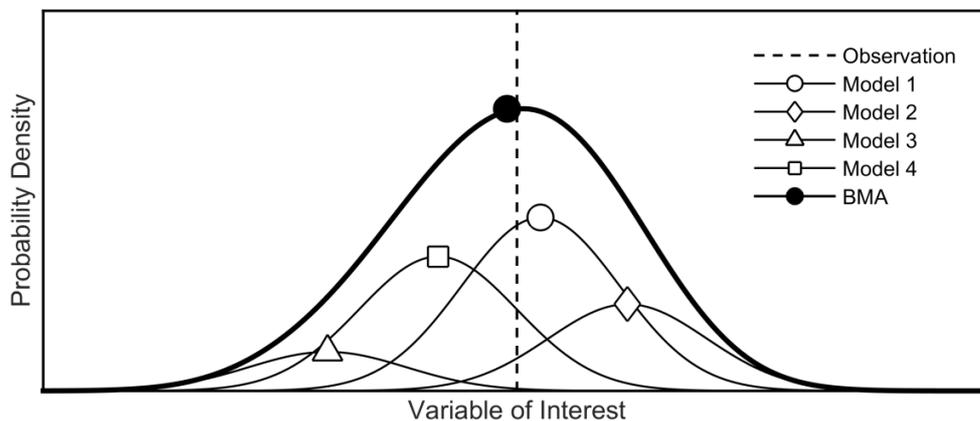


Fig. 16 Illustration of BMA distribution generated by a weighted-average of normal distributions from four competing models.

BMA has recently been applied to hydraulic and sediment transport modeling (Sabatine et al. 2015), but it was constrained to cases that consider only a single variable because the likelihood function of BMA is not easily generalized to account for multiple variables at the same time. If BMA is modeled to maximize the likelihoods of only a single variable, the other variables might be estimated poorly by the BMA model. Moreover, any probabilistic information cannot be obtained for the other variables because the standard deviations for a single variable are not able to be applied to different variables. This limitation is critical because sediment transport models generate multiple output variables (e.g., bed elevations, D50 sizes, and water depths) from a single simulation and these variables are usually investigated at the same time in natural rivers (Duan et al. 2008, Huang and Greimann 2010, Russel et al. 2010, Ahn et al. 2013, Ahn and Yang 2015). Several functions have been suggested to compute likelihoods of multiple objectives or variables (Beven and Binley 1992; Yapo et al. 1998; Mo and Beven 2004; Yang et al. 2004, Chahinian and Moussa 2007; Beven 2011). For example, van Griensven and Meixner (2007) proposed the Global Optimization Criterion to compute the likelihoods of multiple variables by considering qualitative and quantitative differences between the variables. Ruark et al. (2011) extended the Nash-Sutcliffe Coefficient of Efficiency (NSCE) (Nash and Sutcliffe 1970) to calculate the likelihoods by weighting multiple variables based on their sensitivities to model parameters. However, the existing functions would be inappropriate to apply to BMA because they compute the likelihoods based on residuals, which are the differences between model outputs and observations, while BMA likelihoods are calculated using a conditional probability density function (pdf) of observed data for a given model output.

Another limitation in the application to sediment transport models is that the BMA assumption, which applies the same standard deviations to individual model predictions at all locations and

times, cannot reflect the change in the uncertainty scale. The variables such as sediment transport rate, deposition or scour volumes are measured on ratio scales, and the residuals generally increase with respect to their quantities. This implies that the uncertainty in each model prediction depends on the scale of those variables. Thus, the BMA pdf for the quantitative variables might overestimate (or underestimate) the predictive uncertainty if the standard deviations are calibrated using data with large (or small) values and applied to the individual model simulations of the forecast scenario. To consider the heteroscedasticity of uncertainty for BMA, a number of methods including the temporal difference algorithm (Downey and Sanner 2010), geostatistical kriging (Kleiber et al. 2011), sequential data assimilation coupling (Parrish et al. 2012), decay function updating (Veenhuis 2014) have been suggested in the field of hydrologic modeling. However, those methods usually require continuous updating of the calibration dataset, which is not applicable for hydraulic and sediment transport modeling.

The objective of this research is to develop and test a multivariate version of BMA to assess the uncertainty associated with the selection of a transport equation in a one-dimensional sediment transport modeling. The likelihood function suggested in this study is intended to compute the likelihoods of BMA with respect to multiple variables, and the BMA assumption is revised to allow the change in the uncertainty scale for quantitative variables. To test the multivariate BMA, two published flume cases including a depositional experiment (Seal et al. 1997) and an erosional experiment (Pender et al. 2001) with non-cohesive materials are used. Four sediment transport equations, included in the Sedimentation and River Hydraulics – One Dimension (SRH-1D) (Huang and Greimann 2013) program, are used to simulate the flume cases, and each equation is treated as a separate model. The forecast of multivariate BMA is then compared to

existing univariate BMA in both prediction accuracy and observation coverage of their credible bounds.

4.2 Methodology

4.2.1 Existing BMA Method

BMA (Hoeting et al. 1999) defines the uncertainty in the model prediction for variable Δ using a pdf $p(\Delta|\mathbf{M},\mathbf{y})$, which is the posterior distribution of Δ given a set of competing models $\mathbf{M} = \{M_1, M_2, \dots, M_A\}$ and calibration dataset \mathbf{y} :

$$p(\Delta|\mathbf{M},\mathbf{y}) = \sum_{a=1}^A p(\Delta|M_a,\mathbf{y})p(M_a|\mathbf{y}) \quad (28)$$

where A is the number of competing models, $p(\Delta|M_a,\mathbf{y})$ is the posterior distribution from the model a , and $p(M_a|\mathbf{y})$ is the posterior probability that reflects how well model M_a fits the dataset \mathbf{y} . As the posterior probabilities of all competing models add up to one, they can be considered as weights of each model. Thus, the BMA pdf is a weighted average of the posterior distributions given each of the individual models.

BMA has been applied to dynamic models using the assumptions: (1) the forecast O_a from each M_a is the most likely result from that model, and (2) the uncertainty associated with each model can be represented using a normal distribution that is centered on the predictions O_a (Raftery et al. 2005). To obtain the best forecasts, the parameters of each model M_a are optimized using the dataset \mathbf{y} . BMA represents the posterior distribution of Δ given a set of competing model predictions $\mathbf{O} = \{O_1, \dots, O_A\}$ as:

$$p(\Delta|\mathbf{O}) = \sum_{a=1}^A w_a p(\Delta|O_a, \sigma_a) \quad (29)$$

where w_a is model a 's weight, which represents the posterior probability that prediction O_a is the best one among \mathbf{O} , and $p(\Delta|O_a, \sigma_a)$ is the pdf of Δ occurring by model a 's normal distribution with mean O_a and standard deviation σ_a . To generate the BMA pdf, the weights w_a and standard deviations σ_a of competing models are estimated to maximize the BMA likelihood L_{BMA} over all locations and times in the calibration dataset \mathbf{y} . Assuming that the residuals of each model are independent, the log-likelihood function is used for algebraic simplicity and numerical stability:

$$L_{BMA} = \sum_{i=1}^N \log \left[\sum_{a=1}^A w_a g(y_i | O_{ai}, \sigma_a) \right] \quad (30)$$

where $g(y_i | f_{ai}, \sigma_a)$ is the probability density of the observation y_i given model a 's prediction O_{ai} and standard deviation σ_a for the measurement i in the calibration period. An optimization algorithm called Expectation-Maximization (EM) (Dempster et al. 1977) has been used to find the best values of w_a and σ_a . The EM is easy to implement, and its algorithm steps are designed in a way that they always satisfy the constraint that the model weights w_a are positive and add up to one (Vrugt et al. 2008, Givens and Hoeting 2012).

BMA then applies the same weights w_a and standard deviations σ_a obtained from the calibration period to the model predictions at all locations and times in the forecast scenario (Raftery et al. 2005, Sabatine et al. 2015). The mean of each model's normal distribution is changing as the models make their forecasts, but the other quantities all remain fixed. BMA provides a deterministic prediction using an expectation of the BMA pdf, and it is the same as a weighted-average of the predictions from the competing models:

$$E(\Delta|\mathbf{O}) = \sum_{a=1}^A w_a O_a \quad (31)$$

The uncertainty contained in the BMA prediction can be described using credible intervals (CI), which are calculated using specific quantiles of the BMA pdf, and the BMA CIs quantify the predictive uncertainty, which originates from both the uncertainty due to the model selection and the uncertainty associated with each model.

4.2.2 Multivariate BMA Model

Multivariate BMA uses a modified BMA likelihood function to compute the likelihoods of multiple variables within a single formula. This method applies two primary assumptions: (1) each model has same weight w_a for all variables so that the model performance is evaluated with respect to multiple variables and (2) each model has different standard deviations for different variables σ_{aj} to consider the fact that different variables do not necessarily have identical standard deviations. The multivariate BMA likelihood L_{mBMA} is calculated as:

$$L_{mBMA} = \sum_{j=1}^J \left\{ \frac{1}{N_j} \sum_{i=1}^{N_j} \log \left[\sum_{a=1}^A w_a g(y'_{ij} | O'_{aj}, \sigma'_{aj}) \right] \right\} \quad (32)$$

where J is the number of variables considered, N_j is the number of observations for variable j in the calibration dataset, and $g(\cdot)$ represents the probability density of reproducing the normalized observation y'_{ij} given model a 's normalized prediction O'_{aj} with standard deviation σ'_{aj} for the normalized variable j at a location or time i . In Eq. (32), the likelihood of each variable j is divided by N_j to avoid treating highly correlated (or dependent) observations of specific variables (e.g., adjacent bed elevations) as independent. In addition, the normalized values of each variable are used in the likelihood function because $g(\cdot)$ strongly depends on the scale or units of

the variables so the overall BMA likelihoods might be distorted when multiple variables with different scales are included. While BMA assumes normality of the variables, some quantitative variables might not be properly expressed using a single standard deviation, as discussed in the previous paragraph. A non-parametric normalization is then used as follows:

$$y'_{ij} = \frac{y_{ij}}{y_{j,50}} \quad (33)$$

$$O'_{aj} = \frac{O_{aj}}{y_{j,50}} \quad (34)$$

where y_{ij} is the observation, O_{aj} is model a 's prediction for variable j at a measurement i , and $y_{j,50}$ is the median of the observed variable j . By using the normalized variables, the multivariate BMA likelihood L_{mBMA} can be computed independent from the variable's scale or units. As the model standard deviation σ'_{aj} has a normalized scale, it needs to be transformed to the original scale as:

$$\sigma_{aj} = \sigma'_{aj} \times y_{j,50} \quad (35)$$

The w_a and σ_{aj} values from the EM algorithm are then applied to Equation (29) to generate the BMA pdf separately for each variable.

In multivariate BMA, the coefficient of variation (CV) can be used to reflect the change in the uncertainty scale of a variable instead of the standard deviation, following the idea of a linear dependency in standard deviations (Vrugt and Robinson 2007). The CV is a ratio of the standard deviation to the mean. Multivariate BMA assumes that each model has a single value of CV for a quantitative variable, and applies the identical CV to each model's predictions at all locations

and times. By applying the CV, the standard deviation of each model is allowed to vary based on the model predictions as:

$$\sigma'_{aj} = CV_{aj} \times O'_{aj} \quad (36)$$

where CV_{aj} is model a 's coefficient of variation for variable j and σ'_{aj} is model a 's standard deviation for the normalized variable j at measurement i . If the model prediction value is a negative, its absolute value $|O'_{aj}|$ can be used in the Eq. (36) (e.g., the simulated deposition volume would be negative where the results indicate erosion). The BMA likelihood for the quantitative variable q can be calculated as:

$$L_q = \sum_{i=1}^{N_q} \log \left[\sum_{a=1}^A w_a g \left(y'_{iq} \mid f'_{aiq}, \sigma'_{aiq} \right) \right] \quad (37)$$

where N_q is the total number of observations for the variable q . The likelihood L_q then can be divided by N_q and inserted into Eq. (32) in order to compute the multivariate BMA likelihood. After obtaining the best model weights and the CVs, the BMA pdf for this variable can be generated as:

$$p(\Delta \mid \mathbf{Y}) = \sum_{a=1}^A w_a p(\Delta \mid O_a, CV_a) \quad (38)$$

4.3 Application

4.3.1 Flume Experiments

Two flume experiments with non-cohesive sediment transport, conducted by Seal et al. (1997) and Pender et al. (2001), are considered as case studies to evaluate the multivariate BMA. These

experiments are chosen, rather than natural rivers, because flume conditions such as channel geometry, flow rate, sediment supply, and bed materials are well-documented for both cases. That reduces uncertainty about the system configuration. In addition, the computational costs are low because a sediment transport model is able to quickly simulate the flume cases. Scour and sedimentation processes occur separately in each experiment.

The Seal et al. (1997) experiment was designed to investigate downstream fining and sediment sorting during aggradation in narrow channels. The flume was 0.3 m wide, 45 m long with rectangular shape, and initial slope was 0.002. No initial bed material was used and water discharge was steadily applied at 0.049 m³/s for 64 hours. At the upstream end, sediment was supplied with the rate of 0.047 kg/s during the experiment, and the supplied material was a mixture of sand and gravel ranging from 0.125 mm to 65 mm with median diameter of 5 mm. Bed elevations were measured every 4-5 hours at 18 locations, and D50 of deposited materials along the channel distance were measured at five different times. Both bed elevation profile and D50 distribution are considered as variables of interest for multivariate BMA modeling, and the dataset is divided into a calibration period (hours 0 to 32) and a forecast period (hours 32 to 64) (Table 6).

The Pender et al. (2001) experiment was designed to simulate bed degradation and investigate changes in transport rate. The flume was 0.8 m wide, 20 m long, trapezoidal shape with 45° side slope, and initial slope was 0.0026. Initial bed material was a mixture of sand and gravel ranging from 0.25 mm to 22.63 mm. Water discharge was steadily applied at 0.117 m³/s for 84.6 hours, and there was no sediment supply. Bed elevations were measured every 2-3 hours at 21-42 locations, and bed load transport rate was measured at 5 m from the downstream end during the experiment. Both the bed elevation profile and sediment transport rate are considered as

variables of interest in BMA, and the dataset is divided into a calibration period (hours 0 to 32.1) and a forecast period (hours 32.1 to 84.6) (Table 6).

Table 6 Available Observations from Two Experiments for the Calibration and Forecast Periods of BMA Modeling.

Experiment	Period	Duration (hour)	Flow Rate (m ³ /s)	Supplied Sediment (kg/s)	Number of Observations		
					Bed Elevation	D50	Sediment Transport Rate
Seal et al. (1997)	Calibration	0 – 32	0.049	0.094	126	46	-
	Forecast	32 – 64			87	18	-
Pender et al. (2001)	Calibration	0 – 32.1	0.117	0	168	-	17
	Forecast	32.1 – 84.6			203	-	17

4.3.2 BMA Modeling for Sediment Transport Equations

Four equations in SRH-1D are treated as separate sediment transport models for BMA modeling in this study. The following paragraphs describe the three equations for sediment transport capacity considered here in addition to the Wu (2000) equation (Section 2.3.1).

The Parker (1990) and Wilcock and Crowe (2003) (W&C) equations were developed to compute bed load transport capacity using the following form:

$$q_{bk} = \frac{p_k (\tau_g / \rho)^{1.5}}{g [(\rho_s / \rho) - 1]} F(\varphi_k) \quad (39)$$

where q_{bk} is volumetric bed load transport rate per unit width, p_k is percentage of materials available for grain size class k , τ_g is grain shear stress, ρ is density of water, ρ_s is density of the sediment, and φ_k is a measure of shear stress relative to reference shear stress.

In the Parker equation, the empirical function $F(\varphi_k)$ was fit to data from the field with mixed size gravel (2 mm to 64 mm) as:

$$F(\varphi_k) = \begin{cases} 11.93(1-0.853/\varphi_k)^{4.5} & \varphi_k > 1.59 \\ 0.002183 \exp\left[14.2(\varphi_k - 1) - 9.28(\varphi_k - 1)^2\right] & 1 < \varphi_k \leq 1.59 \\ 0.002183\varphi_k^{14.2} & \varphi_k \leq 1 \end{cases} \quad (40)$$

On the other hand, the W&C equation defines the function $F(\varphi_k)$ using a mixture of sand and gravel (62.5 μm to 64 mm) as:

$$F(\varphi_k) = \begin{cases} 14(1-0.894/\sqrt{\varphi_k})^{4.5} & \varphi_k \geq 1.35 \\ 0.002002\varphi_k^{7.5} & \varphi_k < 1.35 \end{cases} \quad (41)$$

The modified Meyer-Peter Muller (MPM) equation (Wong and Parker 2006) was also suggested to compute bed load transport capacity using mixed materials of medium sand to coarse gravel (0.38 mm to 28.65 mm) as:

$$q_{bk} = 3.97 g^{0.5} [(\rho_s/\rho) - 1]^{0.5} d_k^{1.5} \left[\frac{RS_f}{[(\rho_s/\rho) - 1]d_k} - 0.0495 \right]^{1.5} \quad (42)$$

where d_k is median particle diameter of class k , R is hydraulic radius, and S_f is energy slope.

For BMA modeling, the eight parameters of each model were estimated using an optimization algorithm and the dataset from the calibration period (lead of the standard deviations.

Table 7). The optimization was performed using Multivariate Shuffled Complex Evolution Metropolis – Uncertainty Analysis (MSU) (Sabatine et al. 2015), which is able to compute the likelihood of parameter sets using multiple model output variables. The roughness coefficient

has units of $\text{s}\cdot\text{m}^{-1/3}$, and all the other parameters are non-dimensional. Fig. 17 shows the observations and best outputs from individual models for the calibration period of each case. For the Seal et al. (1997) experiment, the outputs from the Parker, W&C, and Wu models are similar in both bed elevations and D50 size distributions, and they seem better than the outputs from the MPM model. For the Pender et al. (2001) experiment, all the models provided similar bed elevation profiles, but the W&C model simulated bed forms along the channel, and noticeable differences between the model outputs for sediment transport rate occur during hours 8 to 20. Using the individual model outputs, two univariate BMA models and one multivariate BMA model were developed for each case by applying standard deviations for two variables for which observations are available from each experiment. An additional multivariate BMA model was also developed for the Pender et al. (2001) case by applying the coefficient of variation for sediment transport rates instead of the standard deviations.

Table 7 Optimized Eight Parameters of SRH-1D Used for Each Model's Simulation of Two Experiments.

Parameter	Seal et al. (1997)				Pender et al. (2001)			
	Parker	W&C	MPM	Wu	Parker	W&C	MPM	Wu
Manning's roughness	0.023	0.024	0.025	0.023	0.015	0.016	0.015	0.015
Reference shear stress	0.066	0.052	0.049	0.046	0.055	0.015	0.049	0.038
Hiding and exposure coefficient	0.958	0.868	0	0.797	0.037	0.006	0	0.003
Active layer thickness multiplier	15.270	11.236	4.934	3.145	3.151	0.130	1.494	2.930
Deposition recovery factor	0.873	0.688	0.680	0.902	0.314	0.718	0.131	0.772
Scour recovery factor	0.077	0.463	0.123	0.519	0.241	0.511	0.334	0.443
Bedload adaptation length multiplier	1.155	3.169	9.594	3.353	21.245	19.329	23.283	21.736
Weight of bedload fractions	0.661	0.474	0.996	0.151	0.687	0.742	0.572	0.655

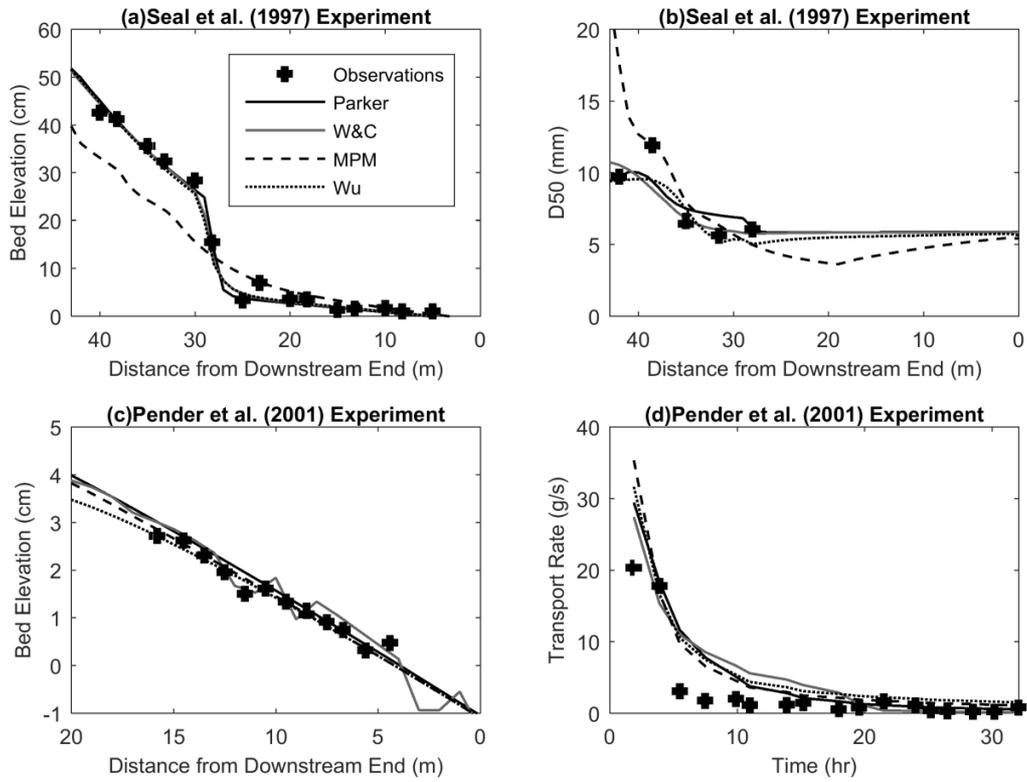


Fig. 17 Observations (crosses) and individual model outputs (lines) of (a) bed elevation profile and (b) D50 distribution at 20 hour of the Seal et al. (1997) experiment; (c) bed elevation profile at 32.1 hour and (d) sediment transport rate during the calibration period of the Pender et al. (2001) experiment.

4.4 Results and Analysis

4.4.1 Model Weights and Standard Deviations

The weights of the four models were determined based on their ability to reproduce the observations for the calibration period, and they varied significantly according to the types of data used for BMA modeling (Fig. 18). No individual model was always the best at all variables and cases.

In the Seal et al. (1997) experiment, the W&C model is the best for BMA using bed elevation data, and the Wu model is the best for BMA using D50 size data. The multivariate BMA model using both bed elevation and D50 data distribute the model weights in a compromise of the distributions from the two univariate BMA models. Specifically, MPM has no weight on bed elevation but it has a meaningful weight on D50 because a couple of D50 data points were captured by this model (Fig. 17). However, the multivariate BMA puts zero weight on MPM. It reflects the fact that the MPM produced worse output profiles for both bed elevation and the D50 distribution than the other models. This result is expected because the range of material sizes used in this experimental case was wider than the sizes used for developing MPM model. In the Pender et al. (2001) experiment, MPM had a weight near one for bed elevation, and Parker was the best for sediment transport rate data. The weights for the multivariate BMA model were also distributed in a compromise of the two univariate BMA models.

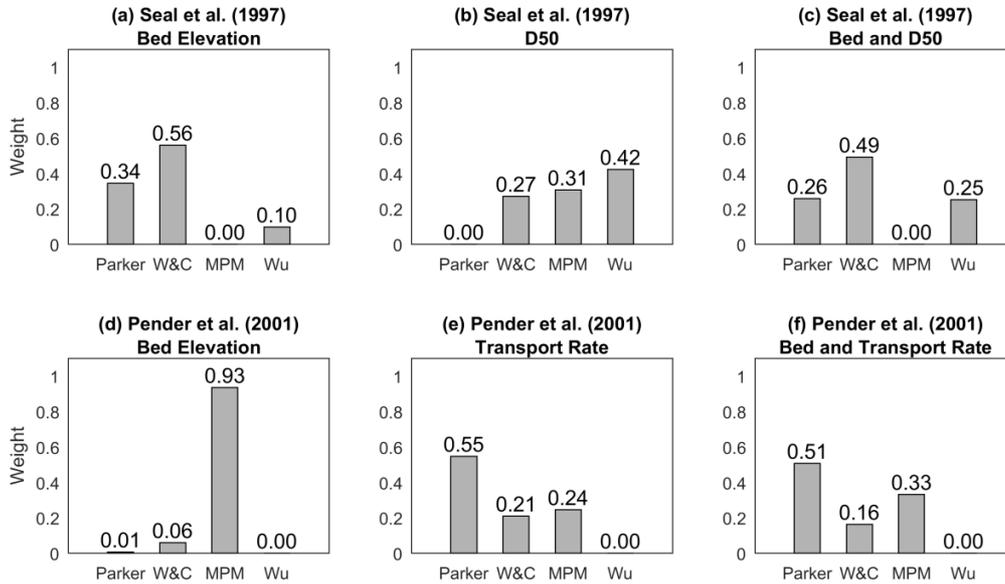


Fig. 18 Weights of individual models determined by BMA using each type of variable from the two experiments.

The standard deviations of each model prediction were determined generally following the reverse order of the model weights (Table 8), implying that predictions from models with low weights contain large uncertainties. The multivariate BMA model was able to specify the standard deviations for both bed elevation and D50 in the Seal et al. (1997) case and for both bed elevation and sediment transport rate in the Pender et al. (2001) case. In addition, the standard deviation for each variable from the multivariate BMA model are similar to the values from each of univariate BMA models.

To evaluate the assumption in which BMA applies the same weights and standard deviations at all times, the observations at a given time were grouped for each variable and the model weights and standard deviations were specified for each of the aggregated dataset. For bed elevations

and D50 sizes in both flume cases, the weights changed every time step with fluctuating patterns, and the average of those fluctuated values followed the weights determined by grouping the observations from all the different times as a single variable. The same patterns occurred for the standard deviations of the individual models, so the assumption is expected to be reasonable for those two variables. However, the sediment transport rates in the Pender et al. (2001) case showed standard deviations that varied following the scales of the observations and model outputs. It suggests that a different approach is required to specify the standard deviations for the variables where such behavior is expected.

4.4.2 BMA Predictions and Uncertainty

The predictive distributions of the BMA models are illustrated using the predictions of bed elevation at one location and time in the forecast period of the Pender et al (2001) case (Fig. 19). Each of the competing models provided a single prediction (white markers), and the deterministic BMA predictions (color filled markers) are the same as a weighted-average of the predictions from the individual models. Both the BMA using bed elevation and the multivariate BMA model generated probabilistic distributions of their predictions, but the prediction from the BMA model using transport rate data was not able to produce the distribution because the standard deviations for model predictions of bed elevation were not specified. The 90% credible intervals (CI), which are expected to include 90% of the data, can be determined by calculating the 5% and 95% quantiles from the BMA pdf at every prediction point.

Table 8 Standard Deviations of Each Model Prediction Obtained from BMA Calibration: Multivariate BMA Model Uses Bed Elevation and D50 Data for the Seal et al. (1997) Case and Bed Elevation and Sediment Transport Rate for the Pender et al. (2001) Case.

	Bed Elevation (cm)				D50 (mm) / Transport Rate (g/s)				
	Parker	W&C	MPM	Wu	Parker	W&C	MPM	Wu	
Seal et al. (1997)									
Bed Elevation BMA	2.93	0.74	12.15	2.20	-	-	-	-	
D50 BMA	-	-	-	-	1.25	0.77	2.30	0.61	
Multivariate BMA	3.05	0.69	12.20	2.12	3.35	0.76	2.47	0.85	
Pender et al. (2001)									
Bed Elevation BMA	0.09	0.13	0.23	0.10	-	-	-	-	
Transport Rate BMA	-	-	-	-	0.54	6.01	3.15	5.84	
Multivariate BMA	0.19	0.15	0.28	0.10	0.53	6.14	3.17	6.57	

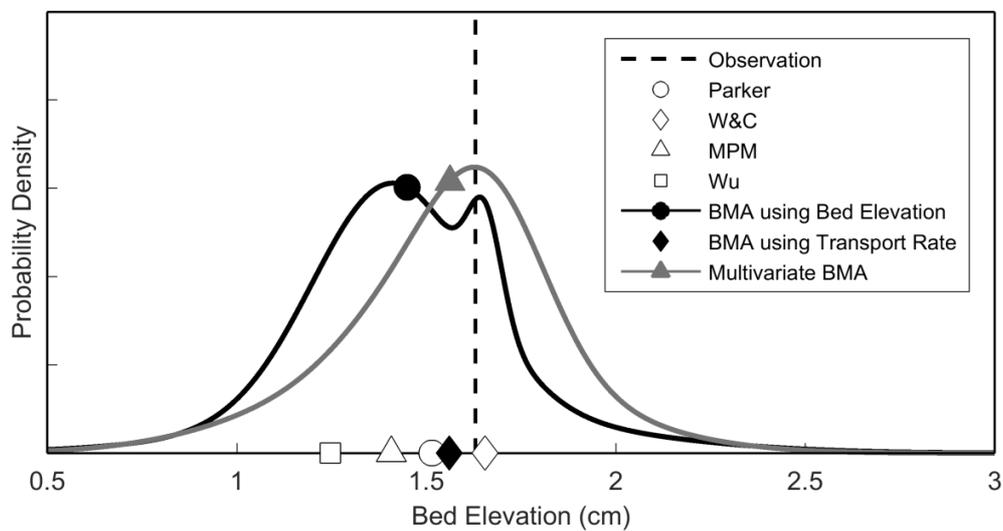


Fig. 19 Observation (dashed vertical line), and predictions from the individual models (white markers) and the BMA model using sediment transport rate (black marker), predictive distributions of BMA using bed elevation (solid black line with filled circle marker) and multivariate BMA model (solid grey line with filled triangle marker) for bed elevation at 10.5 m upstream at 79.6 hour in the Pender et al. (2001) experiment.

In order to examine whether the BMA models defined using calibration data also apply to the forecast period, the predictions and their CIs for the two experimental cases are presented in Fig. 20 and Fig. 21. Overall, the deterministic predictions differ between the BMA models, and the multivariate BMA model was able to produce the CIs for all considered variables at the same time whereas the univariate BMA models provided the CIs only for the considered single variable. For the Seal et al. (1997) depositional case (Fig. 20), the BMA model using D50 data generated lower bed elevation profiles and larger D50 sizes near the upstream end of the flume compared to the other BMA models. It can be interpreted that the amount of deposited material is underestimated and the predicted particle sizes are larger compared to the observed data. For the Pender et al. (2001) erosional case (Fig. 21), the predictions of bed elevation from all the models produce similar profiles, but the transport rates were overestimated by the BMA model using bed elevation. The predictive CIs from both the BMA model using transport rate data and the multivariate BMA covered much wider ranges of transport rate compared to the scale of observed values. Moreover, the lower bounds of the CIs for transport rates were below zero, which is not realistic. The results reinforce the idea that the scale of the uncertainty in the predictions should change according to the predicted values, and this approach is explored in the next section.

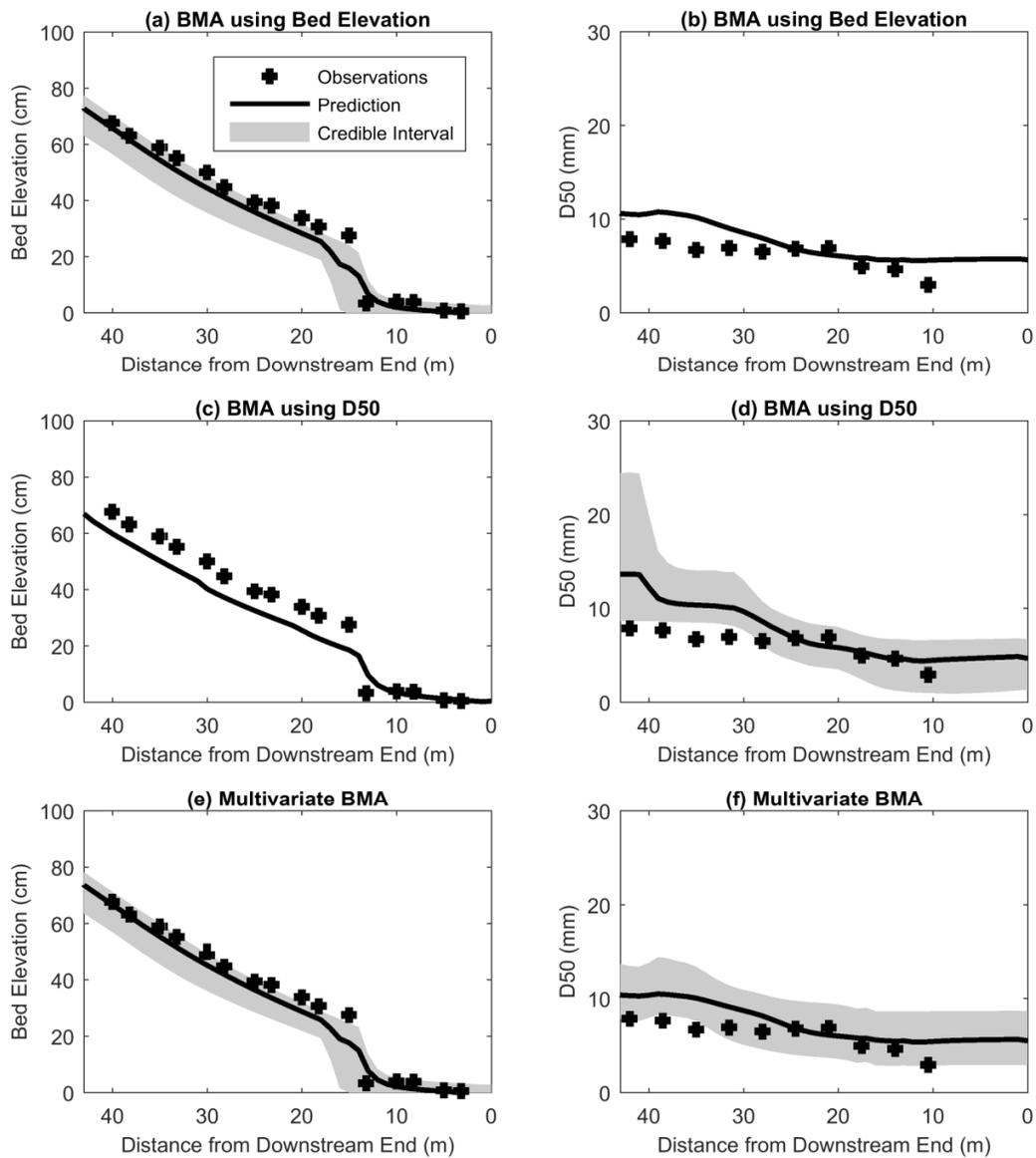


Fig. 20 Observations, predictions and 90% credible intervals from each type of BMA model for bed elevations at 50 hr and D50 at 53 hour in the forecast period of the Seal et al. (1997) experiment.

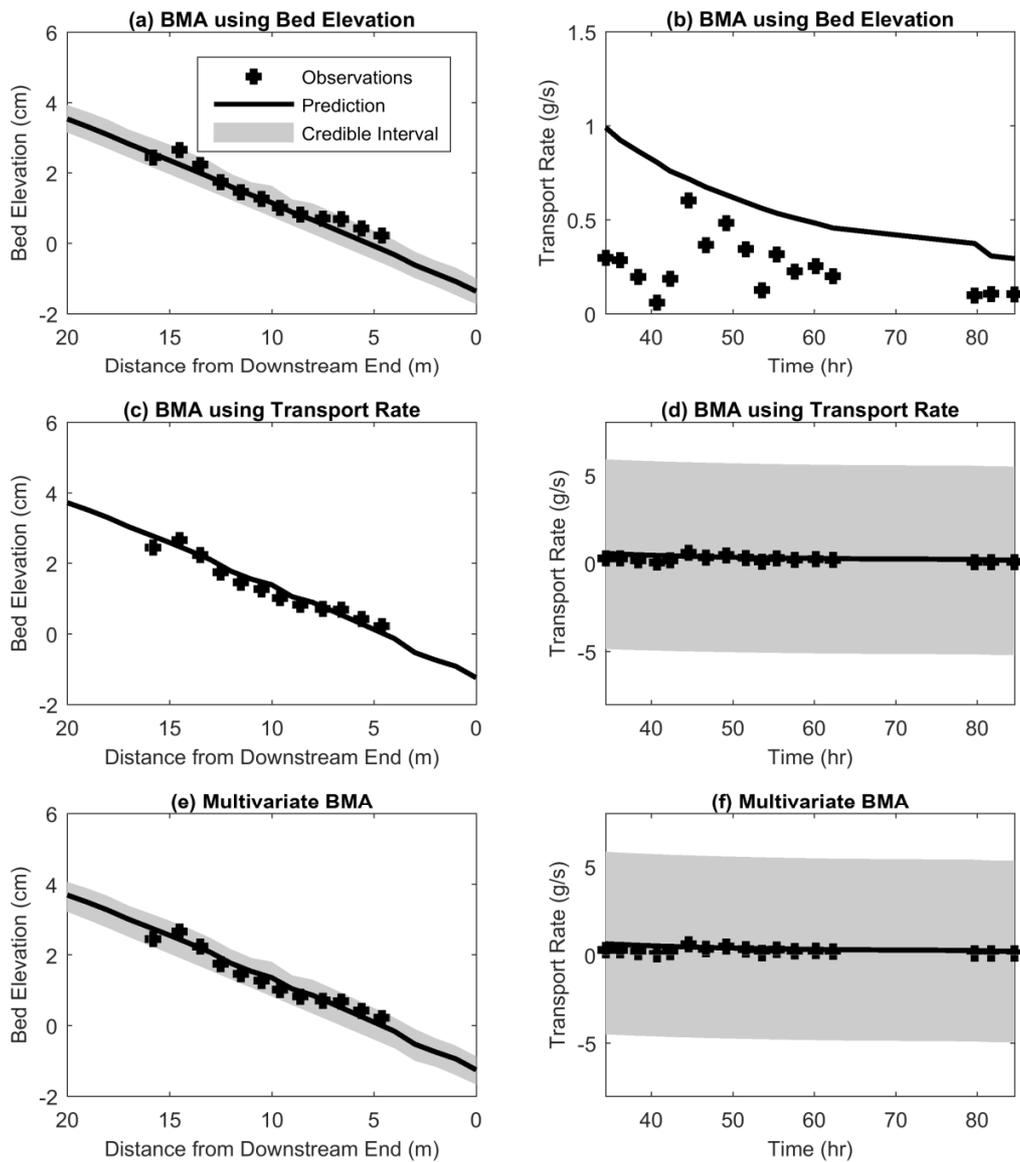


Fig. 21 Observations, predictions and 90% credible intervals from each type of BMA model for bed elevations at 84.6 hr and sediment transport rate during the forecast period of the Pender et al. (2001) experiment.

The model performance was assessed using the Nash-Sutcliffe Coefficient of Efficiency (NSCE), which computes the accuracy of the deterministic predictions from each model. The value of NSCE can range from $-\infty$ to one, and it approaches one when the model reproduces the observations perfectly. For each variable, the NSCE values were calculated for the individual and BMA models (Table 9). For the Seal et al. (1997) case, the NSCE indicated that the multivariate BMA models provided the best predictions for bed elevations in both calibration and forecast periods. Specifically, the BMA using bed elevations and the multivariate BMA had same NSCE in the calibration period, but the multivariate model showed better performance in forecast period. The BMA using D50 data provided the worst forecasts for D50 even though it was the best model in the calibration period. This might result from the fact that MPM gained some weight for the BMA model using D50 (Fig. 18) because MPM captured some observed points of D50 by chance although it produced the worst profiles for both bed elevation and D50 sizes with the lowest NSCE in calibration period (Fig. 17). On the other hand, the performance of the multivariate BMA model, which ignored MPM by considering all available calibration data, is superior to that of the two univariate BMA models. In the Pender et al. (2001) case, the three BMA models provided better estimations for bed elevations than the individual models. However, the BMA model defined by bed elevation had a large negative NSCE, which indicates poor predictions for transport rate in the forecast period. It can be concluded that if the BMA model is developed using only a single variable (limited calibration data), it might provide poor estimates for the forecast period for either its own variable or other variables.

Table 9 Nash-Sutcliffe Coefficient of Efficiency (NSCE) Values for Individual Models and BMA Models from Calibration (Calib.) and forecast (Fore.) Periods of Two Experiments.

Model	Seal et al. (1997)				Pender et al. (2001)			
	Bed Elevation		D50		Bed Elevation		Transport Rate	
	Calib.	Fore.	Calib.	Fore.	Calib.	Fore.	Calib.	Fore.
Parker	0.99	0.81	0.36	-0.99	0.91	0.91	0.63	-0.53
W&C	0.98	0.98	0.43	-0.72	0.89	0.89	0.59	-1.36
MPM	0.84	0.50	0.22	-14.89	0.93	0.92	0.45	-9.47
Wu	0.98	0.97	0.47	-0.24	0.91	0.84	0.54	-34.44
Bed Elevation BMA	0.99	0.96	0.42	-0.70	0.93	0.93	0.47	-7.89
D50 BMA	0.97	0.94	0.53	-2.54	-	-	-	-
Transport Rate BMA	-	-	-	-	0.92	0.93	0.60	-0.80
Multivariate BMA	0.99	0.98	0.43	-0.57	0.92	0.94	0.59	-1.42

The model weights did not exactly follow the order of NSCE, but the BMA models are typically expected to give higher weights to the better performing models. Specifically, the BMA model is specified by optimizing the combination of weights and standard deviations of the competing models so that the single model weight is highly correlated to the weights of the other models and the standard deviations. For example, the Wu model had a zero weight for all BMA models for the Pender et al. (2001) case although the NSCE of this model was not the worst among the competing models for both bed elevations and transport rates in the calibration period. Also, the BMA prediction is located on the expectation of the BMA pdf not the highest probability point. Tests using the mode of the BMA pdf showed that the predictions from the BMA model were generally consistent with the predictions from the individual model with the largest weight, and the model performances were not better than the predictions using a weighted-average. In addition, the BMA pdf had a multi-modal distribution when the predictions from competing

models were widely spread, like the BMA model using bed elevation in Fig. 19. Therefore, the expectation of the BMA pdf is expected to provide a reasonable BMA prediction although it does not possess the highest probability.

The percentages of observations covered by the 90 % CIs of each BMA model are shown in Table 10. The multivariate BMA models cover approximately 90% of the observations in the calibration period of both experimental cases, and they generally covered more observations than the univariate BMA models for the considered variables. Specifically, the multivariate BMA model covers more than 90 % of the observations for all considered variables in the Pender et al. (2001) case. For the Seal et al. (1997) case, only 77 % of the observed bed elevations in the forecast period were covered by the multivariate BMA CIs. Despite missing the observations more than expected, the deterministic predictions for the bed elevations are outstanding because its NSCE value was 0.98. This can be interpreted that the uncertainty in those predictions was underestimated by the multivariate BMA model. The 90 % CIs for the D50 sizes from the multivariate BMA model covered more observations with narrower widths compared to the univariate BMA model using D50 data.

Table 10 Percentage of Observations Covered by 90% Credible Intervals of BMA Models from Calibration and Forecast Periods of Seal et al. (1997) and Pender et al. (2001) Experiments.

Model	Seal et al. (1997)				Pender et al. (2001)			
	Bed Elevation		D50		Bed Elevation		Transport Rate	
	Calib.	Fore.	Calib.	Fore.	Calib.	Fore.	Calib.	Fore.
Bed Elevation BMA	89.6	70.1	-	-	88.7	91.1	-	-
D50 / Transport Rate BMA	-	-	84.7	66.7	-	-	82.4	100
Multivariate BMA	88.9	77.0	89.1	83.3	89.3	94.1	82.4	100

4.4.3 Applying Coefficient of Variation

In order to resolve the problem of overestimation in the uncertainty for the Pender et al. (2001) case, the coefficients of variation were applied to the sediment transport rates whereas the standard deviations were applied to the bed elevations. Table 11 presents the calibration results of the multivariate BMA modeling. Compared to the multivariate BMA using standard deviations for both variables (Fig. 18), the Parker and W&C models gained larger weights and the weight of MPM reduced remarkably. Standard deviations of individual model predictions for bed elevations were not much different compared to Table 8. The coefficients of variation for transport rates follow the reverse order of the model weights. Fig. 22 compares the predictions from the multivariate BMA models for sediment transport rate in both the calibration and forecast periods. By applying the coefficients of variation, the BMA provided more realistic CIs, where the widths vary according to the scale of the predicted values. In addition, there was no negative value for the lower bounds of the CIs. The model performance was also improved as the NSCE values for transport rate increased (Table 12), and the BMA model covered all observations of transport rate from both the calibration and forecast periods.

Table 11 Multivariate BMA Calibration Results from Applying Standard Deviations (STDEV) for Bed Elevation and Coefficients of Variation (CV) for Sediment Transport Rate in the Pender et al. (2001) Case.

	Parker	W&C	MPM	Wu
Model Weight	0.77	0.20	0.03	0.00
STDEV for Bed Elevation	0.24	0.14	0.41	0.01
CV for Transport Rate	0.52	0.53	0.55	0.61

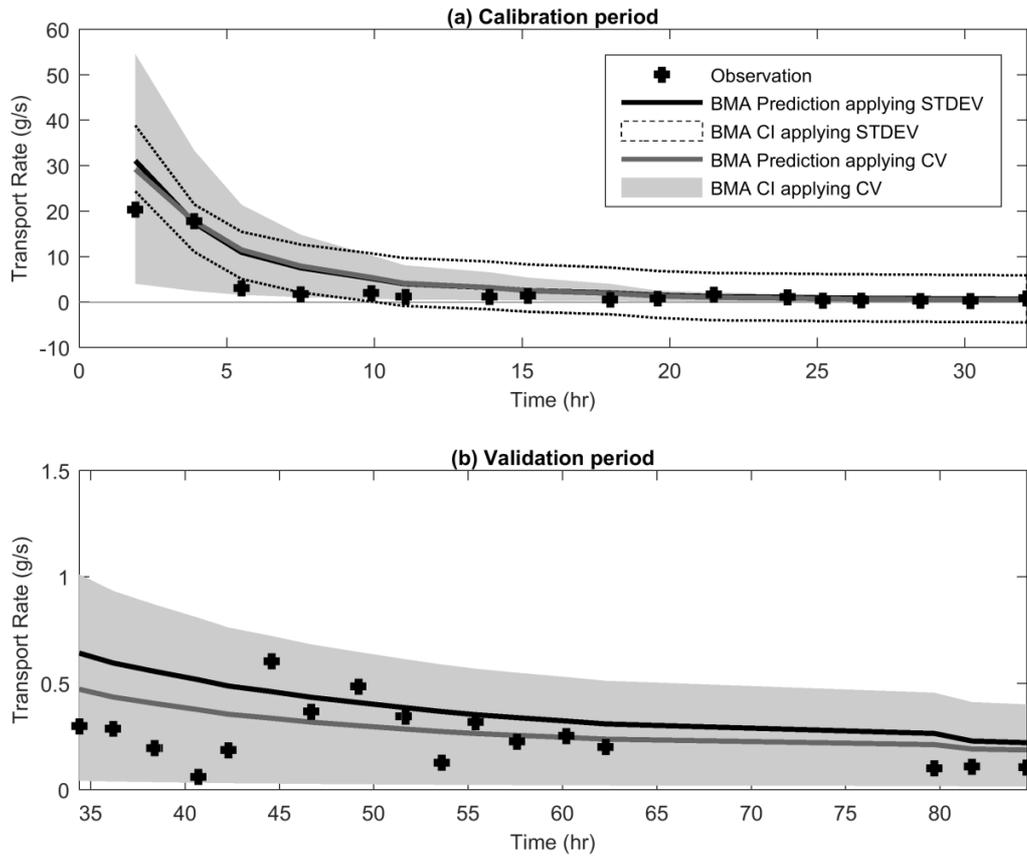


Fig. 22 Observations, predictions and 90% credible intervals for sediment transport rate from the multivariate BMA models applying standard deviations (STDEV) and coefficient of variations (CV) during both the calibration and forecast periods of the Pender et al. (2001) experiment.

Table 12 NSCE and Percentage of Observations Covered by the 90% Credible Intervals of the Multivariate BMA Model Applying CV from Calibration and Forecast Periods of the Pender et al. (2001) Case.

	Bed Elevation		Sediment Transport Rate	
	Calib.	Fore.	Calib.	Fore.
NSCE	0.918	0.922	0.629	-0.155
Coverage (%)	88.1	90.6	100	100

A gamma distribution, which would not allow the values to be negative, was also applied to the transport rate in order to evaluate the assumption of normality. To develop the BMA model, the shape parameter α and scale parameter β of the gamma distribution were determined for each individual model as well as the model weights. This approach showed the similar performance in prediction accuracy and observation coverage as the BMA model using a normal distribution, and the lower bounds of 90% CI were inherently non negative, which satisfies the basic characteristics of the considered variable. However, a couple of distinct weaknesses were found when applying the gamma distribution. It was not easy to implement because the parameters α and β vary every prediction point, and they were not able to be calibrated using the EM method so an alternative (more complex) optimization algorithm was required. In addition, the prediction from individual model, which is used as a mean of the distribution, cannot be treated as the most likely value because the mean and mode are not same in the gamma pdf whereas they are same for the normal pdf. Overall, the results strongly supported applying the coefficient of variation to the normal distributions of model predictions for the ratio-scaled variables in BMA modeling.

4.5 Conclusions

1. The proposed likelihood function enables BMA to assess model structure uncertainty in the predictions of multiple variables from sediment transport models within a single calibration whereas the previous BMA requires separate calibration for each of the considered variables. The multivariate BMA determines the weights and standard deviations of competing models based on the behavior of multiple variables, and it is also able to generate probabilistic predictions for multiple variables in the forecast scenario.

2. The multivariate BMA model provides better model performance than the univariate BMA models. Based on NSCE, the accuracy of the deterministic predictions is improved by considering multiple variables from the calibration dataset. Specifically, the D50 size distribution in the Seal et al. (1997) case was poorly estimated by the univariate BMA even though the model was developed using the D50 size data from the calibration period. This behavior results from MPM, which showed the worst performance for this case among the competing models but gained meaningful weight due to its match for a couple of D50 observations. On the other hand, the multivariate BMA model, which was developed using a larger dataset including both bed elevations and D50 sizes, was able to distinguish that MPM was worse than the other models, and then produced more accurate and reliable forecasts by assigning no weight to MPM. In the Pender et al. (2001) case, the univariate BMA model using the bed elevation data made poor forecasts for sediment transport rates, and the multivariate BMA produced the improved predictions because it was modeled using both bed elevations and transport rate data.
3. Probabilistic distributions of the predictions are also enhanced by using the multivariate BMA model. In the Seal et al. (1997) case, the multivariate BMA model generated narrower CIs, which represent the less uncertainty in predictions, and covered more observations for D50 sizes than the univariate BMA model developed using D50 size data. It can also be inferred that the credibility of the predictions increases by minimizing role of the MPM model in the multivariate BMA modeling. From the Pender et al. (2001) case, approximately 90 % of the observations for both variables were covered by the multivariate BMA CIs, which showed larger percentages than the CIs for the univariate models.

4. Applying coefficients of variation instead of standard deviation allows BMA to vary the uncertainty scale by reflecting the magnitude of model predictions for the variables. Several advantages of using this approach for the transport rates in the Pender et al. (2001) case were found including: (1) the CI widths changed according to the predicted values, (2) their lower bounds did not contain negative values, which is unrealistic, (3) the CIs cover more observations than the CIs determined by the standard deviations, and (4) the EM algorithm is easily applied to find the best values of the coefficient of variation.

Several notable avenues are available for future research. First, the results presented in this study were obtained from the case studies of flume experiments where the sediment transport behavior was well-controlled. In order to establish generality, the proposed method should be expanded to consider other cases that contain natural rivers where the erosion and sedimentation processes occur together and are more complex. Other available sediment transport equations such as Ackers and White (1973), Brownlie (1981), modified Laursen's formula (Madden 1993), and Yang (1996) could be considered. In addition, future research could consider which variables are appropriate to use standard deviations and coefficients of variation to describe their uncertainty. Second, in a natural river, it is expected that the most likely transport model would be different in space and time because the system configuration generally changes in time. The way to describe the variation of model weights needs to be studied in more detail. Third, the underestimates of uncertainty for predicting bed elevations in the Seal et al. (1997) case might result from the magnitude of uncertainty increasing with respect to the length of the forecast period. This issue could be explored. Fifth, the selection of a transport equation is not the only source of uncertainty in sediment transport modeling. Uncertainty might originate from the

errors included in the model parameter values, the data used for model forcing variables, the channel geometry information, and the observations (Ruark et al. 2011).

References

- Ackers, P., and White, W.R. (1973). "Sediment transport: new approach and analysis." *Journal of the Hydraulics Division*, 99(11), 2041-2060.
- Ahn, J., and Yang, C.T. (2015). "Determination of recovery factor for simulation of non-equilibrium sedimentation in reservoir." *International Journal of Sediment Research*, 30(1), 68-73.
- Ahn, J., Yang, C.T., Pridal, D.B., and Remus, J.I. (2013). "Numerical modeling of sediment flushing from Lewis and Clark Lake." *International Journal of Sediment Research*, 28(2), 182-193.
- Anderson, N.H. (1965). "Averaging versus adding as a stimulus-combination rule in impression formation." *Journal of Experimental Psychology*, 70(4), 394.
- Bagnold, R.A., and Barndorff-Nielsen, O. (1980). "The pattern of natural size distributions." *Sedimentology*, 27(2), 199-207.
- Bates, J.M., and Granger, C.W.J. (1969). "The combination of forecasts." *Operational Research*, 20(4), 451-468.
- Bertin, X., Fortunato, A.B., and Oliveira, A. (2007). "Sensitivity analysis of a morphodynamic modeling system applied to a Portuguese tidal inlet." *Proceedings of 5th IAHR Symposium on River, Coastal and Estuarine Morphodynamics*, 11-17.
- Beven, K.J. (2011). *Rainfall-Runoff Modelling: The Primer*. John Wiley & Sons.
- Beven, K.J., and Binley, A. (1992). "The future of distributed models: Model calibration and uncertainty prediction." *Hydrological Processes*, 6(3), 279-298.
- Brownlie, W.R. (1981). *Prediction of Flow Depth and Sediment Discharge in Open Channels*. Ph.D. dissertation, California Institute of Technology, Pasadena, CA.
- Buckland, S.T., Burnham, K.P., and Augustin, N.H. (1997). "Model selection: an integral part of inference." *Biometrics*, 53(2), 603-618.
- Burnham, K.P., and Anderson, D.R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd Edition*. Springer Science and Business Media.

- Chahinian, N., and Moussa, R. (2007). "Comparison of different multi-objective calibration criteria of a conceptual rainfall-runoff model of flood events." *Hydrology and Earth System Sciences Discussions*, 4(3), 1031-1067.
- Clyde, M.A. (1999). "Bayesian model averaging and model search strategies." *Bayesian Statistics* 6, 157-185.
- Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society: Series B*, 39(1), 1-38.
- Downey, C., and Sanner, S. (2010). "Temporal difference Bayesian model averaging: A Bayesian perspective on adapting lambda." *Proceedings of the 27th International Conference on Machine Learning*, ICML-10, 311-318.
- Einstein, H.A. (1950). *The bed-load function for sediment transportation in open channel flows*. U.S. Department of Agriculture, Soil Conservation Service, Technical Bulletin No. 1026.
- Engelund, F., and Hansen, E. (1972). *A Monograph on Sediment Transport in Alluvial Streams*. Teknisk Forlag, Technical Press.
- Gaeuman, D., Andrews, E.D., Krause, A., and Smith, W. (2009). "Predicting fractional bed load transport rates: Application of the Wilcock-Crowe equations to a regulated gravel bed river." *Water Resources Research*, 45, W06409, doi:10.1029/2008WR007320.
- Givens, G.H., and Hoeting, J.A. (2012). *Computational statistic, 2nd Edition*. John Wiley & Sons.
- Granger, C.W., and Ramanathan, R. (1984). "Improved methods of combining forecasts." *Journal of Forecasting*, 3(2), 197-204.
- Hansen, B.E. (2007). "Leats Squares Model Averaging." *Econometrica*, 75(4), 1175-1189.
- Hoeting, J.A., Madigan, D., Raftery, A.E., and Volinsky, C.T. (1999). "Bayesian model averaging: a tutorial." *Statistical Science*, 14(4), 382-401.
- Huang, J.V., and Greimann, B.P. (2010). "Sediment transport and channel morphology model in the San Joaquin River from Friant Dam to Mendota Dam, California." *Proceedings of 2nd Joint Federal Interagency Conference*, 27.
- Huang, J.V., and Greimann, B.P. (2013). *User's Manual for SRH-1D 3.0*. Bureau of Reclamation, U.S. Department of the Interior, Denver.
- Kleiber, W., Raftery, A.E., and Gneiting, T. (2011). "Geostatistical model averaging for locally calibrated probabilistic quantitative precipitation forecasting." *Journal of the American Statistical Association*, 106(496), 1291-1303.
- Laursen, E.M. (1958). "The total sediment load of streams." *Journal of the Hydraulics Division*, 84(1), 1531-1536.

- Madden, E.B. (1993). *Modified Laursen Method for Estimating Bed-material Sediment Load (Contract Report HL-93-3)*. U.S. Army Engineer Waterways Experiment Station, U.S. Army Corps of Engineers.
- Meyer-Peter, E., and Müller, R. (1948). "Formula for bed-load transport." *Proceedings of 2nd International Association for Hydraulic Research*, 39-64.
- Mo, X., and Beven, K. (2004). "Multi-objective parameter conditioning of a three-source wheat canopy model." *Agricultural and Forest Meteorology*, 122(1), 39-63.
- Nash, J.E., and Sutcliffe, J.V. (1970). "River flow forecasting through conceptual models part I—A discussion of principles." *Journal of hydrology*, 10(3), 282-290.
- Parker, G. (1990). "Surface-based bedload transport relation for gravel rivers." *Journal of Hydraulic Research*, 28(4), 417-436.
- Parrish, M.A., Moradkhani, H., and DeChant, C.M. (2012). "Toward reduction of model uncertainty: Integration of Bayesian Model averaging and data assimilation." *Water Resources Research*, 48, W03519, doi:10.1029/2011WR011116.
- Pender, G., Hoey, T.B., Fuller, C., and Mcewan, I.K. (2001). "Selective bedload transport during the degradation of a well sorted graded sediment bed." *Journal of Hydraulic Research*, 39(3), 269-277.
- Pinto, L., Fortunato, A.B., and Freire, P. (2006). "Sensitivity analysis of non-cohesive sediment transport formulae." *Continental Shelf Research*, 26(15), 1826-1839.
- Raftery, A.E., Gneiting, T., Balabdaoui, F., and Polakowski, M. (2005). "Using Bayesian model averaging to calibrate forecast ensembles." *Monthly Weather Review*, 133(5), 1155-1174.
- Ribberink, J. S. (1998). "Bed-load transport for steady flows and unsteady oscillatory flows." *Coastal Engineering*, 34(1), 59-82.
- Rottner, J. (1959). "A formula for bed load transportation." *La Houille Blanche*, 14(3), 285-307.
- Ruark, M.D., Niemann, J.D., Greimann, B.P., and Arabi, M. (2011). "Method for assessing impacts of parameter uncertainty in sediment transport modeling applications." *Journal of Hydraulic Engineering*, 137(6), 623-636.
- Russell, K., Greimann, B.P., Cluer, B., Hepler, T., King, D., O'Meara, S., Simon, A., Godaire, J., and Salas, D. (2010). "Klamath reservoir sediment characterization and drawdown impacts for the dam removal investigation." *Proceedings of 2010 AGU Fall Meeting*, 1053.
- Sabatine, S.M., Niemann, J.D., and Greimann, B.P. (2015). "Evaluation of parameter and model uncertainty in simple applications of a 1D sediment transport model." *Journal of Hydraulic Engineering*, 141(5), DOI: 10.1061/(ASCE)HY.1943-7900.0000992.
- Schmelter, M.L., Hooten, M.B., and Stevens, D.K. (2011). "Bayesian sediment transport model for unisize bed load." *Water Resources Research*, 47, W11514, doi:10.1029/2011WR010754.

- Seal, R., Paola, C., Parker, G., Southard, J.B., and Wilcock, P.R. (1997). "Experiments on downstream fining of gravel: I. Narrow-channel runs." *Journal of Hydraulic Engineering*, 123(10), 874-884.
- Sloughter, J.M., Gneiting, T., and Raftery, A.E. (2010). "Probabilistic wind speed forecasting using ensembles and Bayesian model averaging." *Journal of the American Statistical Association*, 105(489), 25-35.
- van Griensven, A., and Meixner, T. (2007). "A global and efficient multi-objective auto-calibration and uncertainty estimation method for water quality catchment models." *Journal of Hydroinformatics*, 9(4), 277-291.
- van Rijn, L.C. (1989). *Handbook sediment transport by currents and waves*. Delft Hydraulics Laboratory.
- Veenhuis, B. (2014). "A practical model blending technique based on Bayesian model averaging." *Proceedings of 22st Conference on Probability and Statistics in the Atmospheric Sciences*.
- Viallefont, V., Raftery, A.E., and Richardson, S. (2001). "Variable selection and Bayesian model averaging in case-control studies." *Statistics in Medicine*, 20(21), 3215-3230.
- Vrugt, J.A., and Robinson, B.A. (2007). "Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging." *Water Resources Research*, 43, W01411, doi:10.1029/2005WR004838.
- Vrugt, J.A., Ter Braak, C.J., Clark, M.P., Hyman, J.M., and Robinson, B.A. (2008). "Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation." *Water Resources Research*, 44, W00B09, doi:10.1029/2007WR006720.
- Wallingford, H.R. (1990). *Sediment Transport, the Ackers and White Theory Revised, Report SR237*, HR Wallingford, England.
- Wilcock, P.R. (2001). "Toward a practical method for estimating sediment-transport rates in gravel-bed rivers." *Earth Surface Processes and Landforms*, 26(13), 1395-1408.
- Wilcock, P.R., and Crowe, J.C. (2003). "Surface-based transport model for mixed-size sediment." *Journal of Hydraulic Engineering*, 129(2), 120-128.
- Wong, M., and Parker, G. (2006). "Reanalysis and correction of bed-load relation of Meyer-Peter and Müller using their own database." *Journal of Hydraulic Engineering*, 132(11), 1159-1168.
- Wu, W., Wang, S.S., and Jia, Y. (2000). "Nonuniform sediment transport in alluvial rivers." *Journal of Hydraulic Research*, 38(6), 427-434.
- Yang, C.T. (1973). "Incipient motion and sediment transport." *Journal of the Hydraulics Division*, 99(10), 1679-1704.

- Yang, C.T. (1979). "Unit stream power equations for total load." *Journal of Hydrology*, 40(1), 123-138.
- Yang, C.T. (1984). "Unit stream power equation for gravel." *Journal of Hydraulic Engineering*, 110(12), 1783-1797.
- Yang, C.T. (1996). *Sediment Transport: Theory and Practice*, McGraw-Hill Companies, Inc., New York, NY.
- Yapo, P.O., Gupta, H.V., and Sorooshian, S. (1998). "Multi-objective global optimization for hydrologic models." *Journal of Hydrology*, 204(1), 83-97.
- Ye, M., Neuman, S.P., and Meyer, P.D. (2004). "Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff." *Water Resources Research*, 40, W05113, doi:10.1029/2003WR002557.