DISSERTATION

ADVANCED BAYESIAN FRAMEWORK FOR UNCERTAINTY ESTIMATION OF SEDIMENT TRANSPORT MODELS

Submitted by

Jeffrey Youngjai Jung

Department of Civil and Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2018

Doctoral Committee:

Adivsor: Jeffrey D. Niemann

Blair P. Greimann
Pierre Y. Julien
Haonan Wang

ABSTRACT

ADVANCED BAYESIAN FRAMEWORK FOR

UNCERTAINTY ESTIMATION OF SEDIMENT TRANSPORT MODELS

Numerical sediment transport models are widely used to forecast the potential changes in rivers that might result from natural and/or human influences. Unfortunately, predictions from those models always possess uncertainty, so that engineers interpret the model results very conservatively, which can lead to expensive over-design of projects. The Bayesian inference paradigm provides a formal way to evaluate the uncertainty in model forecasts originating from uncertain model elements. However, existing Bayesian methods have rarely been used for sediment transport models because they often have large computational times. In addition, past research has not sufficiently addressed ways to treat the uncertainty associated with diverse sediment transport variables. To resolve those limitations, this study establishes a formal and efficient Bayesian framework to assess uncertainty in the predictions from sediment transport models. Throughout this dissertation, new methodologies are developed to represent each of three main uncertainty sources including poorly specified model parameter values, measurement errors contained in the model input data, and imperfect sediment transport equations used in the model structure. The new methods characterize how those uncertain elements affect the model predictions. First, a new algorithm is developed to estimate the parameter uncertainty and its contribution to prediction uncertainty using fewer model simulations. Second, the uncertainties of various input data are described using simple error equations and evaluated within the parameter estimation framework. Lastly, an existing method that can assess the uncertainty

related to the selection and application of a transport equation is modified to enable consideration of multiple model output variables. The new methodologies are tested with a one-dimensional sediment transport model that simulates flume experiments and a natural river. Overall, the results show that the new approaches can reduce the computational time about 16% to 55% and produce more accurate estimates (e.g., prediction ranges can cover about 6% to 46% more of the available observations) compared to existing Bayesian methods. Thus, this research enhances the applicability of Bayesian inference for sediment transport modeling. In addition, this study provides several avenues to improve the reliability of the uncertainty estimates, which can help guide interpretation of model results and strategies to reduce prediction uncertainty.

ACKNOWLEDGEMENTS

For the last six years, I have received significant help from numerous individuals and institutions in completing my degree, and I am pleased to recognize their support.

I would like to express my sincere appreciation to Dr. Jeffrey Niemann, my doctoral advisor, for his excellent guidance, thoughtful encouragement, generous financial support, and great patience. He is a great scholar who helps refine my ideas and continues to be a source of inspiration for me. I am also grateful to Dr. Blair Greimann, who has cooperated closely with my research. As a supervisor of my project work, he has provided invaluable advice to improve the practicality of my research as well as financial support. I also appreciate Dr. Julien and Dr. Wang, who have broadened the scope of my research and have provided important advice to my dissertation. Finally, thanks to all the people who have been together in A5.

My research has been gratefully sponsored by the Department of Civil and Environmental Engineering, Colorado State University and the U.S. Bureau of Reclamation Science and Technology Program under Project 1596 (R12AC80251) "Design and Testing of Computationally-efficient Methods to Evaluate Parameter and Model Uncertainties."

My family was also a source of encouragement. Jongkook, Sooyeul, Ilwon, and Wonkyung, who are my parents always supported me and believed that I would successfully finish my graduate studies. Lastly, I would love to show my biggest gratitude to my wife Johanna, who is an amazing life partner and my daughter Riley, who always makes me smile.

Thank you very much.

Jeffrey Youngjai Jung

DEDICATION

To Johanna and Riley.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Background

Sediment transport in rivers plays an important role in changing a river morphology, which can affect both infrastructure and water resources systems. Numerous scientists, researchers, and engineers have developed computational models to simulate river system behavior based on mathematical equations that represent the processes of flow and sediment transport. In the past three decades, numerical hydraulic and sediment transport models have been widely used to forecast the potential changes in rivers that might result from climate change and/or human activities. Several commercial and noncommercial releases of the computational models are currently available including: CCHE1D [Vieira and Wu 2002], CCHE2D [Jia and Wang 2001], River2D [Steffler and Blackburn 2002], FaSTMECH [Nelson 2016], SToRM [Simões 2009], Delft3D [Deltares 2014], Nays2DH [Shimizu and Takebayashi 2014], HEC-RAS [Brunner 2016], SRH-1D [Huang and Greimann 2013], and SRH-2D [Lai 2016].

Predictions from those numerical models always possess uncertainty from three sources. First, numerical models apply several simplifications and assumptions to describe the relevant processes using mathematical equations, so no model can perfectly represent the natural system (model structure uncertainty). In addition, a single sediment transport model usually contains several formulas that can compute the transport capacity, and these formulas were empirically developed for certain sediment sizes and flow conditions. Thus, the predictions based on a single model (and/or equation) are inherently uncertain [Wilcock 2001]. Second, numerical models contain parameters that need to be determined by the modeler, and many of those parameters are either difficult or impossible to measure and need to be calibrated (parameter

uncertainty) [Vrugt et al. 2003]. Hence, any errors in the values assigned to the parameters can result in errors in the model predictions. Third, the data used for model inputs are also uncertain because they inherently include measurement errors (input uncertainty). The errors in input data can produce inaccurate predictions even if a suitable model is selected and appropriate parameter values are assigned. In addition, they can also affect the estimation of the uncertain model parameters [Ajami et al. 2007].

Past research has noted the importance of these issues but has not formally identified the impact of those uncertainties. Specifically, most applications have typically evaluated those uncertainties by examining how model predictions spread when constitutive model equations, parameter values, and model input values are varied [Pinto et al. 2006; Bertin et al. 2007; Lai and Greimann 2010]. That approach is relatively simple to implement, but it has the potential to produce inaccurate uncertainty estimates because the variations in the uncertain model elements are informally specified. The informal uncertainty assessment also requires the modeler to interpret the results very conservatively because the model predictions are often used in situations involving potential economic loss, ecological impacts, and/or risks to human health. Such conservatism might lead to over-designed projects, which are expensive and can increase the conflicts between the competing objectives of the project (e.g., endangered species protection, agriculture, infrastructure management, and water quality treatment).

Bayesian inference provides a formal way to assess the uncertainty in the model predictions [Kuczera et al. 2006]. This paradigm represents the uncertainties in model elements using probability distributions. Specifically, Bayesian methods vary the uncertain model elements using distributions that are determined based on the similarity between the corresponding model outputs and calibration data [Green 2001]. The methods then quantify how the distributions of

the uncertain model elements propagate to the model predictions, which describe the prediction uncertainty. Bayesian methods have been increasingly used for models of hydrology, ecology, meteorology, and environmental science in the past two decades [Vrugt et al. 2008; Cressie et al. 2009; Smith et al. 2009; Renard et al. 2010; Wikle and Hooten 2010]. Furthermore, probabilistic forecasts from Bayesian methods can provide significant benefits over the informal uncertainty estimation approaches [Laloy and Vrugt 2012].

Few applications have used Bayesian inference with sediment transport models [Kanso et al. 2005; Wu and Chen 2009; Ruark et al. 2011; Sabatine et al. 2015; Cho et al. 2016]. The limited use of this methodology is largely due to its huge computational demands. Specifically, Bayesian methods often use a large number of model simulations to determine the probability distributions of uncertain elements. Moreover, hydraulic and sediment transport models generally require a long computation time for each simulation depending on the complexity of case considered (e.g., the number of dimensions, the spatial extent and time period that are simulated, and the spatial and temporal resolutions). The sparse use of Bayesian methods is also due to several difficulties in applying Bayesian methods to sediment transport problems [Wu and Chen 2009]. For example, input variables used in sediment transport models (e.g., channel geometry, sediment sizes, and water depth) are obtained using diverse measurement techniques [Bunte and Abt 2005], and their potential errors have not been addressed in the context of the Bayesian uncertainty paradigm [Schmelter et al. 2015]. In addition, Bayesian methods are often limited to a single variable of interest when calculating the similarity between the model results and the calibration data, whereas sediment transport models generate multiple output variables of interest and most applications consider those variables together [Sabatine et al. 2015].

1.2 Objectives

The overall goal of this research is to establish a formal and efficient Bayesian framework to assess uncertainty in the predictions from hydraulic and sediment transport models. This research aims to develop methodologies that retain the formality of the Bayesian approach but require few enough model simulations so that the methodologies can help guide interpretation of model results and strategies to reduce prediction uncertainty. The new methods characterize how the uncertainties from model parameter values, input data, and the model's mathematical structure affect the predictions from a sediment transport model. Those methods are developed with application to the Sedimentation and River Hydraulics - One Dimension (SRH-1D) model [Huang and Greimann 2013], which was developed and has been extensively used by the U.S. Bureau of Reclamation to predict impacts of potential river restoration activities. However, the methods are transferrable to other types of models.

This dissertation comprises three individual papers that accomplished each of three specific objectives in this research, which are briefly addressed below:

(1) Suggest a new method that can estimate the uncertainty in model parameters and its contribution to prediction uncertainty using fewer model simulations. In order to reduce the number of simulations required for specifying parameter probability distributions, the new algorithm is designed to replicate parameter sets that have already been used for model simulations in the Bayesian framework, instead of generating many new but similar parameter sets that require additional simulations. This new method can improve the efficiency of uncertainty estimation so that the computational time of implementing Bayesian inference is more affordable for complex models that demand a long simulation time.

(2) Develop simple error equations for the input data of a sediment transport model and integrate them into an existing Bayesian method of parameter estimation. Adopting an input error model originally proposed for hydrologic modeling [Ajami et al. 2007], input errors are characterized using Gaussian distributions for data such as flow discharges, river topography, and controlled water surface elevations used in a sediment transport model. 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 will allow a modeler to identify the contribution of each uncertain input to the overall uncertainty in the predictions, which can suggest strategies to reduce the uncertainty and improve reliability in the model predictions.

(3) Establish a multivariate version of Bayesian model averaging (BMA) to assess the uncertainty related to the selection and application of a transport equation in sediment transport models. BMA [Raftery et al. 2005] can reduce the effects of imperfections in a single model prediction by combining the predictions from a set of competing equations. The method provides a forecast along with its credible interval to characterize the uncertainty, but it is presently limited to cases that consider a single output variable. To overcome this limitation, the existing BMA method is modified to enable consideration of multiple model output variables and to allow the uncertainty associated with each transport equation to vary with the magnitude of the variables as needed. The multivariate BMA method will be able to generate probabilistic forecasts for multiple variables, which might also provide more accurate estimates.

CHAPTER 2
AN EFFICIENT ALGORITHM TO ESTIMATE PARAMETER UNCERTAINTY WITH APPLICATION TO SEDIMENT TRANSPORT MODELING

2.1 Introduction

Numerical models are widely used to predict the behavior of hydrologic, hydraulic, and water resources systems for conditions that cannot be observed directly. The predictions from these models always possess uncertainty, and one major source of uncertainty is the model parameter values. All numerical models contain parameters that are either difficult or impossible to measure directly. Such parameters are usually calibrated so that the model reproduces the observed system behavior, but no single set of parameter values is expected to perfectly represent the natural system. Thus, it is essential to understand the uncertainty in the parameter values and account for the associated uncertainty in the model predictions.

The Bayesian framework offers a formal way to estimate parameter uncertainty and its impact on model predictions [Clyde and George 2004; Kuczera et al. 2006]. This paradigm treats the uncertain parameters as random variables and represents the uncertainty in the parameter values using a joint probability density function (PDF). A joint prior PDF is specified by the modeler and describes the uncertainty in the parameters before model calibration. Then, the joint posterior PDF is determined based on the likelihood of each parameter set being correct, where the likelihood is computed by comparing the calibration data to the model outputs when each parameter set is used [Christensen et al. 2011]. Bayesian methods then quantify how the posterior PDF of the parameters propagates to the model forecasts. This Bayesian approach has been applied to various hydrologic and water resources modeling cases including: rainfall-runoff estimation [Ajami et al. 2007; Thyer et al. 2009; Sun and Bertrand-Krajewski 2013; Tramblay et

al. 2016], soil moisture assessment [Tolson and Shoemaker 2008; Shen et al. 2012], water quality studies [Kanso et al. 2005; Zheng and Keller 2007], and groundwater modeling [Hassan et al. 2008; Wu et al. 2014].

Generalized Likelihood Uncertainty Estimation (GLUE) [Beven and Binley 1992] is among the earliest and most commonly used Bayesian methods in hydrology. GLUE generates a large sample of independent parameter sets by randomly sampling a prior uniform distribution within specified parameter ranges and implements each parameter set for a calibration period simulation. The resulting likelihoods are used to obtain the marginal posterior distribution for each parameter. Parameter sets are then randomly sampled from the marginal posterior distributions and used for the forecast period to characterize the prediction uncertainty. A key limitation of GLUE is that many model simulations are required to obtain results. For example, case studies with hydrologic models used 50,000 ~ 100,000 model runs, depending on the number of parameters considered [Beven and Freer 2001; Blazkova et al. 2002; Jia and Culver 2006]. Performing large numbers of simulations can be problematic for complex models where each model run is time consuming. For instance, a three-dimensional sediment transport model took 30 min to complete a single steady-state simulation of flow and sediment transport near an intake facility where the simulation domain was about 7,000 m^2 [Ruether et al. 2005]. This case is relatively simple, but implementation of 10,000 parameter sets would require a continuous computation time of 200 days. GLUE is inefficient because the random sampling generates many parameter sets that require simulations but have extremely low likelihoods and contribute little to the uncertainty estimation [van Griensven and Meixner 2007; Blasone et al. 2008]. In addition, GLUE also neglects the correlation between the uncertain parameters because it produces only the marginal parameter distributions. Parameters in numerical models are often

highly correlated, so the uncertainty can be less than GLUE implies [Li and Vu 2013]. Many studies have shown that uncertainty estimates are considerably different when the correlations between the uncertain parameters are included [Vrugt et al. 2003; Wu and Chen 2009; Capaldi et al. 2012; Sabatine et al. 2015].

Three general strategies have been suggested to reduce the computational demands of GLUE. First, parallel processing can be applied because each parameter set is generated independently [Brazier et al. 2000; Freer et al. 2004]. With this approach, the efficiency of the method remains unchanged, but the continuous computation time decreases as the number of available processors increases. Second, sensitivity tests can be used to reduce the number of parameters included in GLUE [Tolson and Shoemaker 2008]. Neglecting less important parameters reduces the dimensionality of the parameter space and allows smaller samples to cover that space. However, sensitivity tests for high-dimensional cases can also require many model simulations. For example, Zak and Beven [1999] used 60,000 model runs for the sensitivity test and another 60,000 simulations for GLUE, and Athira and Sudheer [2015] used 28,000 simulations to reduce the number of parameters considered in SWAT from 13 to 4. Third, Latin Hypercube sampling (LHS) can be applied instead of random sampling to explore the parameter space [Uhlenbrook and Sieber 2005]. LHS has been shown improve the efficiency of GLUE for various case studies [Abbaspour et al. 2006; van Griensven et al. 2006; Ficklin et al. 2013], but these improvements are not large enough to greatly expand the applicability of GLUE.

Markov Chain Monte Carlo (MCMC) methods have also been used for various hydrologic and hydraulic models to assess parameter uncertainty within the Bayesian framework. Among the MCMC methods, Metropolis-Hastings algorithm [Hastings 1970] and Gibbs sampler [Geman and Geman 1984] are most widely used to obtain a sample of parameter values, which is

approximately from posterior distributions. Their parameter set selection is more efficient than GLUE because they identify the high probability region and focus sampling in that region [Vrugt et al. 2003]. Each parameter set is iteratively generated by considering the likelihood information from previous model simulations. Once the MCMC algorithm generates parameter sets from a stationary distribution, the algorithm has converged. A sample generated after convergence conforms to the joint posterior PDF (including parameter correlations) and can be used to assess the implications for prediction uncertainty.

Following the Metropolis-Hastings algorithm and Gibbs sampler, various MCMC algorithms have been proposed to achieve convergence with fewer simulations 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 $\text{DREAM}_{(D)}$ [Vrugt and Ter Braak 2011], $\text{DREAM}_{(ZS)}$ [Laloy and Vrugt 2012], and $\text{DREAM}_{(ABC)}$ [Sadegh and Vrugt 2014]. Despite such efforts, MCMC methods still require many simulations not only for algorithm convergence but also for collecting the posterior sample after convergence. Vrugt et al. [2009] used at least 40,000 simulations of a conceptual watershed model to achieve convergence, and Ajami et al. [2007] generated 20,000 parameter sets after convergence to obtain well-specified histograms of the sampled parameter values. For sediment transport model simulations of flume experiments, Sabatine et al. [2015] found that an MCMC method requires nearly as many simulations as GLUE. Furthermore, parallel computing is not readily implemented for MCMC methods because the parameter set generated in a given iteration depends on those generated previously [Foglia et al. 2009].

Metamodeling approaches have recently received attention in the field of water resources modeling as a way to efficiently characterize parameter uncertainty. Metamodels (also called surrogate models or emulators) approximate the likelihood surface using artificial neural networks, Gaussian process modeling, or radial basis function approximations based on a small sample of model simulations [Jones and Johnson 2009]. These methods have been applied to the modeling of streamflow under climate change [Dehgani et al. 2014; Humphrey et al. 2016], flood inundation [Teng et al. 2017], groundwater levels [Stone 2011], and contaminant transport in porous media [Nouani et al. 2017]. However, the approximation process in metamodeling can become complex as the number of parameters increases, which might also lead to computational inefficiency [Ong et al. 2004]. Moreover, the metamodel is only useful if the functional linkages between parameters and model outputs are properly specified, which can be difficult to achieve [Miller and Lacy 2003].

The main goal of this paper is to develop and test an efficient Bayesian method to assess parameter uncertainty and its contributions to prediction uncertainty. The Evolving Latin Hypercube (ELH) method is proposed, which generates the posterior parameter sample by replicating parameter sets instead of generating new but similar parameter sets that require additional simulations. ELH is evaluated using four synthetic parameter PDFs that were introduced by Vrugt et al. [2003; 2009]. In these cases, the posterior PDF is known exactly, so the accuracy of the method can be examined. The method is also tested by application to a sediment transport model for a 23-km reach of the Tachia River in Taiwan [Lai and Greimann 2010]. In this case, nine parameters in the Sedimentation and River Hydraulics – One Dimension (SRH-1D) model [Huang and Greimann 2013] are treated as uncertain. ELH is compared to both GLUE and SCEM-UA based on: (1) the number of simulations required to

obtain the uncertainty estimates, (2) the accuracy of the estimated posterior parameter PDFs, and (3) the resulting distributions of the model predictions.

2.2 ELH Methodology

In the Bayesian paradigm, the model parameters are treated as uncertain, and their uncertainties are mathematically expressed using a joint posterior PDF:

$$p\left(\boldsymbol{\theta}|\mathbf{y}\right) \propto p\left(\boldsymbol{\theta}\right) L\left(\mathbf{y}|\boldsymbol{\theta}\right) \tag{1}$$

where $p(\boldsymbol{\theta}|\mathbf{y})$ is the joint posterior PDF of a set of parameters $\boldsymbol{\theta}$, which represents the uncertainty in the parameters $\boldsymbol{\theta}$ given a calibration dataset $\mathbf{y}$. The joint posterior PDF can be obtained by combining the joint prior PDF $p(\boldsymbol{\theta})$, which describes the information available for the parameters $\boldsymbol{\theta}$ before calibration is performed, and the likelihood $L(\mathbf{y}|\boldsymbol{\theta})$, which is determined from the similarity between the calibration dataset $\mathbf{y}$ and model outputs when parameters $\boldsymbol{\theta}$ are used [Green 2001].

ELH estimates the joint posterior PDF $p(\boldsymbol{\theta}|\mathbf{y})$ based on the following general steps. First, a sample of parameter sets is generated from a uniform prior PDF, and second, these parameter sets are used in model simulations to determine their likelihoods. Third, a posterior sample of parameter sets is constructed by replicating each parameter set based on its posterior probability density value. Fourth, the parameter space is contracted to remove very low probability regions, and additional parameter sets are generated and used in model simulations. Fifth, a new posterior sample is constructed using all the parameter sets generated so far. Sixth, the stability of the posterior sample is checked (and steps four and five are repeated until stability is reached).

Seventh, the posterior sample is used in the forecast period. The following subsections describe each step in more detail.

2.2.1 Step 1: Generate Prior Sample

The ELH algorithm starts by generating a prior sample of parameter sets from a uniform joint PDF within the feasible ranges of the parameters, which are specified by the modeler. LHS [McKay et al. 1979] is used for the sampling. LHS divides the ranges of the S parameters $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_S]$ into B non-overlapping and equally-sized intervals. One value is randomly selected within each interval for each parameter. The B values of θ_1 are then randomly paired with the values of θ_2 and so forth to produce the prior sample of parameter sets (Fig. 1a).

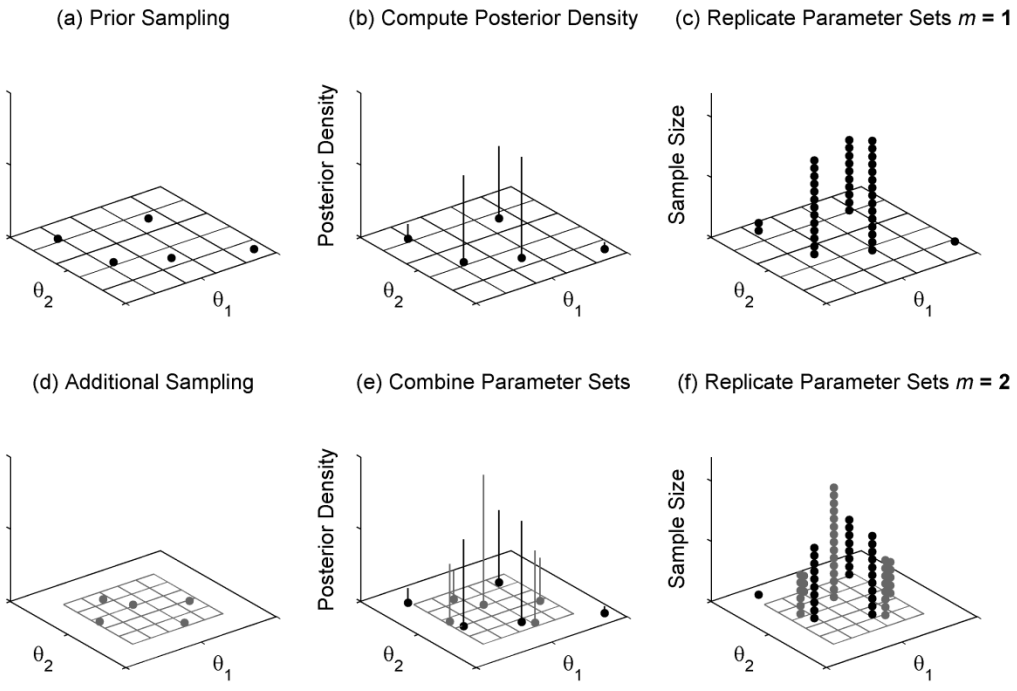

Fig. 1 Illustration of the ELH algorithm using two uncertain parameters ($S = 2$) with LHS sample size $B = 5$.

The LHS method can sample high-dimensional parameter spaces more efficiently than random sampling [van Griensven et al. 2006]. Finer discretization (i.e. larger B values) provides more thorough sampling but increases the computational time because each parameter set requires a model simulation. In general, LHS sample sizes should be proportional to the number of parameters considered [Stein 1987], so in ELH, B is found from $S \times n_s$, where n_s is an ELH algorithmic parameter. $n_s = 10$ has been found to suffice for sensitivity analyses [Chapman et al. 1994; Jones et al. 1998; Loeppky et al. 2009], but ELH results become independent of n_s when n_s is 50 or larger (based on the synthetic case studies presented later). Thus, n_s is set to 50 in all the applications of ELH that follow.

2.2.2 Step 2: Compute Posterior Densities

Each parameter set in the prior sample is used in a model simulation of the calibration period, and its posterior density value $p(\boldsymbol{\theta}_j|\mathbf{y})$ is evaluated using:

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

where j is the index for the parameter sets, $M_i(\boldsymbol{\theta}_j)$ is the model's simulated value when the parameter set $\boldsymbol{\theta}_j$ is used at measurement location and/or time i, y_i is the observed value at i, and N is the number of observations in the calibration dataset $\mathbf{y}$. The right side of Eq. (2) is proportional to the posterior density, but it is called the posterior density for simplicity (Fig. 1b).

Eq. (2) has been used in various MCMC methods because it evaluates the likelihood in a formal and simple way [Tiemann et al. 2001; Vrugt et al. 2003; Vrugt et al. 2008; Vrugt et al. 2009]. However, the model errors must be independent and Gaussian with constant variance in order for Eq. (2) to provide a formal evaluation of the posterior density [Box and Tiao 1973].

2.2.3 Step 3: Construct Posterior Sample

After calculating the posterior densities, ELH constructs a posterior sample of parameter sets that approximates the joint posterior PDF $p(\boldsymbol{\theta}|\mathbf{y})$. The parameter set that has the highest posterior density $\boldsymbol{\theta}_{\text{best}}$ is replicated n_r times (n_r is an ELH algorithmic parameter), and all copies are collected into the posterior sample. Each remaining parameter set $\boldsymbol{\theta}_j$ is then replicated based on the ratio of its posterior density to the most likely parameter set's posterior density:

$$\left\| n_r \frac{p\left(\boldsymbol{\theta}_j \middle| \mathbf{y}\right)}{p\left(\boldsymbol{\theta}_{\text{best}} \middle| \mathbf{y}\right)} \right\| \tag{3}$$

where $\|\cdot\|$ represents the rounding function (Fig. 1c). Through this replication process, the uniformly distributed prior sample is transformed into a posterior sample that conforms to the calculated posterior densities. Parameter sets with very low posterior densities can be excluded from the sample if the number from Eq. (3) rounds to zero. For the synthetic cases presented later, the results of ELH become independent of n_r when n_r is 1000 or larger, so $n_r = 1000$ is used for all the cases. If n_r is too small, then the rounding function removes too many parameter sets and the estimate of the posterior PDF is poor in low probability regions.

The posterior sample can be used to characterize the joint posterior PDF. For example, the marginal posterior PDF of each parameter, which describes the uncertainty in the parameter that remains after calibration, can be approximated using the histogram of the parameter values included in the posterior sample. In addition, the covariance structure of the joint posterior PDF can be explored using the correlation coefficients between the parameters in the posterior sample.

The replication of parameter sets in ELH essentially replaces the generation of additional parameter sets in MCMC methods. In most MCMC methods, the Metropolis algorithm

[Metropolis et al. 1953] is used to iteratively sample the parameter values that conform to the posterior distribution. The Metropolis algorithm generates trial parameter sets near existing parameter sets and keeps the trial parameter sets based on the ratio of the trial parameter set's posterior density and the current parameter set's posterior density. As the procedure iterates, the parameter space is explored and more likely parameter sets are retained. MCMC can also replicate parameter sets under certain conditions where the trial parameter set is rejected and the current parameter set is repeatedly used for comparing to another trial parameter set in the next iteration [Robert et al. 2010].

2.2.4 Step 4: Specify New Parameter Ranges for Additional Sampling

The initial sampling is designed to span the entire feasible parameter space, but it is possible that many generated parameter sets fall in regions with very low posterior density. In that case, the posterior PDF will be poorly characterized by the initial sampling. Thus, the initial posterior sample is used to update the parameter ranges. For each parameter, the new range is determined based on the interquartile range (IQR) of the posterior sample as follows:

$$\left[25\%\text{ quantile value} - \text{IQR}*n_w \qquad 75\%\text{ quantile value} + \text{IQR}*n_w\right] \tag{4}$$

where n_w is an ELH algorithmic parameter that adjusts the sampling range based on the posterior sample. The new range is not allowed to surpass the feasible range that was specified by the modeler. In some cases, the IQR can be zero if most members of the posterior sample are replicated from a single parameter set. In such a case, ELH maintains the previous range for generating new parameter sets.

The ELH parameter n_w must be specified to use Eq. (4). The purpose of that equation is to exclude regions with very low probability density so that reducing sampling ranges does not lose

any information about the main body of the PDF nor decrease the resulting uncertainty. If the underlying posterior distribution is Gaussian, Eq. (4) includes more than 99.99% of the probability density when $n_w = 2.5$ [Frigge et al. 1989]. However, posterior distributions can be non-Gaussian, which might require different n_w values. To examine this issue, Pearson distributions were generated that consistently have a mean of zero and a standard deviation of one but have varying skewness γ (from -1 to 1) and kurtosis κ (from 1.8 to 10). Fig. 2 shows the IQRs and the 99.99% probability bounds of those distributions. In this figure, even though the IQR remains constant for all these distributions, the upper and lower probability bounds change asymmetrically as the skewness changes. The bounds also widen as the kurtosis increases.

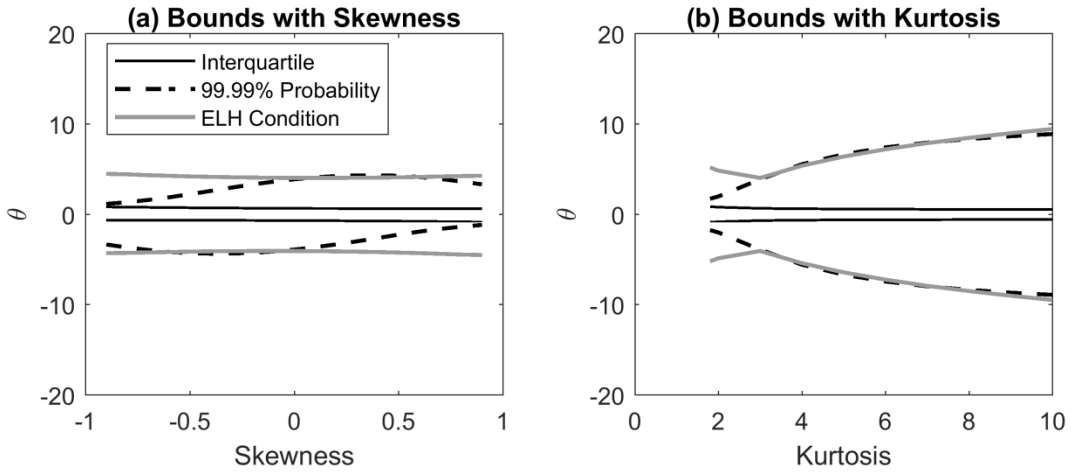

Fig. 2 Limits of the IQR, 99.99% probability, and ELH condition plotted by (a) varying skewness and (b) varying kurtosis for a Pearson distribution with zero mean and a standard deviation of one.

To account for the dependence on kurtosis, n_w is calculated from the kurtosis of the posterior parameter sample. When the kurtosis is below 3, n_w is set to 2.5 to avoid narrowing the ranges more than necessary. The skewness is not considered for simplicity and because it has a smaller effect than kurtosis. Specifically:

$$n_w(\kappa) = \begin{cases} 2.5 \times \left[2\dfrac{\log(\kappa)}{\log(3)} - 1 \right] & \text{for} \quad \kappa > 3 \\ 2.5 & \text{for} \quad \kappa \le 3 \end{cases} \tag{5}$$

, and Fig. 2 shows that the probability bounds estimated from Eq. (5) closely follow those of the Pearson distributions.

Once the new ranges are established from Eqs. (4) and (5), parameter sets are generated within the new ranges using a uniform joint PDF and LHS (Fig. 1d). A uniform joint PDF is used to avoid double counting the increased posterior density of the parameters in the new range (more likely parameters will be emphasized through the replication process). Once the new parameter sets have been generated, they are used in model simulations for the calibration period, and the posterior density of each parameter set is calculated using Eq. (2).

2.2.5 Step 5: Update Posterior Sample

To update the posterior sample, the copies of the parameter sets from the previous replication process of ELH are removed, and the new parameter sets are added to the previous (unique) parameter sets (Fig. 1e). The replication process is then performed using the entire collection of the parameter sets (Fig. 1f). The most likely parameter set might change with the inclusion of the new parameter sets, and additional parameter sets with very low likelihoods might be

removed from the posterior sample. ELH then iterates by calculating the new sampling ranges (repeating Step 4) and updating the posterior sample (repeating Step 5).

2.2.6 Step 6: Check Stability of Posterior Sample

The iterative process stops when the posterior sample converges to a stable distribution, which means the posterior sample has stable moments. To evaluate stability, the standard deviation (SD) for each parameter is first calculated from the posterior sample (with replications) in each iteration of ELH. Then, the variability of the SD values between iterations is considered. Specifically, the coefficient of variation (CV) of the second half of the collected SD values ($\mathbf{SD}_{s,m}$) is computed as:

$$\mathrm{CV}_{s,m} = \frac{\sqrt{\mathrm{Var}\left(\mathbf{SD}_{s,m}\right)}}{E\left(\mathbf{SD}_{s,m}\right)} \tag{6}$$

where

$$\mathbf{SD}_{s,m} = \left\{\mathrm{SD}_{s,\lfloor m/2 \rfloor+1},\ \cdots,\ \mathrm{SD}_{s,m-1},\ \mathrm{SD}_{s,m}\right\} \tag{7}$$

and $\mathrm{SD}_{s,m}$ is the SD of parameter s in the posterior sample from the m th iteration. The sample is considered adequately stable if the CVs of the SDs for all parameters are lower than 0.01. The posterior sample available at this point is considered to be consistent with the joint posterior PDF. Note that the CV is first calculated after 10 iterations (m = 10) so that sufficient SDs are available for calculating the variance in Eq. (6).

SD is used to evaluate stability because it quantifies the spread of the posterior sample, which is an important element of uncertainty quantification. Other statistical properties (e.g., IQR, total range, or kurtosis) could be used as the diagnostic variable in Eq. (6), but tests revealed that the

SD provides more consistent evaluations of stability than these other metrics. A collection of SDs is used because trends are difficult to detect if only consecutive SD values are considered. Only the second half of the SDs are used (similar to MCMC methods [Gelman and Rubin 1992; Geweke 1991]) to focus on the recent behavior of the ELH sampling. The SDs from early iterations usually vary drastically. The CV is used to require less variation in the SD for parameters that are better constrained (more certain) and to allow more variation for parameters that are poorly calibrated.

2.2.7 Step 7: Simulate Forecast Period

After the posterior sample has been finalized, its parameter sets are used to simulate the forecast period. The posterior sample includes many replicated parameter sets, but each parameter set needs to be simulated only once. The model forecasts are then replicated according to the number of parameter sets in the posterior sample, and the resulting collection of forecasts can be used to develop credible intervals, etc.

2.3 Case Studies with Synthetic Distributions

Four synthetic case studies are used to evaluate the performance of ELH. For each case, the correct posterior PDF for the parameters $p(\boldsymbol{\theta}|\mathbf{y})$ is specified in advance using an equation. These equations cover diverse problem features including a uniform PDF (Case 1), a narrow bimodal PDF (Case 2), a multivariate joint PDF (Case 3), and a complex banana-shaped bivariate PDF (Case 4). Those equations have been used for testing a variety of MCMC methods [Vrugt et al. 2003; Vrugt et al. 2009; Laloy and Vrugt 2012]. For these synthetic cases, the posterior density values are not calculated by comparing model results to observations (from Eq. (2)). Instead, the

posterior density is calculated by inserting each generated parameter set directly into the selected PDF equation. Thus, each evaluation of the PDF equation is analogous to a model simulation.

ELH is compared to two existing Bayesian methods: GLUE [Beven and Binley 1992] and SCEM-UA [Vrugt et al. 2003]. To implement GLUE, 100 parameter sets are generated from the pre-specified parameter ranges using LHS, and the posterior density value of each parameter set is calculated from the given PDF equation. Each posterior density value is then divided by the sum of all the posterior density values. Finally, the marginal posterior cumulative distribution for each parameter is produced by summing the relative posterior density values along the parameter ranges. Because GLUE does not produce a posterior sample, the stability diagnosis used in ELH cannot be applied. Instead, the sample size is increased by increments of 100 until the IQRs of the cumulative distributions change by less than 1%.

SCEM-UA starts by generating 100 parameter sets from a uniform joint prior PDF and computes the posterior density value of each parameter set using the PDF equation. SCEM-UA then partitions the parameter sets into 4 complexes, and the members in each complex are updated using the Metropolis algorithm, which iteratively generates a new parameter set to explore the parameter space (see details in Vrugt et al. [2003]). After 5 updates, the members of all complexes are recombined, shuffled, and re-divided into complexes. When the parameter sets are sampled from a stationary distribution, the algorithm has converged. Convergence is evaluated using the scale reduction score (SRS) [Gelman and Rubin 1992]. SCEM-UA discards the parameter sets generated before convergence, and uses a sample produced after convergence as the posterior sample. To check whether enough parameter sets are included in the posterior sample, the stability diagnosis used in ELH (Eq. (6)) is applied after every 100 parameter sets are added.

2.3.1 Univariate Distributions (Cases 1 and 2)

Case 1 considers a single parameter θ that has a uniform posterior PDF, which is expressed as:

$$p(\theta|\mathbf{y}) = \frac{1}{b-a} \tag{8}$$

where a and b are specified as zero and one, respectively. The feasible range for the parameter is also from 0 to 1 when implementing the three uncertainty methods.

Case 2 considers a single parameter with a bimodal posterior PDF, which can be written:

$$p(\theta|\mathbf{y}) = \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\theta^2\right] + \frac{2}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(2\theta-8)^2\right] \tag{9}$$

This PDF is the sum of two Gaussian distributions and has modes at $\theta = 0$ and 4. 99.99% of its probability occurs between -4 and 8, but the feasible range is from -20 to 20 to examine how the methods work if the feasible range is much wider than the main body of the posterior PDF.

Table 1 shows the simulations required to reach stability for GLUE, SCEM-UA, and ELH. GLUE needs 4,000 and 7,000 simulations to obtain stable estimates for Cases 1 and 2, respectively. SCEM-UA converges in 100 and 500 simulations and reaches stability in 1,200 and 2,500 simulations for Cases 1 and 2, respectively. SCEM-UA obtains stable results using fewer simulations than GLUE, but it should be noted that GLUE necessarily uses a different stability condition than the other methods. ELH produces stable estimates in 500 simulations for both cases, which are only 42% and 20% of the simulations required by SCEM-UA for Cases 1 and 2, respectively. Both SCEM-UA and ELH use the same stability diagnosis.

Table 1 Numbers of simulations required to provide stable posterior samples of the parameter sets, errors in the estimated marginal posterior CDFs, and errors in the estimated marginal posterior PDFs for the four synthetic cases. Numbers within brackets for SCEM-UA results indicate the simulations required to reach convergence (i.e. the burn-in period).

Case	Para-meter	Simulations for Stability			CDF Errors			PDF Errors		
		GLUE	SCEM-UA	ELH	GLUE	SCEM-UA	ELH	GLUE	SCEM-UA	ELH
1	θ	4,000	1,200 (100)	500	0.005	0.009	0.002	-	0.08	0.07
2	θ	7,000	2,500 (500)	500	0.05	0.05	0.07	-	0.23	0.19
3	θ_1	500,000	68,000	27,500	0.05	0.02	0.03	-	0.07	0.09
	θ_2		(3,000)		0.07	0.03	0.03	-	0.10	0.11
	θ_3				0.06	0.03	0.08	-	0.09	0.11
	θ_4				0.05	0.06	0.06	-	0.12	0.16
	θ_5				0.09	0.06	0.10	-	0.11	0.14
4	θ_1	400,000	275,200	77,600	0.22	0.22	0.29	-	0.08	0.13
	θ_2		(6,000)		0.29	0.36	0.44	-	0.09	0.10

Fig. 3a illustrates the progress of GLUE, SCEM-UA, and ELH by plotting the error in the estimated CDF as a function of the number of simulations. The CDF error measures the area between the target posterior CDF and the estimated posterior CDF. The CDF error for ELH becomes stable much quicker than the other methods for both cases. The improvement in computational efficiency is more notable in Case 2 than Case 1. In Case 1, the feasible range matches the target uniform distribution exactly, but for Case 2, the feasible range is much wider than the target distribution. For Case 2, ELH reduces the sampling bounds to -10 and 12 after the first sampling, and this new range changes little in subsequent iterations. On the other hand, both GLUE and SCEM-UA keep the initial limits of -20 and 20 during the entire process, so they run many simulations in regions with posterior density values near zero, which are not very useful for estimating the posterior distribution.

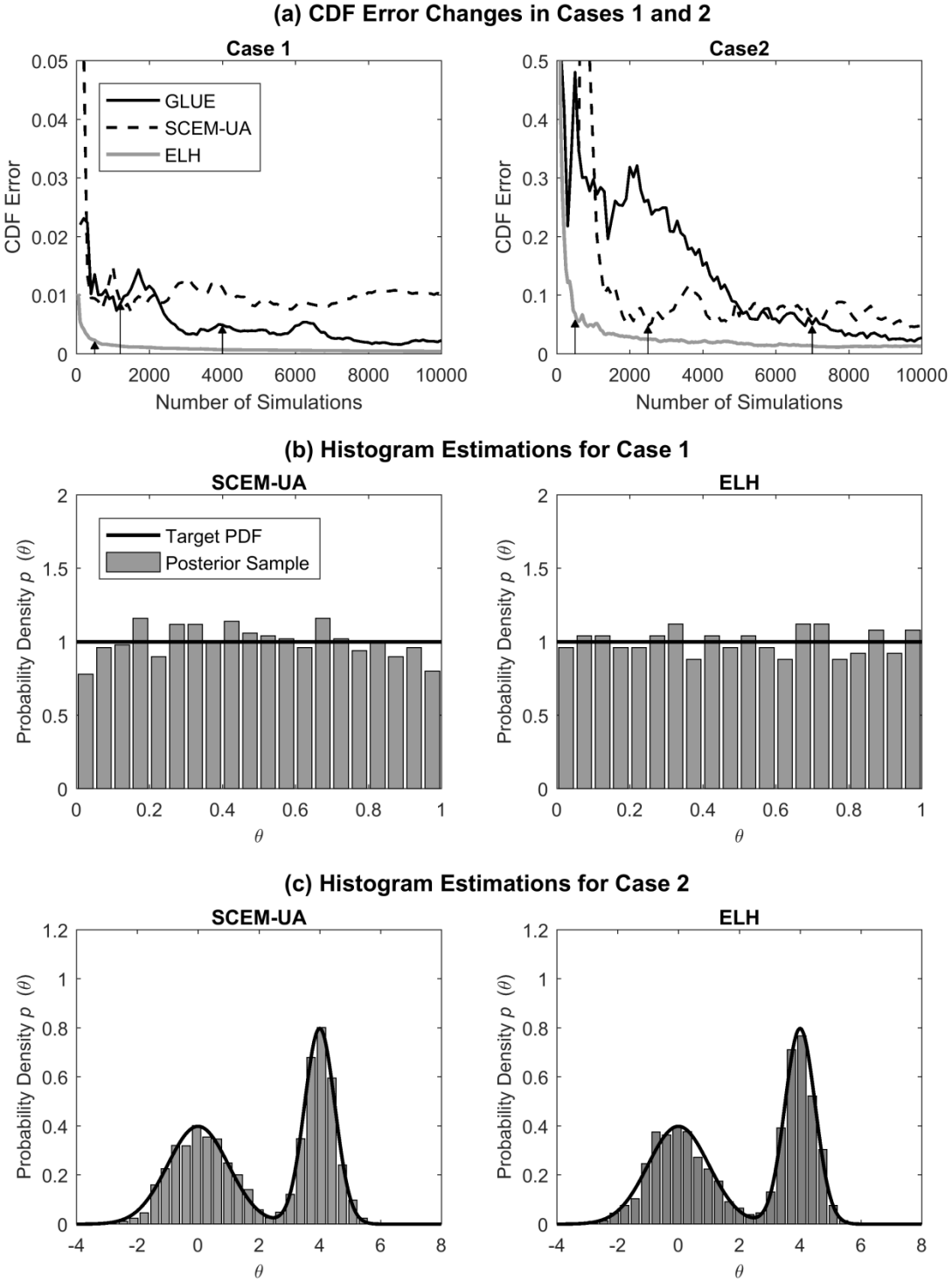

Fig. 3 (a) CDF errors as a function of the number of simulations performed for Cases 1 and 2, (b) target posterior PDFs (black line) and histograms (grey bars) for Case 1, and (c) target posterior PDFs and histograms for Case 2. The vertical arrows in (a) indicate the points where each method reaches stability.

Table 1 also summarizes the CDF and PDF errors at the point which each methods reaches stability. To calculate the PDF error, a histogram is constructed using the posterior sample. Then, the vertical distances between the histogram bars and the target bar heights are measured, and the sum of the vertical distances is multiplied by a histogram bar width. The histogram is not available for GLUE because the method does not provide a posterior sample. Overall, the errors are similar among all three methods for Cases 1 and 2. When considering the CDFs, ELH has the lowest error for Case 1 and the highest error for Case 2. For the PDFs, ELH has lower errors than SCEM-UA for both Cases. The histograms produced by SCEM-UA and ELH (once stability is reached) are shown in Fig. 3b and Fig. 3c for Case 1 and Case 2, respectively. Overall, both methods successfully reproduce the target posterior PDFs.

We also evaluated the uncertainty methods using a case where the two modes are very far apart. To test this situation, all the conditions remain the same as Case 2 but the modes are located at θ = -12 and 15. For this case, the CDF estimates from ELH not only become stable much faster but also have lower errors than GLUE and SCEM-UA. Next, the uncertainty methods were also implemented for the same target distribution but enlarging the feasible range from [-20 20] to [-100 100]. This case was used to examine the effectiveness of the range reduction process in ELH. Overall, ELH provides better estimates for this case than the other two methods, but the improvement in accuracy of ELH is not significant compared to the previous cases. Specifically, the CDF estimates indicate that the errors from ELH fluctuate within 50~80% of those from GLUE and SCEM-UA during the simulations for this wide range case whereas the earlier case shows that ELH errors become less than 5% compared to the others. Because the distance between the two peaks is very wide, a wide IQR is produced for the posterior distribution. Thus, the sampling range of ELH remains wide during the analysis, and the bin size for LHS is often

wider than the body of each peak in the target PDF. For this reason, the parameter values within the main body of the target PDF are not sampled sufficiently, which brings less accurate estimation. To overcome this limitation, a larger value for the sample size multiplier n_s (with narrower width) of ELH can be used. For both widely-spaced bimodal test cases, the diagnosis of posterior sample stability using Eq. (6) does not work well. Specifically, both SCEM-UA and ELH present very irregular histogram if their posterior samples are determined based on Eq. (6), and they require much more simulations (at least 20,000) to acquire smoother distributions that reproduce the target posterior PDF well. Because the two modes of the target PDF are far apart, the SD of the posterior sample reflects only the spread between those two peaks (not the spread of the parameter values within each peak). These results suggest the need to find a more robust way to determine the stability of a posterior sample for such cases. However, such widely-spaced bimodal distributions are not expected to be common in real modeling problems.

2.3.2 Multivariate Normal Distribution (Case 3)

Case 3 uses a five-dimensional Gaussian distribution with correlated parameters. The target posterior PDF for the parameters $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5]$ can be expressed as:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^5 |\boldsymbol{\Sigma}|}} \exp\left[-\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}) \right] \tag{10}$$

where:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 & \mu_2 & \mu_3 & \mu_4 & \mu_5 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{11}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0.5\sigma_1\sigma_2 & 0.5\sigma_1\sigma_3 & 0.5\sigma_1\sigma_4 & 0.5\sigma_1\sigma_5 \\ - & \sigma_2^2 & 0.5\sigma_2\sigma_3 & 0.5\sigma_2\sigma_4 & 0.5\sigma_2\sigma_5 \\ - & - & \sigma_3^2 & 0.5\sigma_3\sigma_4 & 0.5\sigma_3\sigma_5 \\ - & - & - & \sigma_4^2 & 0.5\sigma_4\sigma_5 \\ - & - & - & - & \sigma_5^2 \end{bmatrix} \tag{12}$$

$$\sigma_1^2 = 1, \quad \sigma_2^2 = 2, \quad \sigma_3^2 = 3, \quad \sigma_4^2 = 4, \quad \sigma_5^2 = 5 \tag{13}$$

$\boldsymbol{\mu}$ is the mean vector, $\boldsymbol{\Sigma}$ is the covariance matrix, and μ_s and σ_s^2 are the mean and variance of the s^{th} parameter, respectively. The distribution is centered on zero for all five parameters. The s^{th} parameter has a variance of s, and every pair of parameters has a correlation coefficient of 0.5. The feasible range for all five parameters is from -10 to 10.

For this case, GLUE requires a very large number of simulations (500,000) due to the high dimensionality of the parameter space (Table 1). Moreover, GLUE does not estimate the correlation coefficients because it only produces the marginal posterior distribution for each parameter. SCEM-UA needs 68,000 simulations to reach stability, which is far fewer than GLUE, but ELH needs only 27,000 simulations (about 40% of the simulations required by SCEM-UA). Fig. 4 shows the estimated variances and correlation coefficients of the parameters for the first 100,000 simulations of SCEM-UA and ELH. The variances from SCEM-UA change rapidly during the first 10,000 iterations and then gradually increase until about 60,000 simulations. In contrast, the variances from ELH become stable at about 30,000 simulations.

Upon reaching stability, the CDF and PDF errors are similar for all three methods, but both errors are slightly larger for ELH (Table 1). However, as shown in Fig. 4a, SCEM-UA underestimates the variances for the 3rd, 4th, and 5th parameters, which have relatively high target values. The ELH estimates of variance are usually closer to the target values. SCEM-UA might

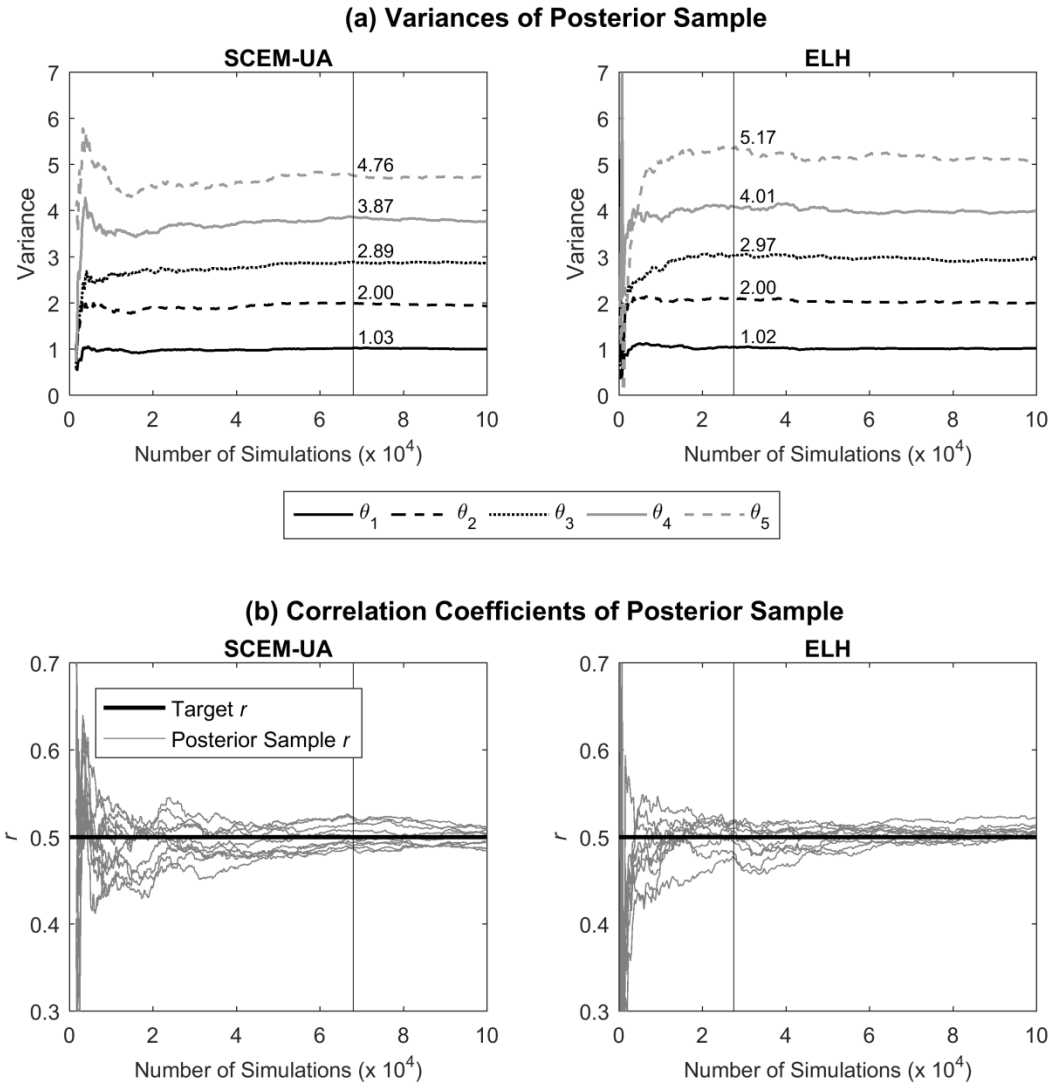

Fig. 4 (a) Parameter variances and (b) correlation coefficients from posterior sample as a function of the number of simulations performed for Case 3. The vertical lines indicate the points where each method reaches stability. The numbers above each line in (a) provide the variance of the corresponding parameters in the posterior sample where stability is reached. Each gray line in (b) corresponds to a pair of parameters in the posterior sample.

underestimate these variances because it violates the detailed balance principle, which requires MCMC methods to use a continuous updating sequence. SCEM-UA disrupts such continuity when it removes outlier trajectories in order to find the high probability parameter region quicker

[Laloy and Vrugt 2012], so it might misrepresent the posterior distributions. The correlation coefficients from both SCEM-UA and ELH become stable before 50,000 simulations at about 0.5 (Fig. 4b), which is the target value.

2.3.3 Banana-Shaped Distribution (Case 4)

Case 4 considers a banana-shaped posterior PDF, which represents a strong but highly nonlinear relationship between two parameters $\boldsymbol{\theta} = [\theta_1, \theta_2]$. The target posterior PDF is written as:

$$p\left(\boldsymbol{\theta}|\mathbf{y}\right) = f \circ \Phi\left(\boldsymbol{\theta}\right) \tag{14}$$

where f is a two dimensional normal distribution that is centered on zero:

$$f \sim N\left(0, \boldsymbol{\Sigma}\right) \tag{15}$$

and its covariance matrix is:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix} \tag{16}$$

To yield the nonlinear relationship, the function $\Phi(\boldsymbol{\theta})$ is defined as:

$$\Phi\left(\boldsymbol{\theta}\right) = \left(\theta_1,\ \theta_2 + 0.1\theta_1^2\right) \tag{17}$$

Following Vrugt et al. [2003], the feasible range for both parameters is from -100 to 100.

Similar to the other cases, GLUE requires the most simulations, and ELH requires the fewest simulations to acquire stable posterior samples (Table 1). In particular, ELH uses about 30% of the simulations needed by SCEM-UA. The required number of simulations increases significantly for both ELH and SCEM-UA compared to Case 3 because the target posterior PDF

is more complex (even though the number of parameters is reduced). GLUE neglects the complex relationship between the parameters, so it requires fewer simulations for this case than Case 3.

Fig. 5a shows scatterplots of the parameter sets included in the posterior samples from SCEM-UA and ELH, and the area within the bold lines includes 99.99% of the target PDF's probability (from Eq. (14)). The density of points within this area is notably different between two methods due to differences in their construction of the posterior sample. As described earlier, ELH generates parameter sets approximately uniformly within its sampling bounds and replicates the parameter sets based on their posterior densities. On the other hand, SCEM-UA generates parameter sets that explore the parameter space and collects them based on their posterior densities.

Fig. 5b and Fig. 5c show the estimated marginal PDFs for parameters θ_1 and θ_2, respectively. ELH produces an uneven histogram (especially for θ_1), probably due to its replication of parameter sets. This behavior also causes the errors in the CDF and PDF estimates to be larger for ELH than the other methods for this case (also for Case 3) (Table 1). However, the overall shapes of the marginal PDFs are well estimated by both methods, and the difference between the errors for SCEM-UA and ELH are small.

For this case, the range adjustment multiplier n_w plays an important role. The target marginal posterior PDF for θ_2 has a kurtosis near 11, and its 99.99 % probability region spans from -90 to 14 (Fig. 5c). If kurtosis was not considered when re-specifying the ranges (Eq. (5)), ELH would miss the long tail in the marginal posterior PDF of θ_2.

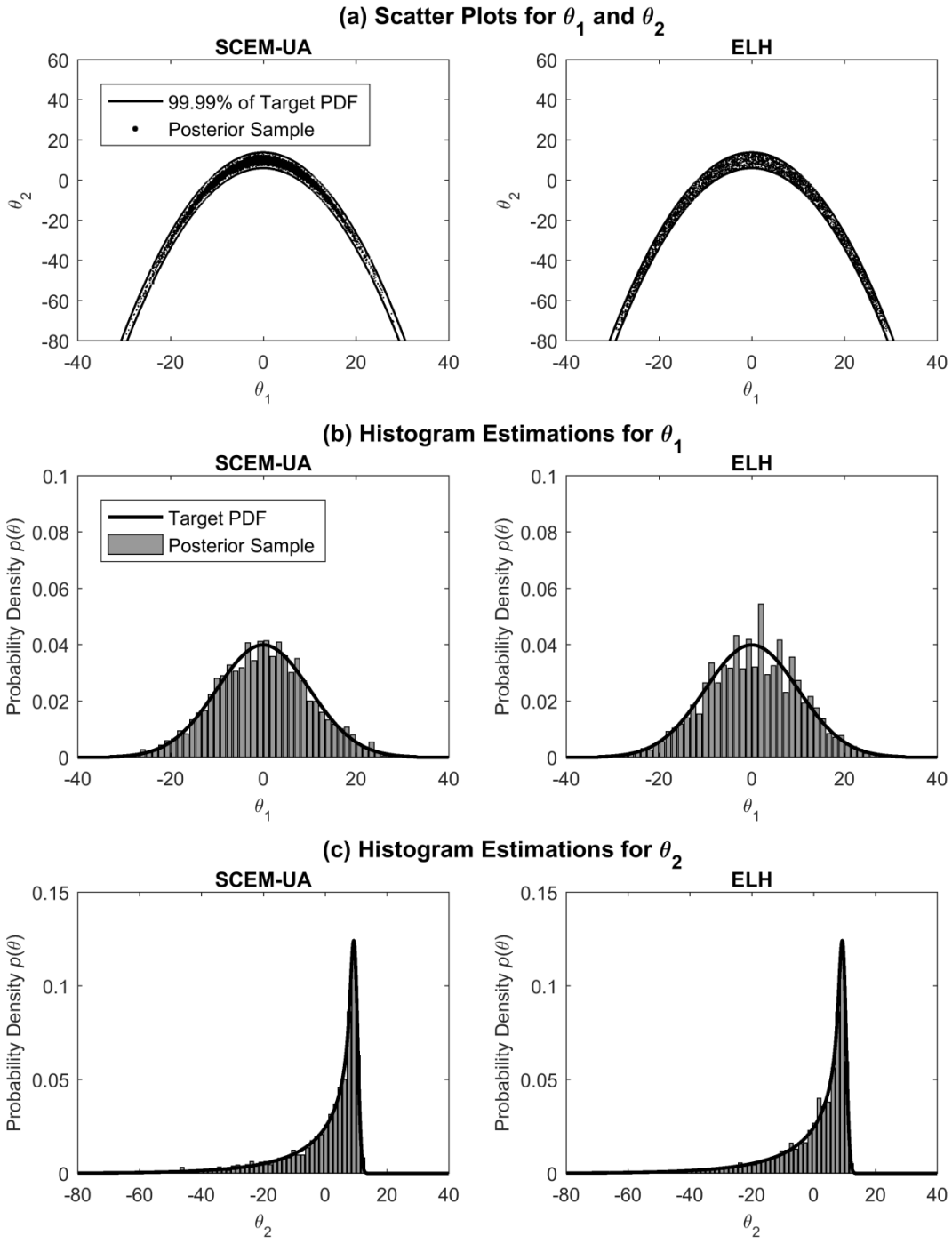

Fig. 5 (a) Scatterplots for the parameter sets included in the posterior sample from SCEM-UA and ELH along with the bounds of the 99.99 % probability area from the target posterior PDF, (b) marginal target posterior PDF (solid line) and histogram (grey bars) for θ_1, and (c) marginal target posterior PDF and histogram for θ_2 in Case 4.

2.4 Case Study with a Sediment Transport Model

The final case study evaluates ELH for an application of SRH-1D to the Tachia River in Taiwan. SRH-1D was developed and has been extensively used by the U.S. Bureau of Reclamation [Huang and Greimann 2013] to predict impacts of potential river restoration activities. The following sections briefly introduce the nine SRH-1D parameters considered here and the application to the Tachia River. The results are then discussed based on: (1) computational times for uncertainty estimates, (2) estimated posterior PDFs of model parameters, and (3) accuracy of prediction uncertainty estimates.

2.4.1 SRH-1D Model Parameters

SRH-1D computes flow hydraulics by solving the energy equation for steady, gradually varied flow. Given a time series of incoming flow rates, standard step method is used to compute the water surface elevations at each cross-section along the channel. The energy equation between adjacent cross sections i and $i + 1$ is expressed as:

$$z_{i+1} + \alpha_{i+1} \frac{U_{i+1}^{\ 2}}{2g} = z_i + \alpha_i \frac{U_i^{\ 2}}{2g} + h_f + h_c \tag{18}$$

where z represents the water surface elevation, α is the kinematic coefficient, U is the cross-sectional average velocity, g is gravitational acceleration, h_f is friction loss, and h_c is contraction or expansion loss. To find the friction loss in Eq. (18), Manning's equation is used, which requires specifying the roughness coefficient n (a model parameter).

For sediment transport computations, Exner equation routing is used to calculate the changes of the sediment volume in the bed. The Exner equation expresses mass conservation as:

$$\frac{\partial Q_s}{\partial X} + (1-\phi)\frac{\partial A_d}{\partial T} - q_{sl} = 0 \tag{19}$$

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. To solve this equation, the sediment porosity ϕ (a model parameter) need to be specified.

The volumetric sediment discharge is computed by calculating the transport capacity separately for each predefined grain size class. Several transport capacity equations are available in SRH-1D, but the Wu et al. [2000] equation is used here because it applies for wide range of sediment sizes (0.01 mm to 128 mm) and usually provides better performance in natural river simulations than other available equations [Lai and Greimann 2010]. The Wu et al. [2000] equation computes total bed material load for each grain size class by combining the bed load and suspended load, and this equation requires specification of two parameters: the non-dimensional reference shear stress θ_r and the hiding and exposure coefficient λ. The parameter θ_r is used to calculate the critical shear stress, which strongly affects the transport capacity of total bed material load because it controls the initiation of particle movement. The critical shear stress for class t ($\tau_{c,t}$) is computed as:

$$\tau_{c,t} = \theta_r \left[(\rho_s / \rho) - 1 \right] d_t \xi_t \tag{20}$$

where ρ_s is the density of the sediment, ρ is the density of water, and d_t is the median diameter of class t. The hiding and exposure function ξ_t describes the extent to which small particles are hidden by large particles in the bed. Such hiding leads to an increase in the critical shear stress for small particles and a reduction in the critical shear stress for large particles. This function ξ_t is calculated:

$$\xi_t = \left(\frac{p_{h,t}}{p_{e,t}} \right)^{\lambda} \tag{21}$$

where $p_{h,t}$ and $p_{e,t}$ are the hidden and exposed probabilities of particles in class t. If the parameter λ is zero, the mobility of each class is determined by its median size. If λ is one, then the different-sized particles have equal probability to transport (equal mobility).

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

$$L_{total} = (1 - f_s) L_b + \frac{\zeta f_s}{w_{f,t}} Uh \tag{22}$$

where f_s is the fraction of suspended load, $w_{f,t}$ is the fall velocity of particles in class t, U is the average velocity for the cross-section, and h is the hydraulic depth. The bed load adaptation length L_b is calculated as:

$$L_b = b_L h \tag{23}$$

where b_L is the bed load adaptation length multiplier (a parameter). The suspended sediment recovery factor ζ in Eq. (22) possesses different values depending on whether deposition (ζ_d) or scour (ζ_s) occurs at a given location, where ζ_d and ζ_s are parameters.

To simulate bed material mixing, SRH-1D partitions the channel bed into one active layer, which is the upper zone containing sediment particles available for transport, and several inactive layers below the active layer. The active layer thickness multiplier n_{alt} (a parameter) is applied to the geometric mean of the largest grain size class to determine the thickness of the active layer. In addition, the size distribution of the bedload material that is transferred from the active layer to

the inactive layers during deposition is controlled by the weight of bed load fraction χ (a parameter).

2.4.2 Application to the Tachia River

The 23-km reach of the Tachia River stretches from the Shih-Gang Dam to the ocean (Fig. 6). Within the considered reach, the bankfull widths are from 300 to 1200 m and the average slope is about 0.011. The bed material sizes observed in 2007 range from 0.125 mm to 512 mm (sand to boulders) with a median size (D50) of 108 mm, and the dominant substrate includes cobbles and gravels. From 2001 to 2009, severe erosion occurred 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. The Tachia River has been previously simulated using SRH-1D by Lai and Greimann [2010] and Jung et al. [2017].

Fig. 6 Satellite photo of Tachia River in Taiwan [adapted from Google Maps 2018].

For the present study, simulation of 2001 to 2005 is used for calibration and simulation of 2005 to 2009 is considered the forecast (prediction). Channel geometry data are available for both 2001 and 2005. A time series of measured discharges at the Shih-Gang Dam from 2001 to 2009 is used as the upstream boundary condition (BC), but no incoming sediment flow is applied. Flow and sediment inputs from tributaries are not considered due to lack of data [Jung et al. 2017]. The downstream BC is specified using the critical depth obtained from HEC-RAS simulations based on the 2005 geometry [Lai and Greimann 2010]. Sediment size gradation data from 2007 are averaged along the channel and used as the initial bed material distribution. This case is expected to include both scour and sedimentation at different locations and/or times within a single simulation.

Table 2 summarizes the feasible ranges for the nine SRH-1D parameters, which are determined based on following literature. The range of Manning's roughness n reflects 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 steams [Mueller et al.

Table 2 Key model parameters in SRH-1D and their feasible ranges for the uncertainty analysis. All parameters are dimensionless except as noted.

Parameter	Feasible Range
Manning's roughness coefficient, n	0.035 ~ 0.045 s $m^{-1/3}$
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

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 [Huang and Greimann 2013].

The model output of interest is the net sediment deposition/erosion at cross-sections along the reach. Observations are available for this variable at 43 cross-sections for the calibration period and 48 cross-sections for the forecast period. The posterior distributions for the nine model parameters are estimated using GLUE, SCEM-UA, and ELH during the calibration period. The forecast period is then simulated using the posterior parameter sets, and the estimated prediction uncertainty is compared to the observations. All three uncertainty methods calculate the posterior density value of each parameter set using Eq. (2) (based on the model simulation). Prior to the analysis, the model errors were investigated to determine whether the underlying assumptions of Eq. (2) apply. To perform this test, the model errors were calculated using a manually calibrated SRH-1D model. The residual lag plot showed a random pattern, so each error can be considered independent. The Kolmogorov-Smirnov test also indicated that the model errors are approximately Gaussian with constant variance. For the implementation of the methods, GLUE was conducted by increasing the LHS sample size by 1,000 until the results became stable, and stability was evaluated using the same approach that was applied to the synthetic cases. For SCEM-UA, an initial population of 250 parameter sets was generated and divided into 10 complexes. These values are appropriate for cases where numerous uncertain parameters are expected to be correlated [Vrugt et al. 2003]. ELH used the same algorithmic parameters as the synthetic cases.

2.4.3 Computational Times for Uncertainty Estimates

The total simulations used for the calibration and forecast periods in each method are compared in Fig. 7(a). For GLUE, the IQRs of the posterior cumulative distributions for all parameters are stable by 70,000 simulations. For the forecast period, the sample of parameter sets was generated from those posterior distributions using LHS. When the sample size is 1,000 or more, the IQRs for all parameters have less than 1% errors compared to the IQRs of the posterior distributions, which were obtained from calibration period. Thus, 1,000 simulations were used for the forecast period for GLUE. For SCEM-UA, convergence is checked every 50 simulations, and 3,450 simulations are required to reach convergence based on the SRS threshold (the calibration burn-in period). 13,400 more simulations are required to generate a stable posterior sample based on Eq. (6). Because only 3,282 parameter sets are unique in the posterior sample (the total size is 13,400), SCEM-UA requires 3,282 simulations of the forecast period. Thus, the total number of simulations required by SCEM-UA is 20,132. For ELH, stability is checked every 450 simulations and all parameters reach stability based on Eq. (6) at 9,000 simulations. At that point, the ELH has produced a posterior sample that includes 95,197 parameter sets but only 2,318 unique parameter sets. Thus, only 2,318 forecast simulations are required for ELH, and the total number of simulations required for ELH is 11,318. Overall, ELH requires 16% and 55% of the simulations needed by GLUE and SCEM-UA, respectively.

The simulations were performed using four parallel 3.50 GHz Intel i5-4690 processors with 8 GB of RAM. A single simulation of the calibration period required 20 sec, and a single simulation for the forecast period needed 30 sec. The continuous computation times were about and 118 hr for GLUE, 31 hr for SCEM-UA, and 17 hr for ELH (Fig. 7(b)). High performance computing systems with large numbers of processors would reduce the computation time for all

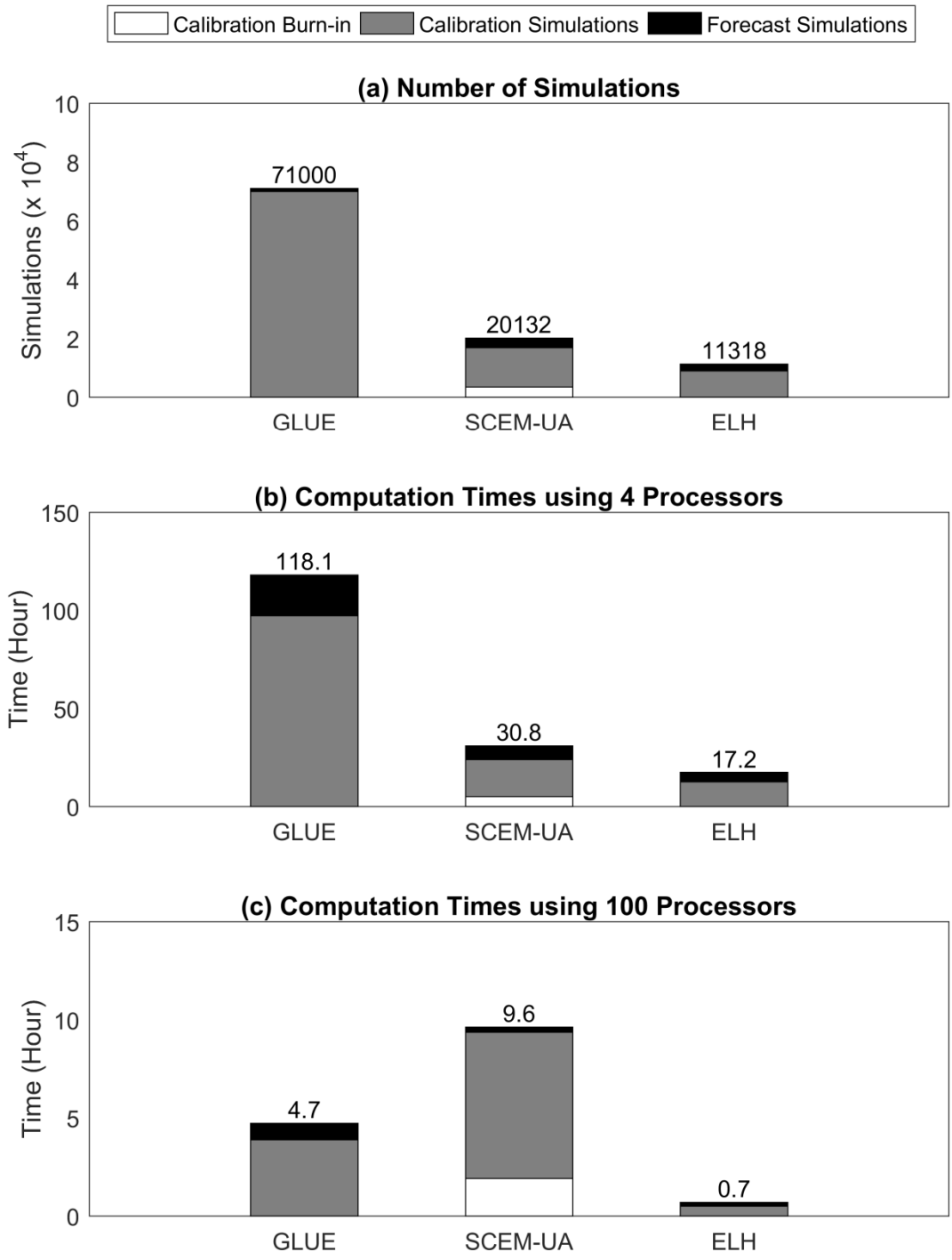

Fig. 7 (a) Number of simulations required to obtain prediction uncertainty estimates, (b) computation times for those simulations using 4 parallel processors, and (c) computation times using 100 parallel processors for GLUE, SCEM-UA, and ELH.

three methods. GLUE can use all processors efficiently because each parameter set is generated independently. With 100 processors, the computation time would decrease to 4.7 hr (Fig. 7(c)). SCEM-UA would not have a similar reduction because the algorithm updates the parameter sets depending on the last retained parameter set. For this case, SCEM-UA used 10 parallel updating sequences so only 10 simulations could be completed at the same time, and the total computation time would reduce to 9.6 hr. Unlike SCEM-UA, ELH would be able to run the 100 parameter sets that are generated independently from LHS simultaneously if 100 parallel processors were available, and it would require only 0.7 hr for all simulations.

2.4.4 Estimated Uncertainty in Model Parameters

Fig. 8 plots the marginal posterior CDFs of the parameters from the three methods, and Table 3 provides the medians and IQR ratios from these distributions. The IQR ratio is the ratio of posterior IQR to prior IQR and can be interpreted as the fraction of the initial uncertainty that remains in the parameter after calibration. Among the parameters, the reference shear stress θ_r is most constrained by the calibration data because it has the steepest CDF (Fig. 8) and the smallest IQR ratio. The parameter θ_r relates the flow velocity to the overall susceptibility of bed materials to erosion. Thus, it strongly affects the evolution of the bed profile, which is associated with the net deposition volume. The median values for θ_r are around 0.048~0.049 (Table 3), which are realistic for the bed material sizes found in the Tachia River [Lai and Greimann 2010]. The Manning's roughness n, hiding and exposure coefficient λ, and active layer thickness multiplier n_{alt} are also constrained to some degree by the data (Fig. 8), and their IQR ratios range from 11% to 62% (Table 3). In contrast, the other five parameters (ζ_d, ζ_s, b_L, χ, and ϕ) remain unconstrained after calibration (Fig. 8). They have nearly linear CDFs, which

imply nearly uniform posterior PDFs. Large uncertainty remains in their values because they have little impact on the model simulation in this case.

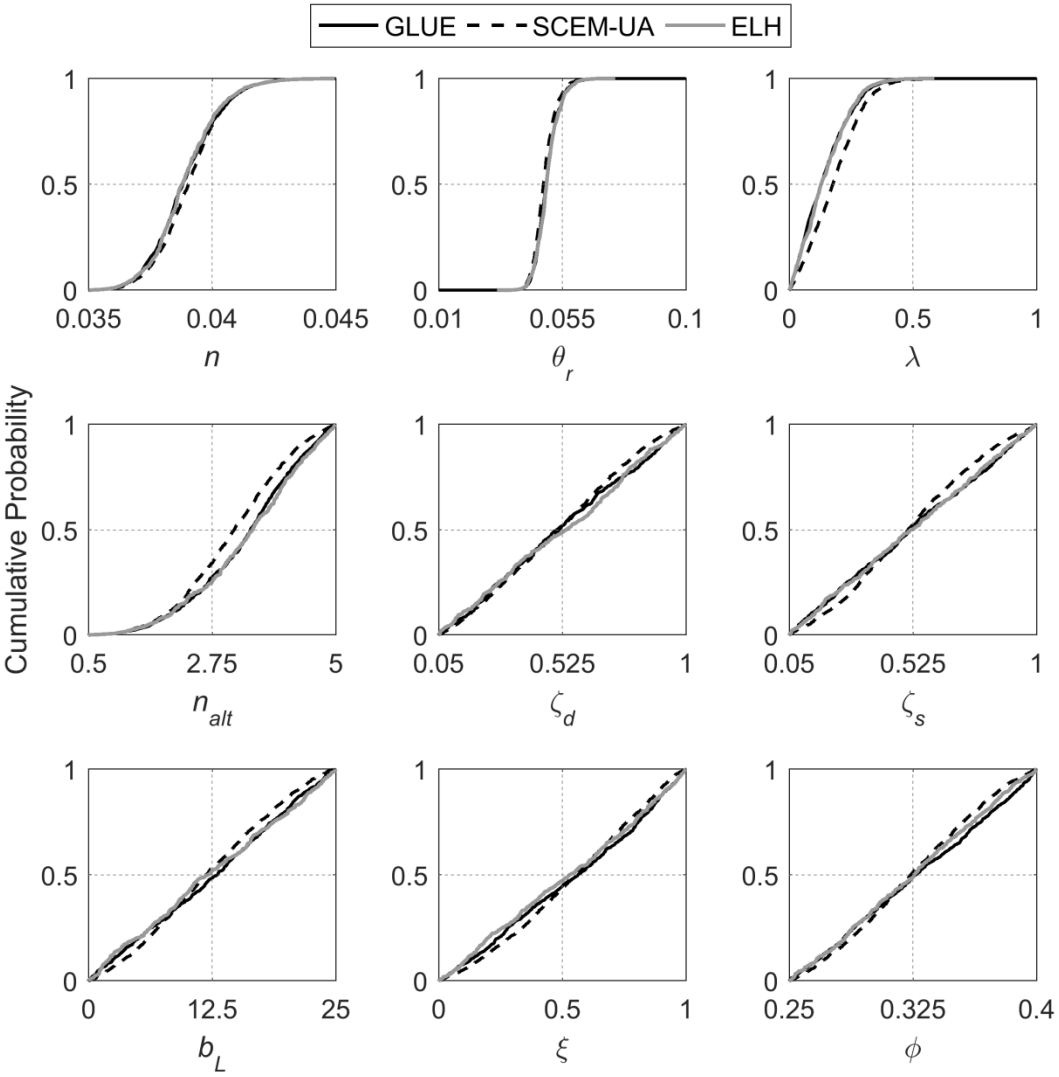

Fig. 8 Marginal posterior CDFs for the SRH-1D model parameters as estimated by GLUE, SCEM-UA, and ELH for the Tachia River case.

Table 3 Medians of the posterior parameter PDFs for the SRH-1D model parameters and the ratios of the IQRs from the posterior and prior parameter PDFs for the Tachia River case.

	Median			IQR Ratio		
Parameter	GLUE	SCEM-UA	ELH	GLUE	SCEM-UA	ELH
n	0.039	0.039	0.039	36.3%	33.3%	34.7%
θ_r	0.049	0.048	0.049	11.6%	11.3%	11.1%
λ	0.129	0.176	0.129	29.1%	31.1%	28.5%
n_{alt}	3.445	3.147	3.483	62.6%	61.2%	61.5%
ζ_d	0.504	0.517	0.536	100.1%	83.7%	100.5%
ζ_s	0.509	0.505	0.515	103.9%	78.7%	99.7%
b_L	13.056	11.883	12.080	97.9%	81.9%	99.6%
χ	0.563	0.557	0.529	105.5%	84.5%	106.7%
ϕ	0.327	0.324	0.326	102.7%	80.5%	96.8%

For most parameters, the three uncertainty methods provide similar marginal posterior cumulative distributions (Fig. 8). In fact, GLUE and ELH show nearly identical distributions for all parameters, but SCEM-UA differs from the other two methods for hiding and exposure coefficient λ and active layer thickness multiplier n_{alt}. The medians for λ are 0.13 and 0.18 for ELH and SCEM-UA, respectively (Table 3). According to Eq. (21), the lower value of λ from ELH indicates more influence of bed material gradations on the sediment transport process compared to SCEM-UA. All the three methods calculate the highest posterior densities when λ is about 0.2, so SCEM-UA produces a posterior sample that is more concentrated on the most likely value. This behavior is also seen with the IQR ratios of the five poorly constrained parameters (ζ_d, ζ_s, b_L, χ, and ϕ). Specifically, GLUE and ELH both produce IQR ratios around 100%, but SCEM-UA provides IQR ratios around 80% for those parameters (Table 3). This behavior is also consistent with Case 3 where the SCEM-UA underestimated the largest target

variances. Overall, these results may suggest that SCEM-UA underestimates the variance of posterior PDFs with large variances. However, the correct distributions are not known for this case.

Table 4 compares the correlations between the four well-calibrated parameters (n, θ_r, λ, and n_{alt}) as estimated by SCEM-UA and ELH (GLUE does not consider parameter correlation). The poorly-constrained parameters exhibit correlations near zero (not shown). Overall, the values from SCEM-UA and ELH show similar tendencies in Table 4. For example, the reference shear stress θ_r indicates notable correlations with three other parameters. The strong relationship between θ_r and λ is expected given the mathematical structure of Wu et al. [2000] equation. In particular, both parameters affect the critical shear stress $\tau_{c,t}$ as described in Eqs. (20) and (21).

Table 4 Correlation coefficients among the four well-constrained model parameters of SRH-1D for the Tachia River case.

	SCEM-UA				ELH			
Parameters	n	θ_r	λ	n_{alt}	n	θ_r	λ	n_{alt}
n	1	0.31	-0.05	-0.09	1	0.28	-0.02	-0.13
θ_r	0.31	1	-0.37	0.48	0.28	1	-0.40	0.48
λ	-0.05	-0.37	1	0.03	-0.02	-0.40	1	0.02
n_{alt}	-0.09	0.48	0.03	1	-0.13	0.48	0.02	1

2.4.5 Estimated Prediction Uncertainty

The net sediment deposition volumes for the forecast period were simulated using the posterior parameter sets generated from each method. Fig. 9 shows the mean prediction and the 99% credible interval for each method. The mean prediction from all methods consistently underestimates the erosion in the upstream portion of the reach (15-23 km) and the deposition at

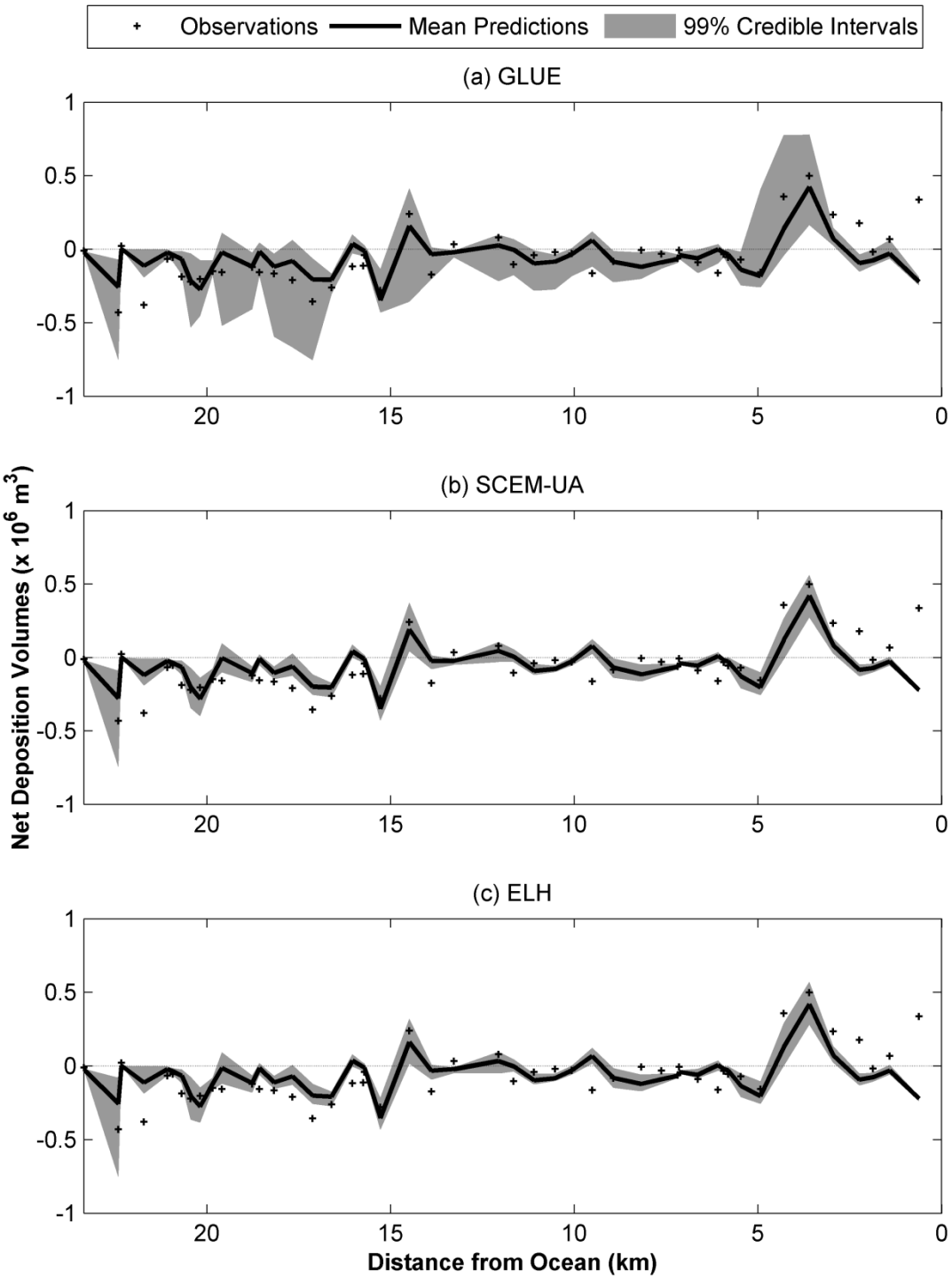

Fig. 9 Observations, mean predictions, and 99% credible intervals from GLUE, SCEM-UA, and ELH for net deposition volume for the forecast period of the Tachia River case.

the downstream end of the reach (0-4 km). GLUE produces a wide credible interval along the reach likely because it neglects the correlations between the parameters when generating the parameter sample for the forecast period. Wu and Chen [2009] and Sabatine et al. [2015] have previously demonstrated that estimates of prediction uncertainty substantially change depending on whether the correlations between sediment transport model parameters are included. SCEM-UA and ELH provide similar credible intervals despite their different posterior distributions for λ and n_{alt}. The differences in those two parameters might be attenuated by a slight change in the reference shear stress θ_r, which had a median value of 0.048 from SCEM-UA and 0.049 from ELH (Table 3). Table 4 shows that θ_r has a negative correlation with λ and a positive correlation with n_{alt}.

For each method, the performance of the mean prediction is evaluated using the Nash–Sutcliffe coefficient of efficiency (NSCE) [Nash and Sutcliffe 1970]. NSCE can range from $-\infty$ to 1, where 1 means the model predictions match the observations exactly (note that none of the three estimation methods uses NSCE in their algorithms). Fig. 10a shows that all three methods reproduce the observations better for the calibration period than the forecast period. In addition, no meaningful difference is observed in the performance of the three methods for either period.

The estimates of prediction uncertainty are evaluated using the continuous ranked probability score (CRPS) [Brown 1974; Matheson and Winkler 1976]. The CRPS is based on the principle that a probabilistic forecast should maximize the sharpness of the predictive distribution while still reproducing the observations [Gneiting et al. 2003]. It essentially calculates the area between the CDF of the model prediction and a step function at each observation. In this study, the CRPS is calculated at each cross-section and then averaged. Specifically:

$$CRPS = \frac{1}{N}\sum_{i=1}^{N}\int\left[G_i(M) - 1\{M \geq y_i\}\right]^2 dM \tag{24}$$

where $G_i(M)$ denotes the cumulative distribution of the model predictions M at a cross-section i, and $1\{M \geq y_i\}$ is the Heaviside function that attains the value 1 if $M \geq y_i$ and the value 0 otherwise [Hersbach 2000]. A lower CRPS indicates better performance, and its minimum of

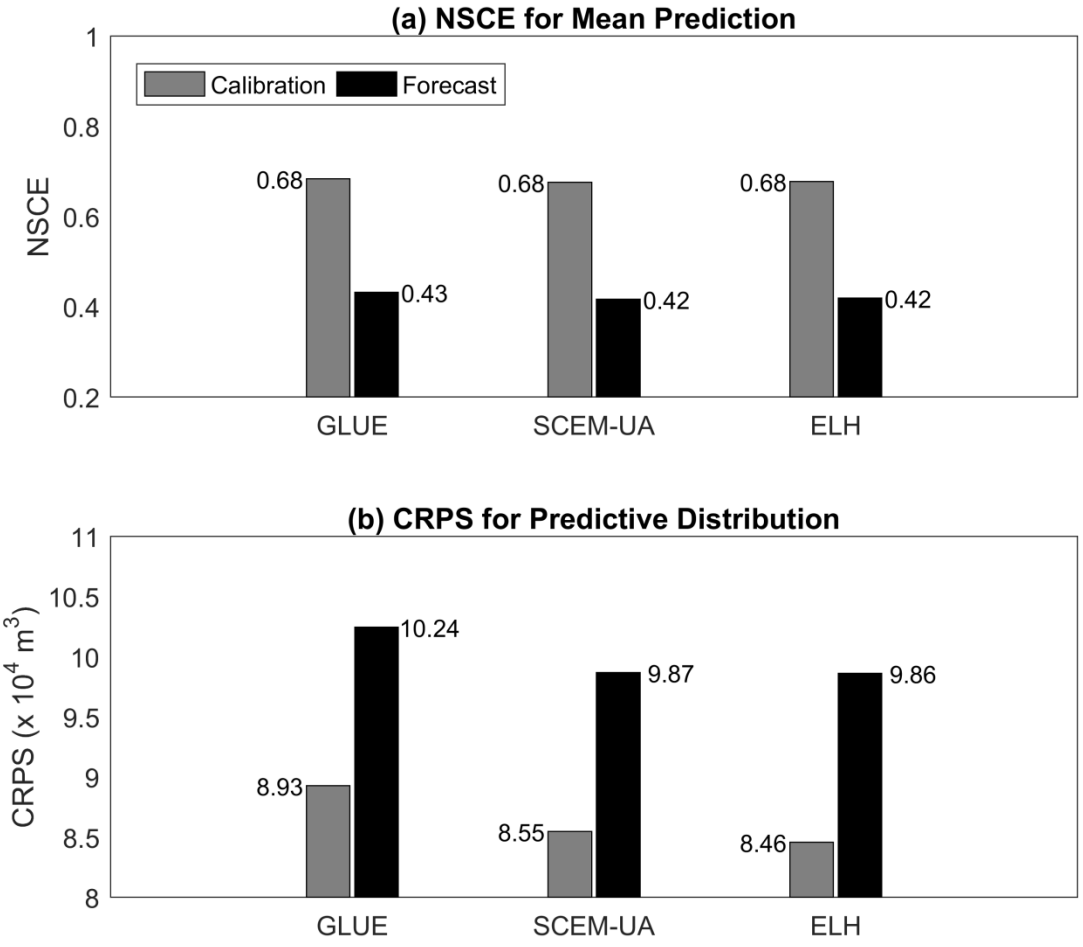

Fig. 10 (a) Nash–Sutcliffe coefficient of efficiency (NSCE) values for the mean prediction, and (b) continuous ranked probability score (CRPS) values for the predictive distribution from GLUE, SCEM-UA, and ELH.

zero is only achieved when the model predictions perfectly match the observations with no spread in the distribution. In Fig. 10b, GLUE performs worse than the other methods for both the calibration and forecast periods, which supports that its wide credible intervals are unrealistic. SCEM-UA and ELH have similar values for the CRPS. The credible intervals for SCEM-UA and ELH miss many of the observations (Fig. 9), which suggests that important sources of uncertainty have been neglected (e.g., uncertainty in the observations of net depositional volume, the model input data, the mathematical structure of the model).

In order to examine the impact of the other uncertainty sources, we also implemented the three uncertainty methods using SRH-1D model outputs as the calibration data. For this test, the calibration period was simulated using the single parameter set, which was identified as the most likely from the previous analysis. The model results for net deposition volume during the calibration period at the cross sections of interest were collected. Then, the posterior distributions of the model parameter were estimated by using the collected outputs as calibration dataset. This case includes errors only from the parameter values and removes other sources of uncertainty. For this case, ELH still shows an advantage in computational time to reach stable estimates compared to GLUE and SCEM-UA, but ELH did not exactly capture the correct parameter values (especially for the less important parameters) whereas SCEM-UA did. ELH applies the LHS method to sample parameter values for a range that includes the best point (not to find the exact best parameter values). As a result, it is more difficult for ELH to find those correct parameter values in this case where the target posterior distributions of the parameters are extremely narrow. In spite of such limitation, ELH provided the mean prediction that can reproduce the (synthetic) observations as well as GLUE and SCEM-UA did. The resulting 99% credible intervals are much narrower than those from the original case (using real observations),

but both SCEM-UA and ELH cover more than 80% of the available observations, which is about 40% more than the original results. In addition, the CRPS values for the predictive distributions also indicate that both SCEM-UA and ELH perform better than GLUE. However, some of the above results imply that this test is not the best way to consider the only parameter uncertainty. In particular, all the three methods provide very narrow posterior distributions for unimportant parameters that merely affect the model outputs, but they should be wide and/or uniformly distributed over their ranges because their high likely values are hard to be identified. Moreover, narrow credible intervals from both SCEM-UA and ELH still miss some (about 20%) of the available observations for both calibration and forecast periods. Those limitations occur likely because the use of the synthetic observations violates the underlying assumption of Eq. (2). Recall that the model errors must be independent and normally-distributed with constant variance in order for Eq. (2) to provide a formal evaluation of the posterior density [Box and Tiao 1973]. This test synthetically generated the calibration data using the model outputs, which were simulated using pre-specified parameter values. Hence, the model can perfectly reproduce the calibration data when the correct parameter set is used. In such a case, the model errors are all zero so that the assumption of normally-distributed errors is broken. Moreover, the posterior density value computed from Eq. (2) would toward infinity as the sampled parameter values approach the correct parameter set. Thus, all the three uncertainty methods are not able to formally assess the posterior density of each parameter set in this test case.

2.5 Conclusions

This paper proposed the ELH methodology to estimate parameter uncertainty and its impacts on prediction uncertainty. By application to synthetic posterior parameter PDFs and a sediment transport model, the following conclusions can be made.

1. For the cases considered, ELH is able to estimate the joint posterior PDF of the parameters and the prediction uncertainty using fewer model simulations than GLUE or SCEM-UA. Specifically, ELH required about 16% to 55% of the simulations required by SCEM-UA and provided an even greater advantage in simulations relative to GLUE. The reduction in required simulations was greatest for cases where the pre-specified parameter ranges are much wider than the main body of the joint posterior PDF. The improvement in efficiency occurs because ELH replicates parameter sets in the posterior sample rather than generating similar parameter sets that require additional simulations.

2. ELH can estimate the parameter variance better than SCEM-UA when the posterior parameter distribution has a large variance. In the case using a multivariate Gaussian target distribution, SCEM-UA underestimated large posterior variances, while ELH accurately estimated such variances. In the sediment transport model case, SCEM-UA also provided smaller IQR ratios than both ELH and GLUE for several model parameters that were poorly constrained by the calibration (and thus had large variances).

3. ELH produces more erratic histograms of posterior parameter sample because of its replication of parameter sets. That tendency caused higher errors in the estimated PDFs and CDFs for the synthetic cases with the multivariate Gaussian PDF and the banana-shaped PDF. Yet, overall, both SCEM-UA and ELH described the overall shapes of the joint posterior PDFs well for all synthetic cases considered.

4. SCEM-UA and ELH can provide similar estimates for the prediction uncertainty that arises due to the parameter uncertainty, and these results can be more accurate than GLUE. For the sediment transport model case, the mean forecasts from all three methods exhibited almost

the same NSCE value. However, the CRPS values indicated that the prediction uncertainty from GLUE provided unrealistic credible intervals, likely because it disregards parameter correlations.

The investigation of ELH should be expanded along several avenues in the future. First, ELH should be evaluated using additional tests. For example, ELH could also be applied to models that simulate watersheds, soil moisture profiles, water quality, or pipe networks. Specifically, distributed hydrologic models can have many more parameters to be determined by a modeler, which could have a higher dimension (e.g., $S > 20$) of the joint posterior PDF and/or include several optimal regions. The impact of the size of the calibration dataset on the performance of ELH estimates is also worth investigating. Second, the algorithmic variables included in ELH should be determined more formally to obtain robust selection criteria. Specifically, a larger LHS sample size B (or sample size multiplier n_s) will produce a finer discretization of the parameter space, which might reduce the histogram errors, but it will also increase the required simulations. In addition, the range adjustment multiplier n_w can be specified using a different formula that considers other posterior sample properties like the standard deviations and/or skewness. Third, the posterior density function used in ELH might not be formally applicable to some models. That equation has been criticized because model errors are often correlated, nonstationary, and non-Gaussian [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 algorithmic parameters that must be determined. In addition, the posterior density function could be generalized to consider multiple variables [Sabatine et al. 2015; Jung et al. 2018] because models in water resources engineering usually

generate multiple outputs of interest (e.g., discharge, soil moisture, water table elevations, etc.). Fourth, the method could be generalized to accept a non-uniform prior distribution in case the modeler has some prior knowledge of the parameter values. Fifth, future work could also focus on other sources of uncertainty such as the model's mathematical structure (including the selection of the transport equation), model input errors, and calibration data errors.

CHAPTER 3
MODELING INPUT ERRORS TO IMPROVE UNCERTAINTY ESTIMATES FOR A ONE-DIMENSIONAL SEDIMENT TRANSPORT MODEL[1]

3.1 Introduction

Numerical sediment transport models are widely used to make predictions of river morphological changes that result from natural and/or human influences. Those predictions are often used to help manage water systems, protect endangered species, etc. Uncertainty in the model predictions is important because it forces conservative designs that increase costs and potential conflicts between competing project goals. Traditionally, uncertainty has been assessed by examining how model predictions spread when variations are included in model inputs like 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 and critical shear stress [Chang et al. 1993; Yeh et al. 2004; Lai and Greimann 2010; Corazza et al. 2012], and constitutive model equations such as the transport formula [Wilcock 2001; Davies et al. 2002; Camenen and Larroudé 2003; Bertin et al. 2007]. Those traditional approaches are relatively simple to implement, but the distributions that are imposed on the model inputs are not directly derived from comparisons to calibration data. Inaccuracies in the imposed distributions have the potential to produce inaccuracies in the estimated uncertainty of the model predictions.

Bayesian inference provides a formal way to assess the uncertainty in the model predictions where the distributions for uncertain model elements are determined by comparing to calibration

---

[1] Jung, J. Y., J. D. Niemann, and B. P. Greimann (2017), Modeling input errors to improve uncertainty estimates for one-dimensional sediment transport models, Stochastic Environmental Research and Risk Assessment, 1-16, doi:10.1007/s00477-017-1495-8.

data. 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], and cohesive sediment behavior [Cho et al. 2016]. These Bayesian methods treat model parameters as random variables. The modeler selects a prior probability density function (PDF) for each parameter (usually uniform), which characterizes the uncertainty in the parameter value before model calibration is performed. Then, the posterior PDF is derived by evaluating the ability of parameter sets that are sampled from the prior PDF to reproduce available calibration data. Parameter sets that generate results that are similar to the observations have a higher likelihood of being correct. Various Markov Chain Monte Carlo (MCMC) algorithms can be used to generate and implement the parameter sets and iteratively obtain a numerical approximation for the posterior PDF [Green 2001]. Once an MCMC algorithm has performed sufficient simulations, the parameter sets are sampled from a stationary distribution, which is an estimate of the posterior PDF. Those generated parameter sets are then used in the forecast scenario to quantify how the uncertainty in the parameter values affects the model predictions. Key advantages of MCMC algorithms are that: (1) their sampling processes are efficient, which reduces computation time [Vrugt et al. 2003], (2) they use formal likelihood functions to generate the posterior PDF [van Griensven and Meixner 2007], and (3) they can infer correlations between model parameters, which can significant impact the uncertainty assessments [Li and Xu, 2014; Sabatine et al. 2015].

A key limitation in the Bayesian methods that have been used for sediment transport modeling is that they neglect other potential sources of uncertainty. Specifically, the forcing variables, boundary conditions (BCs), and geometric data used in these models such as discharges,

sediment supply rates, and channel geometries always include uncertainty due to their inherent unsteadiness, heterogeneity, and difficulty being measured [Wilcock 2001; Bunte and Abt 2005; Gaeuman et al. 2009]. For example, discharge data are considered to have 5% to 15% errors [Owen-Joyce and Raymond 1996], and flow depths can have errors with a standard deviation (SD) of 10% of the measured value in natural rivers [Sauer and Meyer 1992]. Accurate estimation of those variables is fundamental to successful 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, neglecting errors in the input data has been shown to lead to poor parameter uncertainty estimates as well as unreliable model predictions in various fields [Ajami et al. 2007; Vrugt et al. 2008; De Risi et al. 2017]. Similar issues may arise in sediment transport modeling.

Schmelter et al. [2011, 2012, and 2015] and Schmelter and Stevens [2013] developed a Bayesian method that considers input uncertainty in sediment transport modeling. They used a single transport equation, rather than a full sediment transport model, to predict the sediment transport rate for fluvial beds, and they applied a variance to the predictions in order to address the uncertainty due to mathematical approximations, input errors, and random variations. The variance parameter was calibrated jointly with a dimensionless shear parameter included in the formula using an MCMC algorithm. However, this approach cannot evaluate how much uncertainty in the model predictions comes from the uncertainty in the input data or identify the error associated with each forcing variable (because all sources are lumped together in a single variance parameter). Thus, the method is not well-suited for developing strategies to reduce the uncertainty in model predictions.

Other Bayesian methods have been developed to explicitly consider the errors in forcing variables for hydrologic models [Kavetski et al. 2002; Huard and Mailhot 2006; Ajami et al. 2007]. Kavetski et al. [2002] and Huard and Mailhot [2006] included forcing errors in their Bayesian methods, but their methods do not update the error distributions for the forcing variables based on the calibration data. In contrast, Ajami et al. [2007] modeled the input errors using a multiplicative approach and Gaussian distributions where the means and standard deviations (SDs) are treated as uncertain parameters and are estimated with the model parameters in the MCMC algorithm. However, Ajami et al. [2007] only applied the method to one uncertain forcing variable (rainfall), while sediment transport models have numerous uncertain inputs. Considering many uncertain inputs increases the dimensionality of the space that the MCMC algorithm must explore, which could greatly increase the computational time and make the approach impractical. In addition, the large number of parameters that are produced by this approach might make the model prone to over-calibration and thus less reliable for forecasting. Furthermore, the uncertain inputs in sediment transport modeling are also diverse, ranging from BCs to channel geometries, which makes the error modeling more complex.

The objectives of the present study are: (1) to assess the feasibility of the Ajami et al. [2007] method for determining the impacts of parameter and input uncertainty on sediment transport model predictions and (2) to determine whether uncertain inputs contribute substantially to the overall uncertainty in the predictions from these models. The uncertainty methodology is applied to the Sedimentation and River Hydraulics - One Dimension (SRH-1D) model and used to simulate net deposition volumes along the 23-km reach of the Tachia River in Taiwan. Nine model parameters and five types of inputs are considered uncertain in the experiments. Experiments are then performed to compare cases when input errors are neglected and included.

Feasibility is judged by examining: (1) the number of simulations required for MCMC algorithm convergence and (2) the performance of the model forecasts when input uncertainty is included. The impacts of input uncertainty are judged by considering: (1) the posterior PDFs of the model parameters, (2) the posterior PDFs of the input error parameters, and (3) the prediction ranges for the model forecasts.

3.2 Methodology for Uncertainty Assessment

3.2.1 Existing Parameter Uncertainty Method

In Bayesian inference, uncertain parameters are treated as random variables having a joint PDF:

$$p\left(\boldsymbol{\theta}|\mathbf{y}\right) \propto L\left(\mathbf{y}|\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right) \tag{25}$$

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 calibration dataset $\mathbf{y}$ when parameters $\boldsymbol{\theta}$ are used, and the prior PDF $p(\boldsymbol{\theta})$ summarizes the available information about the parameters $\boldsymbol{\theta}$ before considering any calibration data.

The Shuffled Complex Evolution Metropolis – Uncertainty Analysis (SCEM-UA) method is employed to estimate the joint posterior PDF $p(\boldsymbol{\theta}|\mathbf{y})$ in this study (see Vrugt et al. [2003] for a detailed discussion of this method). The SCEM-UA algorithm has been shown to require fewer model simulations to infer the posterior PDFs than other MCMC methods in various hydrologic applications [Laloy and Vrugt 2012]. This algorithm starts by generating an initial sample ($\approx$ 250) of parameter sets from a uniform prior joint PDF under the assumption that no information is available about the parameter values aside from their feasible limits before calibration. Then,

the parameter sets are used in model simulations of the calibration period, and the posterior density of each parameter set $\boldsymbol{\theta}_j$ is calculated as:

$$p\left(\boldsymbol{\theta}_j \middle| \mathbf{y}\right) \propto \left[\sum_{i=1}^{N}\left(M_i\left(\boldsymbol{\theta}_j\right) - y_i\right)^2\right]^{-\frac{1}{2}N} \tag{26}$$

where j is the index for the parameter sets, $M_i(\boldsymbol{\theta}_j)$ is the model's simulated value when the parameter set $\boldsymbol{\theta}_j$ is used at measurement location and/or time i, y_i is the observed value at i, and N is the number of observations in the calibration dataset $\mathbf{y}$. The equation contains a primary assumption that the model errors are mutually independent and Gaussian with constant variance [Box and Tiao 1973]. The value from the right side of Eq. (26) 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. (26). After that, the sample of parameter sets is partitioned into a number of complexes or groups (≈ 10), and the parameter sets in the complexes are updated in parallel using the Metropolis algorithm [Metropolis et al. 1953]. The highest posterior density parameter set in each complex is treated as the starting point of an 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 (see Vrugt et al. [2003]) with a covariance structure that is inferred from the parameters in the complex. The posterior density of the generated parameter set is then calculated by using it in the model and comparing the results to the calibration data. If the posterior density of the new parameter set $\boldsymbol{\theta}^*$ is larger than the current one $\boldsymbol{\theta}$, then the current one is replaced by the new parameter set. Otherwise, the current parameter set is retained, and the lowest posterior density member is replaced by the new parameter set with a specified probability. After a few iterations (≈ 5) of the updating procedure, the members of all

complexes are recombined, shuffled, and re-divided into complexes. As the procedure iterates, the parameters are being calibrated because the parameter values with higher posterior densities are sampled more frequently.

The iterative SCEM-UA algorithm converges when all complexes have approximately the same statistical properties. To diagnose convergence, the Scale Reduction Score (SRS) [Gelman and Rubin 1992] is calculated for each parameter. 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 achieved [Vrugt et al. 2003]. At that point, the samples are generated from an approximately stationary distribution, which is an estimate of the joint posterior PDF of the uncertain parameters.

3.2.2 Input Uncertainty Method

Input uncertainty is included in SCEM-UA following Ajami et al. [2007]. In this approach, an observed input value is assumed to be the true (but unknown) value but corrupted by a random error at each location and/or time. The random errors are modeled using a Gaussian distribution where the mean represents the measurement bias and the SD reflects the independent error in each measurement (Fig. 11). The Gaussian distribution is selected because individual errors are likely summed to produce the overall error in each measurement. In addition, the Gaussian distribution requires only two parameters, whose meanings are easy to interpret within the context of this analysis. A single distribution is used for all measurements of a given variable because space-time variations in the error distribution are not expected to be known in practice. This model can be written as:

$$\varepsilon_{ik} \sim N\left(m_k, \sigma_k\right) \tag{27}$$

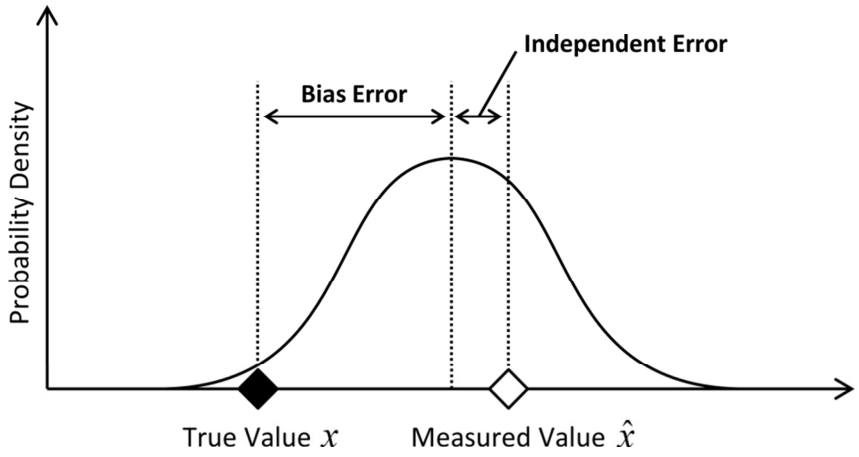

Fig. 11 Conceptual diagram for modeling input data error using a Gaussian distribution.

where ε_{ik} is the random error in measurement i for input variable k, which has a mean m_k and SD σ_k.

Five types of inputs that are commonly used for one-dimensional sediment transport models are considered uncertain in this study. First, discharge data are usually used as an upstream BC. The Owen-Joyce and Raymond [1996] stated that "fair" measurement performance has 95% of measured values within 10% of the true discharge value. This criterion suggests that discharge errors typically depend on the magnitude of the true discharge (larger discharge measurements exhibit larger errors). Such errors can be represented using a multiplicative model [Ajami et al. 2007] as follows:

$$\hat{x}_{ik} = x_{ik}\varepsilon_{ik} \tag{28}$$

where $\hat{x}_{ik}$ is the observed value and x_{ik} is the true value for measurement i of input variable k.

Second, rating curves are used to specify the water surface elevation for a given discharge at internal and/or downstream BCs. Rating curves inherently possess large uncertainties because they can vary in time as the channel geometry changes. An additive model is selected for rating curve data because it allows a possible shift in the curve due to bed degradation or aggradation. This model can be written:

$$\hat{x}_{ij} = x_{ij} + \varepsilon_{ij} \tag{29}$$

The third and fourth inputs are the horizontal and vertical coordinates of the channel cross-section points, respectively. The shape of the channel cross-section is obtained by measuring horizontal and vertical distances from a benchmark point. The horizontal distances along the cross-section can be over-estimated if the survey is not conducted perfectly orthogonal to the channel direction. This type of error can be represented by the following multiplicative model:

$$\hat{x}_{ij} = x_{ij} \frac{1}{\cos\left(\varepsilon_{ij}\right)} \tag{30}$$

where ε_{ij} is the angle between the surveyed line and the true orthogonal direction, and $\hat{x}_{ij}$ and x_{ij} represent the measured and true distances from the benchmark point, respectively. For vertical distances, errors might originate from the existence of dunes or antidunes, the uneven distributions of large bed materials (cobbles, rocks, and boulders), or an actively mobile bed. The Sauer and Meyer [1992] suggested that such depth measurements can have errors with a SD of 10% of the local flow depth in natural rivers, which implies that a multiplicative error model is appropriate. Thus, Eq. (28) is used to represent these errors.

Fifth, bank elevations are commonly used to define the longitudinal profile of the stream. Because those elevations are defined relative to an arbitrary datum (e.g., sea level), the errors in

such data are expected to be independent of the true value's magnitude. Accordingly, the additive model (Eq. (29)) is used to account for the errors in this variable.

The above error models have introduced a set of error parameters $\boldsymbol{\eta} = \{m_1, \sigma_1, \ldots, m_K, \sigma_K\}$, where K is the total number of input variables considered as uncertain. These new parameters are included in the main equation of the Bayesian uncertainty method (Eq. (25)), which becomes:

$$p(\boldsymbol{\theta}, \boldsymbol{\eta} | \mathbf{y}) \propto L(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\eta}) \, p(\boldsymbol{\theta}, \boldsymbol{\eta}) \tag{31}$$

where $p(\boldsymbol{\theta},\boldsymbol{\eta}|\mathbf{y})$ describes the uncertainty associated with model parameters and input data simultaneously given calibration dataset. To specify the posterior density $p(\boldsymbol{\theta},\boldsymbol{\eta}|\mathbf{y})$ in the SCEM-UA algorithm, the model is simulated using a sample of model parameters $\boldsymbol{\theta}$ and input variables x_{ik}. The input variables are determined by adjusting the measured values based on independent errors generated from Eqs. (28), (29), and (30) and the associated input error parameters $\boldsymbol{\eta}$.

3.3 Sediment Transport Model and Application Site

3.3.1 Sediment Transport Model

SRH-1D was developed and is extensively used by the U.S. Bureau of Reclamation to simulate flows and sediment transport in open channels and river networks. This section highlights the key model parameters (see Huang and Greimann [2013] for model details).

SRH-1D simulates flow hydraulics by solving the energy equation for gradually-varied flow, and it applies the standard step method and uses a time series of flow rates as an upstream BC. Manning's roughness n is a model parameter that must be specified to calculate the friction loss between two adjacent cross-sections in the energy equation.

Given the flow conditions, the Exner equation is used for sediment routing, and it calculates the changes in the sediment volume of the channel bed based on mass conservation, which includes the sediment porosity ϕ as a parameter. The volumetric sediment discharge Q_s is found by calculating the transport capacity separately for each predefined grain size class. The Wu et al. [2000] transport equation is used in this study because it is applicable for a wide range of sediment sizes (0.01 mm to 128 mm) and has better performance for the study site than the other formulas [Lai and Greimann 2010].

The Wu et al. [2000] equation computes total bed material load for each grain size class by combining the bed load and suspended load. The bed and suspended loads depend on the dimensional critical shear stress, which is computed as:

$$\tau_{c,t} = \theta_r \left[\left(\rho_s / \rho \right) - 1 \right] d_t \xi_t \tag{32}$$

where $\tau_{c,t}$ is the dimensional critical shear stress for class t, θ_r is the non-dimensional reference shear stress (a model parameter), ρ_s is the density of the sediment, ρ is the density of water, and d_t is the median diameter of class t. ξ_t 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. The hiding and exposure function is calculated:

$$\xi_t = \left(\frac{p_{h,t}}{p_{e,t}} \right)^{\lambda} \tag{33}$$

where λ is a hiding and exposure parameter and $p_{h,t}$ and $p_{e,t}$ are the hidden and exposed probabilities of particles in class t (see Huang and Greimann [2013] for details). The bedload

adaptation length multiplier b_L is a parameter used to determine the length over which the computed transport capacity is reached for bedload. The suspended sediment recovery factors for deposition (ζ_d) and scour (ζ_s) (both are parameters) determine the adaptation length for suspended load, depending on whether deposition or scour occurs at a given location.

SRH-1D simulates bed material mixing by dividing the bed into one active layer, which is a thin upper zone containing the bed material available for transport, and several inactive layers below the active layer. The thickness of the active layer is calculated by multiplying the geometric mean of the largest grain size class by a user-specified parameter, which is the active layer thickness multiplier n_{alt}. In addition, the weight of bed load fraction parameter χ controls the weighting of the grain size distribution for bed load when material transfers from the active layer to the inactive layers during deposition.

3.3.2 Application to the Tachia River

A 23-km reach of the Tachia River in Taiwan (Fig. 6) is used as a case study to evaluate the effects of input uncertainty. As a natural river, this case has more uncertainty from its inputs than flume experiments, which have been considered in past uncertainty analyses [Wu and Chen 2009; Ruark et al. 2011; Schmelter et al. 2011; Schmelter and Stevens 2013; Sabatine et al. 2015]. From Lai and Greimann [2010], net deposited/eroded sediment volumes were measured during two periods (2001 to 2005 and 2005 to 2009) at cross-sections along the selected reach from the dam to the ocean. This study uses the first period (2001 to 2005) as the calibration period and the second (2005 to 2009) as the forecast period, and the net deposition volume at each cross section is considered as the output variable of interest for the uncertainty analysis. All model and input uncertainty parameters are estimated by comparing simulation results to observations from the calibration period. Then, the calibrated parameters are applied to the

simulation of forecast period and the estimated prediction uncertainty is compared to the actual spread of the observations.

For the SRH-1D simulations, a time series of measured discharges at the Shih-Gang Dam provides the upstream BC (Fig. 12a). Flow and sediment inputs from tributaries downstream of Shih-Gang Dam are not considered here due to lack of the associated data. An internal BC is applied near the middle of the reach, where a grade control structure is located to slow the erosion. This BC is represented by a rating curve computed from HEC-RAS simulations and the 2005 geometry [Lai and Greimann 2010] (Fig. 12b). The downstream BC is specified using the critical depth obtained from the same HEC-RAS simulations (Fig. 12c). Channel geometry data were collected in both 2001 and 2005, and these datasets provide the initial bed geometry. The cross-sections were measured at 43 and 48 locations along the stream in 2001 and 2005, respectively. Each cross-section includes 32 to 205 elevation points where the horizontal and vertical coordinates are defined relative to a benchmark point (Fig. 12d). The benchmark is the ground elevation that corresponds to the bankfull water surface. The benchmark also provides the bank elevation that is used to define the longitudinal profile of the reach (Fig. 12e). 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.

Seven modeling cases are considered in this study. Case 0 implements the uncertainty analysis including only parameter uncertainty as a control case. Table 5 summarizes the feasible ranges for the nine SRH-1D parameters, which are determined following the section 2.4.2. Cases 1-6 also treat the model parameters as uncertain, but they include the uncertainty from a single, different input as shown in Table 6. In each of those cases, two input error parameters are added

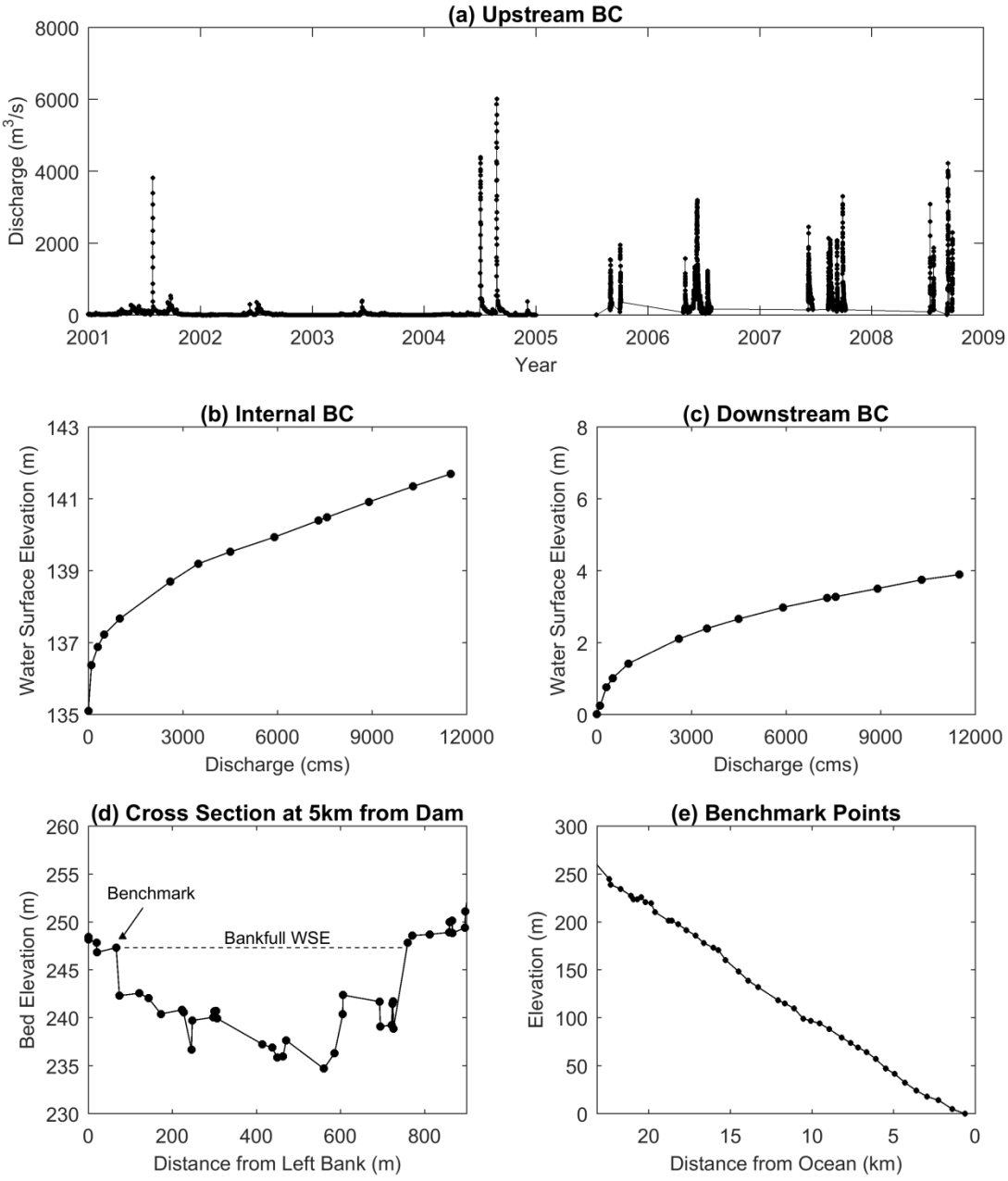

Fig. 12 Input data used in SRH-1D simulations of the Tachia River where the black circles represent the measured values. Subplot (d) also indicates the benchmark station, which corresponds to the bankfull water surface elevation (WSE).

Table 5 Uncertain model parameters of SRH-1D and their feasible ranges. All parameters are dimensionless except as noted.

Parameters	Ranges
Manning's roughness coefficient n	0.035 ~ 0.045 s $m^{-1/3}$
Reference shear stress θ_r	0.01 ~ 0.10
Hiding and exposure factor 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

Table 6 Error models used for the uncertain inputs, the feasible ranges for the error model parameters, and the SDs of the output produced by variations in the uncertain inputs.

Case	Uncertain Input Variables	Error Models	Ranges for m	Ranges for σ	Averaged SD of Outputs (m^3)
0	None				
1	Discharges at Upstream BC	$\hat{x} = x\varepsilon_1$	0.95 ~ 1.05	0 ~ 0.1	5,600
2	Rating Curve at Internal BC	$\hat{x} = x + \varepsilon_2$	-2 ~ 2	0 ~ 0.05	4,200
3	Rating Curve at Downstream BC	$\hat{x} = x + \varepsilon_3$	-2 ~ 2	0 ~ 0.05	11,700
4	Horizontal Points at Cross-sections	$\hat{x} = x/\cos\varepsilon_4$	0 ~ 15	0 ~ 15	8,700
5	Vertical Points at Cross-sections	$\hat{x} = x\varepsilon_5$	0.9 ~ 1.1	0 ~ 0.1	24,200
6	Benchmark Elevations	$\hat{x} = x + \varepsilon_6$	-2 ~ 2	0 ~ 0.5	44,400
7	All Data Used in cases 1~6				

to the nine model parameters. Subscripts will be used on the error parameters to denote the associated case (e.g., m_1 and σ_1 apply to Case 1 and describe the discharge error distribution as shown in Table 6). Case 7 includes the uncertainty from all nine model parameters and all six inputs so that this case involves a total of 21 parameters. The feasible ranges for the input error parameters (the means and SDs of the error distributions) are also shown in Table 6. These ranges are based on the available information about each type of measurement. Unrealistically wide ranges are avoided because they slow the convergence of the SCEM-UA algorithm and cause large computation times. Discharge data typically have 5% to 15% errors [Owen-Joyce and Raymond 1996], so the parameter ranges allow up to a 5% bias and a 10% SD. The rating curves were obtained from HEC-RAS simulations rather than measurements, so the ranges for their mean errors allow large biases. However, their SD ranges are narrow to ensure that the rating curves always monotonically increase. The ranges for the cross-section angle error parameters were selected to avoid angles greater than 45° with 95% confidence. The ranges for the vertical cross-section error parameters assume that the errors are typically up to 10% of the measured depth in natural rivers [Sauer and Meyer 1992]. Benchmark elevations can be measured using a handheld global positioning system (GPS) or more accurate surveying methods. Because the method used for the Tachia River is unknown, the parameter ranges allow relatively large uncertainties in the benchmark elevations.

To facilitate interpretation of the uncertainty analyses, simple sensitivity tests were conducted to determine the inputs that have the greatest impact on the net deposition volumes. For each inputs variable, one hundred combinations of the input error parameter values (m and σ) were generated within their feasible limits using factorial design [Saltelli et al. 1995]. For each m and σ combination, 10 sets of inputs errors were generated and used in model simulations. The

resulting SD of the simulated net deposition volumes (averaged among all cross-sections) was then calculated using all 1,000 simulations for a given input variable (Table 6). The benchmark elevations (Case 6) produce the largest variation in the model results. The associated input error parameters (m and σ) have relatively wide feasible ranges, and the elevations affect the channel slope, which plays a large role in determining the sediment transport.

3.4 Results and Discussion

3.4.1 Required Number of Simulations

The number of simulations required for SCEM-UA convergence for each case was determined based on the SRS (Fig. 13). Each bar in the figure shows the number of simulations required for the SRS of one parameter to go below 1.2. Comparing to Case 0, which only considers model parameter uncertainty, Cases 1, 2, 3, 4 and 6 require only about 1.5 to 2.5 times more simulations. These increases are small given that two input error parameters are added in each case, which implies two additional dimensions in the parameter space for the SCEM-UA algorithm to explore. The small increases suggest that the added variables have relatively predictable impacts on the model results (e.g., few local optima). In contrast, Case 5, which considers the vertical coordinates of the cross-sections, requires about 5 times more simulations to converge. In particular, σ_5 needs many simulations to converge because it only affects the independent errors at individual points in the cross-section, which have a small effect on the model behavior.

Case 7 requires many more simulations to converge because it includes many more uncertain variables than the previous cases, and some input error parameters (e.g., those associated with Case 5) are slow to converge. The SD for the vertical measurement error σ_5 requires the largest number of simulations to converge for the same reason described earlier. In addition, the mean

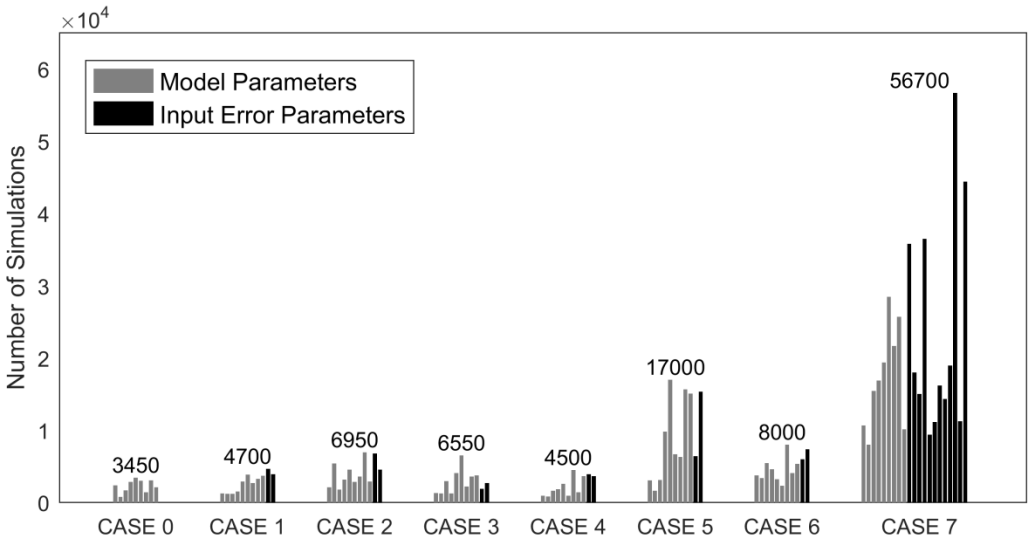

Fig. 13 Number of simulations required for SCEM-UA to converge for each model and input error parameter for the cases described in Table 2. The gray bars consider the model parameters, which are arranged in the same order as in Table 1. The black bars consider the input error parameters with m shown before σ in each case. The numbers above the bars indicate the number of simulations required for all parameters to converge in that case (i.e. the maximum bar height in each case).

for the discharge errors m_1, the SD for the internal rating curve errors σ_2, and the SD for the benchmark elevation errors σ_6 converge only after more than 35,000 simulations, which is substantially longer than any of the model parameters. In contrast, when those input errors are considered individually in Cases 1, 2, and 6, their input error parameters converge about as fast as the model parameters. The slower convergence in Case 7 likely occurs because those parameters interact with other parameters or have similar effects as other parameters in the model, which makes the posterior PDF more difficult to infer. For example, the parameter m_1 changes the flow rate, while σ_6 determines the local benchmark elevation errors, which affect the slopes between adjacent cross-sections. Therefore, those two parameters both affect the flow velocity, which ultimately affects sediment transport.

3.4.2 Uncertainty in Model Parameters

The marginal posterior PDF for each parameter describes the uncertainty in the parameter value that remains after model calibration is complete. It can be examined using the histogram of the parameter sets that are generated after SCEM-UA convergence. Fig. 14 presents the histograms for Manning's roughness along with the median value and the interquartile range (IQR), which is the difference between the 25% and 75% quantiles. The percentage shown in the figure is the ratio of the posterior and prior IQRs. The prior IQR is always 50% of the feasible range (Table 5) because the prior PDF is chosen to be uniform. The percentage can be interpreted as the fraction of the initial uncertainty that remains in the parameter after calibration.

For all the cases, the median value for Manning's roughness is around 0.039 (Fig. 14), which is reasonable for the conditions found in the Tachia River (i.e. the median grain size, vegetation,

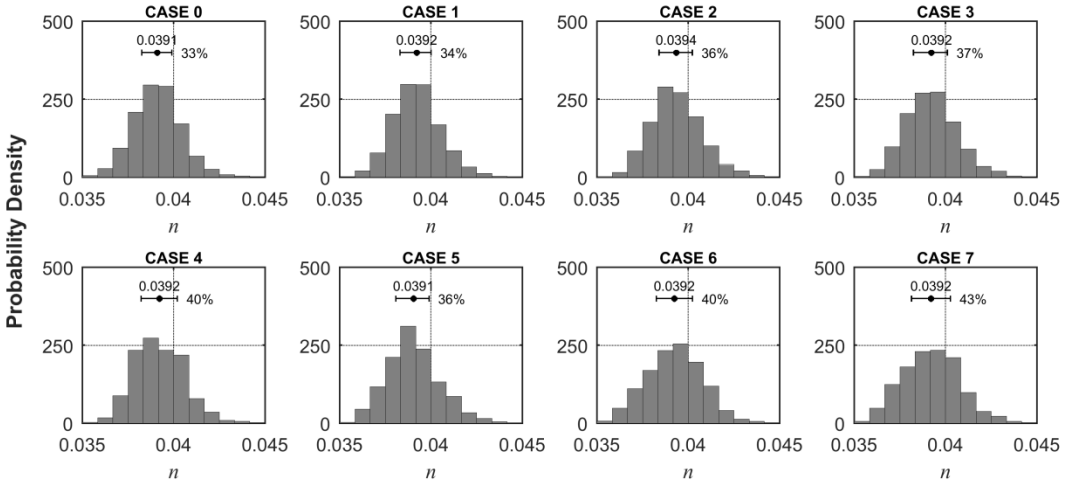

Fig. 14 Approximated posterior distributions for Manning's roughness using the histograms of 10,000 parameter sets generated after convergence. Also shown are the median value (black dot) and the interquartile range (horizontal line) of the roughness in each case. The percentage is the ratio of the posterior to the prior interquartile range.

and changes in cross-section shape) [Chow 1959; López and Barragán 2008]. The IQR ratios are between 33% and 43% for the cases, which suggests that Manning's roughness is reasonably well constrained by the calibration data. The reference shear stress θ_r, hiding and exposure coefficient λ, and active layer thickness multiplier n_{alt} are also well constrained and have the IQR ratios about 11%, 29%, and 58%, respectively (not shown). In previous research, these three parameters were found to have the highest impact on SRH-1D simulations of bed profile elevations in erosional flume experiments [Ruark et al. 2011], which are related to the deposition volumes considered here. Thus, the calibration data are expected to constrain their values. 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. The other five parameters (ϕ, ζ_d, ζ_s, b_L, and χ) have IQR ratios larger than about 80% (not shown), 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.

The histograms of Manning's roughness from Cases 1-3 are almost identical to the histogram for Case 0 (Fig. 14). These results could occur because those inputs contain little additional uncertainty, but this interpretation is contradicted by results presented later. They could also occur if the input errors in Cases 1-3 have little effect on the model outputs. The sensitivity test indicates that the changes in flow rate produce an average SD in net deposition volume of 5,600 m^3. However, a similar change can be achieved by varying the reference shear stress θ_c within 0.047~0.049, which is a small adjustment compared to the feasible range for that parameter. In addition, the rating curves considered in Cases 2 and 3 only affect the model results near those BCs.

The histograms for Manning's roughness from Cases 4-6 exhibit somewhat wider peaks and larger IQR ratios compared to Case 0 (Fig. 14). Case 7 shows the widest peak and largest IQR ratio among the cases considered. Specifically, the IQR ratio increases from 33% in Case 0 to 43% in Case 7. Similar behavior was observed for the other model parameters (not shown). These results suggest that model parameter uncertainty is somewhat underestimated when the effects of input uncertainty are neglected. The increase in model parameter uncertainty occurs because variations in model parameters can be accommodated somewhat by variations in the model inputs. Thus, a wider range of parameter values can still produce good model performance when input uncertainty is considered.

3.4.3 Uncertainty in Input Variables

Fig. 15 shows the posterior distributions for the error parameters that describe the bias in each input (m_1 to m_6). The solid lines show the results when the uncertainty of each input is considered independently in Cases 1 to 6, and those results are discussed first. The posterior distribution for the mean discharge error (m_1) centers around one, and a value of one indicates negligible bias in this input because the error is multiplicative. The mean internal BC rating curve error (m_2) is centered on zero, which also indicates negligible bias because this error is additive. Its high IQR ratio (88%) suggests that substantial uncertainty remains in m_2 after calibration. In contrast, the mean downstream BC rating curve error (m_3) is relatively well-constrained with an IQR ratio of 55% because it has a larger impact on the deposition volumes than m_2 (see sensitivities in Table 6). Changes in the downstream water surface elevation have a larger impact than changes in the internal water surface elevation because the channel width is larger downstream. Thus, the elevation changes produce larger changes in cross-sectional area and therefore larger changes in flow velocity and net depositional volume. The posterior

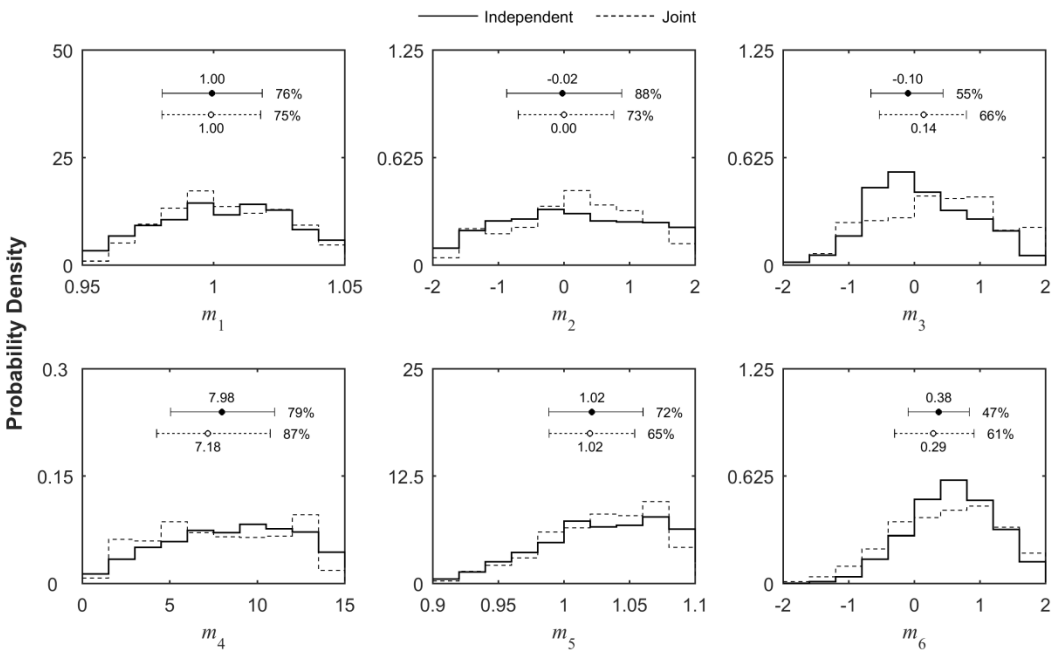

Fig. 15 Approximated posterior distributions for mean input error parameters m_1 to m_6 using the histograms of 10,000 parameter sets after convergence. Above each histogram are the median value (circle) and the interquartile range (horizontal line) for the error parameter. The percentage is the ratio of the posterior to the prior interquartile range.

distribution of m_3 is centered at -0.10, which suggests that the true water surface elevations in this rating curve are on average 0.10 m higher for any given discharge. An increase in the water surface elevation at the downstream BC reduces the water surface slope and likely deposits more sediment in the lower portions of the reach. The mean cross-section angle error (m_4) indicates a bias of about 8º, which implies the true channel widths are on average about 99% of the measured widths. This small change produces only minor modifications in the model behavior. The mean vertical measurement error (m_5) is centered on values larger than one, which indicates that the true vertical distances might be smaller than measured. This reduction would imply that the actual cross-sections are smaller than indicated by the original data, which would tend to increase the flow speed and erosion throughout the channel profile. The mean benchmark (or

bank) elevation error (m_6) is well-specified (IQR ratio of 47%) and larger than zero. Based on the analysis, the true elevations are on average about 0.38 m lower than the survey data. Lower elevations cause lower channel slopes upstream of the fixed BCs, which is expected to promote deposition in those sections of the reach.

When all input variables are jointly considered as uncertain in Case 7 (dashed lines in Fig. 15), most of the posterior distributions remain similar to the previous cases. However, the mean downstream BC rating curve error (m_3) changes noticeably. Specifically, the median for m_3 moves from -0.10 to 0.14, which suggests that the true water surface elevations in the downstream BC rating curve are lower than the measurements. This shift in m_3 occurs because of an interaction with m_6. When m_3 and m_6 were considered independently, both were estimated to promote deposition in the lower end of the reach. When they are considered together, m_6 still promotes deposition in the lower end of the reach, but m_3 no longer needs to serve this role and actually moves in the opposite direction.

The posterior distributions for the SD parameters σ_1 to σ_6 are shown in Fig. 16. None of these posterior distributions concentrate near zero, so the analysis suggests that all the considered inputs contain some independent errors. However, the IQR ratios for the SD parameters are typically about 80%, which means that the magnitude of the independent errors is uncertain. This result is expected. For example, the SD for the benchmark elevation error (σ_6) alters the local bed slopes by varying the elevations of adjacent cross-sections independently. Such changes affect the local patterns of erosion and deposition in the reach, but they are more difficult to infer than changes that affect large portions of the reach (e.g., m_6). The wide IQRs for the SD parameters are important to note. They suggest that the Bayesian method is unable to infer the distributions for these parameters from the calibration data and the prior distributions

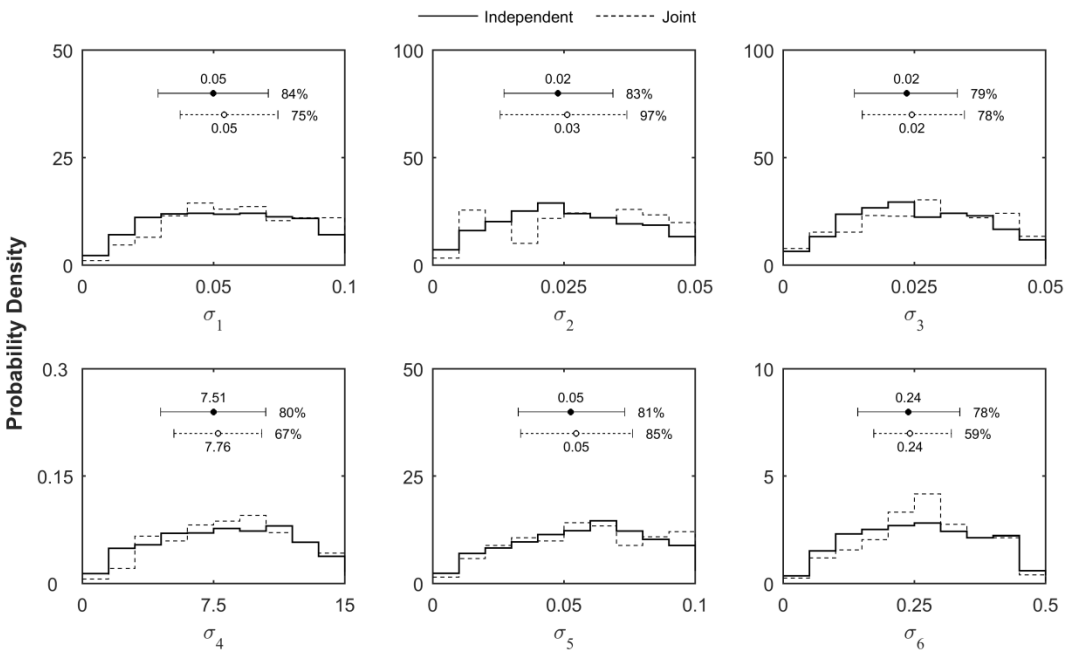

Fig. 16 Approximated posterior distributions for standard deviation error parameters σ_1 to σ_6 using the histograms of 10,000 parameter sets. Above each histogram are the median value (circle) and the interquartile range (horizontal line) for the error parameter. The percentage is the ratio of the posterior to the prior interquartile range.

continue to have a strong effect on the estimated prediction uncertainty. When all the input uncertainties are considered together (Case 7 and dashed lines in figure), the posterior distributions change only slightly. The largest change is seen for σ_6, which is better constrained when the parameters are considered together. It is also worth noting that the scale of the SD for the benchmark elevation errors (σ_6) is much larger than the scales of the other SDs.

3.4.4 Uncertainty in Predictions and Model Performance

The sediment deposition volumes for the forecast period (2005 to 2009) were simulated using a sample of 10,000 parameter sets generated after convergence for each case. 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. The simulations then use the estimated true inputs x, which are computed from the input error models using the sampled input error parameters. For each case in Fig. 17, the individual simulations are shown with gray lines, so the prediction uncertainty is visible from the width of the gray region.

When the input uncertainty is ignored (Case 0), the model consistently underestimates 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). The predictions also possess similar uncertainty throughout the reach (i.e. the width of the prediction range remains relatively constant). The estimated uncertainty from Case 1 is almost the same as Case 0 because the discharge data were found to be unbiased and have only small independent errors. Cases 2 and 3 have wider uncertainty ranges only near the cross-sections where the internal and downstream BCs are applied, respectively. Unlike Case 0, the range of predictions in Case 3 includes some deposition at the downstream end of the reach. As noted earlier, the mean downstream BC error m_3 tends to reduce the water surface slope in this portion of the reach, which enhances deposition. Case 4 again shows similar uncertainty to Case 0 because the inferred uncertainty in the cross-section angles is small. The prediction ranges from Case 5 are wider than Case 0 throughout the reach, and this additional uncertainty originates from the bias in the vertical cross-section elevations that was seen in Fig. 15. The uncertainty in Case 6 is much wider than the previous cases and suggests that deposition could occur for much of the river's length below the internal BC. The large uncertainty in this case occurs because of the substantial (0.38 m) bias that was seen in Fig. 15 and the relatively large scale of the independent errors (Fig. 16). The benchmark elevations also play an important role in determining erosion and deposition in the model simulations as shown by the sensitivity analysis in Table 6. As the benchmark elevations change, the slopes

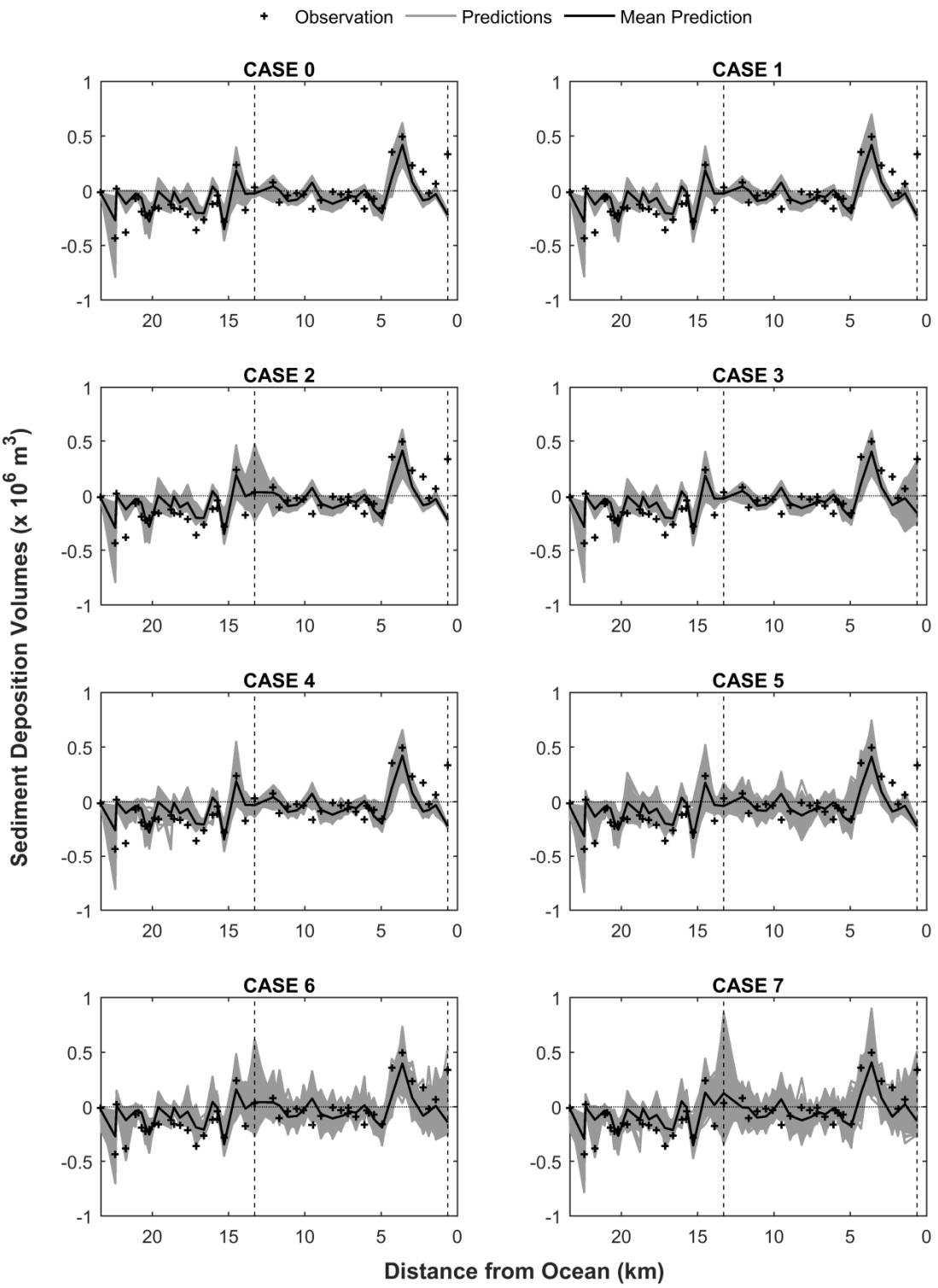

Fig. 17 Observations and 10,000 model predictions for sediment deposition volume in the forecast period. The dashed vertical lines show the locations of the internal and downstream BCs. The vertical width of the gray region is the prediction range and the black line is the mean prediction.

between adjacent channel cross-sections change. The slopes affect the bed shear stresses and thus the deposition volumes. Case 7 shows a blend of the prediction ranges for the previous cases. It has the generally wider range of Case 6 but larger uncertainty near the internal BC as seen in Case 2.

Fig. 18a shows the Nash-Sutcliffe Coefficient of Efficiency (NSCE) [Nash and Sutcliffe 1970] for the mean prediction of each case (shown as black lines in Fig. 17) during the calibration and forecast periods. NSCE can range from $-\infty$ to 1, where 1 means the model prediction perfectly matches the observations. For the calibration period, the NSCE typically increases when input uncertainty is included. This result is expected because the MCMC method has more parameters that it can adjust when input uncertainty is included. For the forecast period, the NSCE also typically improves as input uncertainty is considered. This result is more important because it implies that the increased flexibility is not leading to over-calibration and less reliable forecasts.

Fig. 18b compares the estimated prediction uncertainty from each case. To generate this figure, the width of the prediction range for each case was averaged over the reach. Then, the average width for each case was divided by the width from Case 0. The resulting values indicate the increase in the estimated uncertainty when each type of input uncertainty is included. When all input uncertainties are included (Case 7), the prediction uncertainty increases by 72% and 199% for the calibration and forecast periods, respectively. Thus, input uncertainty substantially contributes to the overall prediction uncertainty. Input uncertainty also makes larger contributions in the forecast period than the calibration period for all cases. This tendency might be due to the discharge data (Fig. 12a), which includes three or four large flood events every year during the forecast period (2005 to 2009) but only about one flood event each year during the calibration period (2001 to 2005). More sediment transport occurs during the forecast period

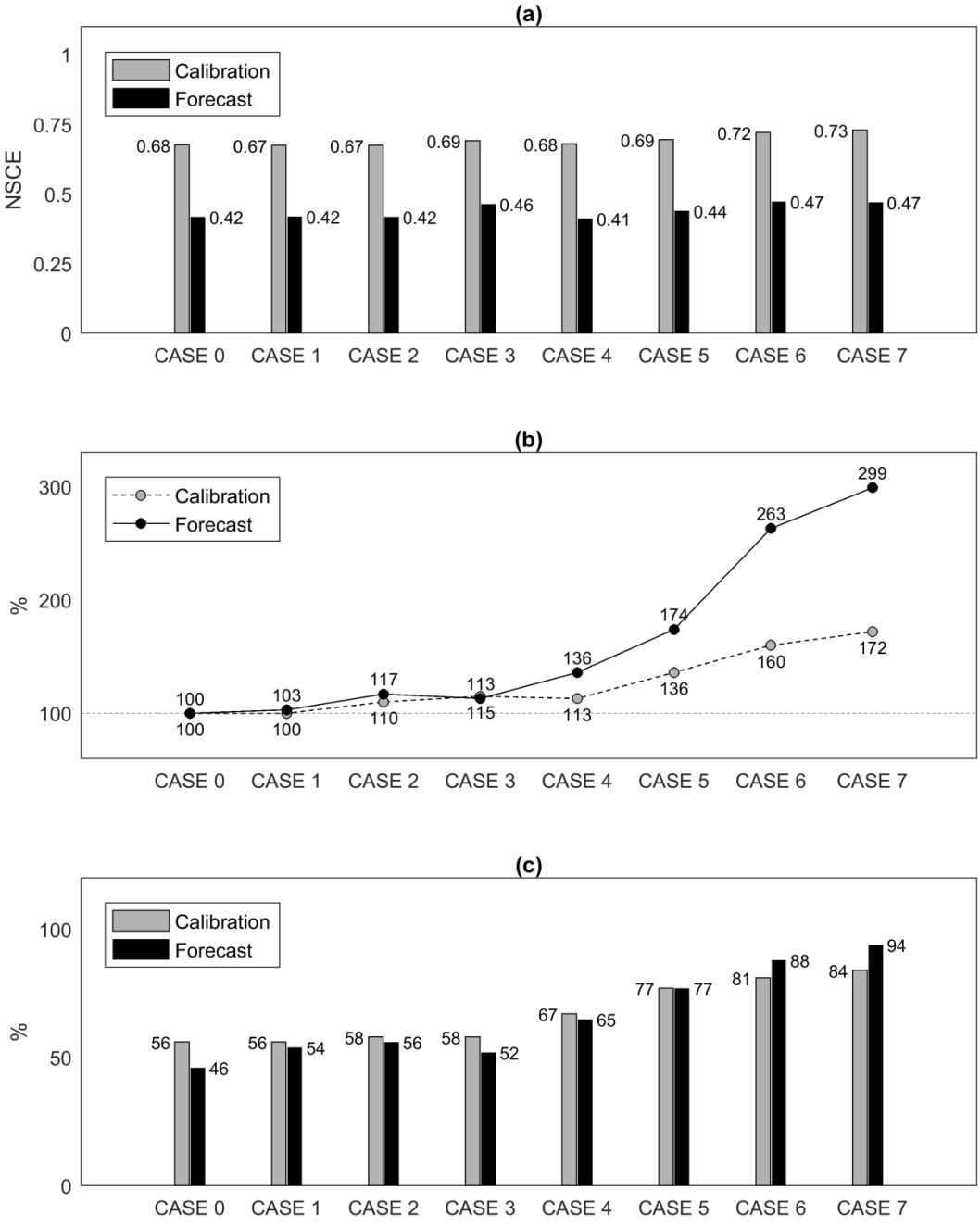

Fig. 18 (a) Nash-Sutcliffe Coefficient of Efficiency (NSCE) values for the mean prediction in each case, (b) averaged width of the prediction ranges from 10,000 simulations for each case relative to the average width for Case 0, and (c) percentage of observations covered by the prediction ranges of the same simulations.

than the calibration period, so the impacts of input uncertainty are generally larger in the forecast period.

Among the inputs considered, the benchmark elevation error (Case 6) makes the largest contribution to the prediction uncertainty in both the calibration and forecast periods (Fig. 18b). When only this single source of input uncertainty is considered, the prediction uncertainty increases by 60% and 163% for calibration and forecast periods, respectively, compared to Case 0. These increases are similar to those for Case 7, which includes all the uncertain inputs. This result suggests that the best way to reduce the uncertainty in the model predictions is to improve the reliability of the elevation data.

The percentages of the observations for deposition volume that are covered by the prediction ranges are shown in Fig. 18c. If the uncertainty has been accurately estimated and the net deposition volume observations contain no error, the prediction ranges should cover all the observations (more strictly, 99.98% or 9,998 of 10,000 observations given the sample size of 10,000 simulations and using Weibull plotting position [Nelson 1982]). Case 0 neglects substantial uncertainty because it only covers 56% and 46% of the observations for the calibration and forecast periods, respectively. The prediction ranges for Case 7 cover 84% and 94% of the observations in the calibration and forecast periods, respectively, which is the almost twice coverage of Case 0. These results suggest that considering the input uncertainties not only widens the prediction intervals but also improves the accuracy of uncertainty estimates. Case 6, which only includes the single largest source of input uncertainty, covers the 81% and 88% of the observations in the calibration and forecast periods, respectively, which makes it a reasonable approximation of Case 7. The remaining uncovered observations in Case 7 (about 6%) might be due to errors in the measurements of net deposition volume, which are not considered in this

research. Other reasons could be deviations from the assumptions used in the SCEM-UA algorithm or the input error models or flaws in the hydraulic and sediment transport model itself (e.g., the use of the Wu [2000] equation for sediment transport capacity).

3.5 Conclusions

This research evaluated the feasibility of the Ajami et al. [2007] method to include input uncertainties when estimating the uncertainty in the predictions from a one-dimensional sediment transport model. It then assessed the contribution of input uncertainty to the overall prediction uncertainty. While the magnitude of the input uncertainty will vary from case to case, the methodology presented here can be used to evaluate this uncertainty. For the cases considered in this study and given the assumptions of input uncertainty, the following conclusions can be drawn:

1. Including input uncertainty can require many more model simulations than considering parameter uncertainty alone, but the required number of simulations can be reduced by only considering the most important sources of input uncertainty. When all the types of input uncertainty were included in the analysis, the SCEM-UA algorithm required 15 times more model simulations to converge than when considering only parameter uncertainty. This large increase in simulations is partly due to difficulty inferring the posterior PDFs for inputs that have little impact on the model results. When only the single most important source of input uncertainty was included in the analysis, the required number of simulations increased only by a factor of 2. Furthermore, the estimated prediction uncertainty for this simplified case was similar to the case that considered all the uncertain inputs.

2. Including input uncertainty does not necessarily reduce the reliability of the mean model predictions for the forecast period. When input uncertainty is considered, it requires estimating up to 21 parameters, which is 12 more than when only parameter uncertainty is included. This large degree of flexibility might result in over-calibration of the model to the available observations. However, for the cases considered, the NSCE of the forecasts increases slightly when input uncertainty is included.

3. Considering input uncertainty does not necessarily change the estimated values of the model parameters, but it can modify the estimated uncertainty of the parameter values. For example, the median estimate of Manning's roughness was around 0.039 for all cases considered, and similar stability was observed for the values of other important parameters. However, the ratio of the posterior and prior IQRs for Manning's roughness was 33% when input uncertainty was neglected and 43% when all types of input uncertainty were included. This increase occurs because variations in the inputs allow wider ranges of parameter values to produce good model performance.

4. Including input uncertainty can produce much larger estimates of the prediction uncertainty. When all the uncertain inputs were included, the prediction uncertainty increased by 72% and 199% for the calibration and forecast periods, respectively. For the Tachia River model, the primary source of input uncertainty was the benchmark elevations, which control the longitudinal profile of the reach. When only this uncertain input was included, the prediction uncertainty increased by 60% and 163% for the calibration and forecast periods, respectively, compared to the case without input uncertainty.

5. Including input uncertainty can produce much more accurate estimates of the prediction uncertainty. When only parameter uncertainty was included for the Tachia River, the range of predictions for the forecast period covered only 46% of the deposition volume observations. However, when all six uncertain inputs were included, the range of predictions covered 94% of the observations for the forecast period.

Overall, the results support the feasibility of using the Ajami et al. [2007] approach for models with larger numbers of uncertain inputs. Furthermore, it also supports including input uncertainty when assessing the prediction uncertainty for sediment transport models. The research described in this paper should be expanded along several avenues in the future. First, the impact of uncertainty in other inputs such as bed material size, bed elevation control structures, sediment supply, and tributary inflows, etc., can be evaluated by applying the proposed input error models or developing new ones. Second, this approach can also be applied and tested for other case studies such as river networks, where the several reaches are linked and interact, or reservoir sedimentation cases. Third, future work should also consider other sources of uncertainty such as the model's mathematical structure (including the selection of the transport equation) and the observations used for calibration. Fourth, future research should focus on improving the efficiency of the uncertainty method. The MCMC algorithm requires too many simulations for use with complex models given present computational capabilities. A single simulation of the Tachia River took only 20 seconds to run, but 7 days of continuous computation time with four parallel computing processors were required to achieve SCEM-UA convergence for the case that considered all six uncertain inputs. The computational times could increase tremendously when assessing the uncertainty in the predictions for more complex and higher dimensional modeling cases.

CHAPTER 4
COMBINING PREDICTIONS AND ASSESSING UNCERTAINTY FROM SEDIMENT TRANSPORT EQUATIONS USING MULTIVARIATE BAYESIAN MODEL AVERAGING[2]

4.1 Introduction

Sediment transport models are widely used to predict the impacts of potential river restoration activities, but the predictions from these models always possess uncertainty. One major source of uncertainty is the mathematical equation that is used to compute sediment transport capacity. Several equations have been empirically developed for different fluvial conditions by applying various simplifications and assumptions. For example, some equations are applicable for a certain ranges of sediment size such as sand [Yang 1973; Yang 1979], gravel [Yang 1984], and sand with a high concentration of fine particles [Yang 1996]. Some equations are also designed for different transport types including bed load [Meyer-Peter and Müller 1948; Einstein 1950; Rottner 1959; Bagnold 1980; Parker 1990; Ribberink 1998; Wilcock and Crowe 2003] and total material load [Laursen 1958; Engelund and Hansen 1972; Ackers and White 1973; Brownlie 1981; van Rijn 1989; Wu et al. 2000]. In addition, some equations are revisions of earlier equations to accommodate data from additional flume experiments [Wallingford 1990; Wong and Parker 2006] and natural rivers [Madden 1993; Yang 1996; Gaeuman et al. 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 shown that different equations provide different predictions even when the equations are calibrated using same data [Wilcock

---

[2] Jung, J. Y., J. D. Niemann, and B. P. Greimann (2018), Combining predictions and assessing uncertainty from sediment transport equations using multivariate Bayesian model averaging, Journal of Hydraulic Engineering, 144(4), 04018008, doi:10.1061/(ASCE)HY.1943-7900.0001436.

2001; Pinto et al. 2006; Bertin et al. 2007; Schmelter et al. 2011], but little consideration has been given to quantifying the uncertainty originating from the sediment transport equation in model forecasts.

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. These methods usually combine the predictions from a set of competing models and provide an average forecast along with its credible interval (CI) to characterize the uncertainty. Several multi-model averaging techniques are available including: 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]. Among the available methods, BMA has been shown to produce the most accurate predictions and more realistic descriptions of the predictive uncertainty in various hydrologic modeling studies [Clyde 1999; Viallefont et al. 2001; Ye et al. 2004; Sloughter et al. 2010]. BMA typically represents the uncertainty associated with each competing model using a normal distribution that is centered on that model's prediction, and it combines the distributions of the models by a weighted average (Fig. 19). The weights and standard deviations for the normal distributions are estimated to maximize the combined likelihood of the observed (calibration) values of the variable of interest. BMA assigns higher weights to better performing models, and it accounts for both the uncertainty due to the model selection (represented by the weights and the spread in the model predictions) and the uncertainty associated with each model (represented by the normal distributions) [Vrugt and Robinson 2007].

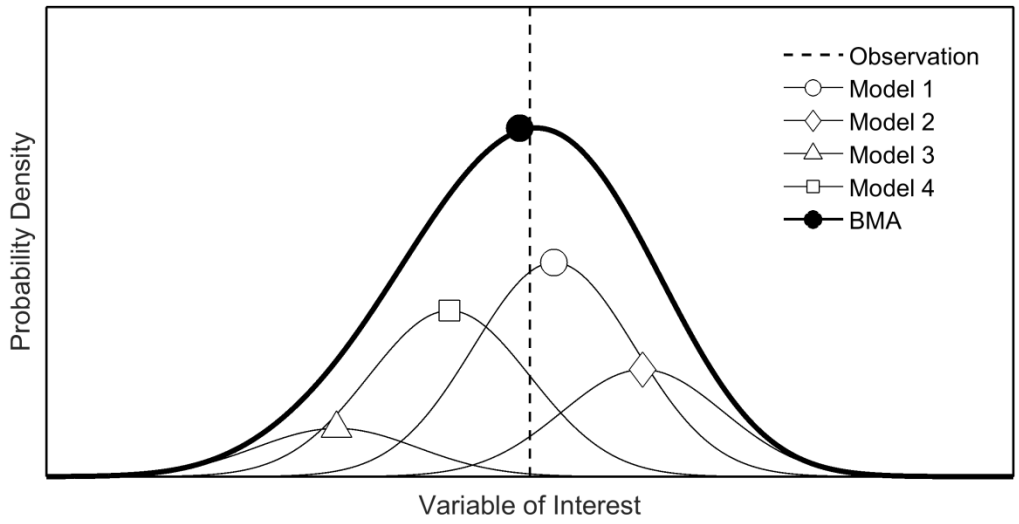

Fig. 19 BMA distribution generated by a weighted-average of normal distributions from four competing models. The white markers are the predictions from the competing models, the black marker is the BMA prediction, which is same as the weighted-average of individual model predictions, and the lines present the probability density function of each prediction (adapted from Raftery et al. [2005]).

Sabatine et al. [2015] recently applied BMA to sediment transport modeling, but they only considered cases with a single output variable because the BMA likelihood function is not easily generalized to multiple variables. However, if BMA maximizes the likelihood for a single variable, other variables might be estimated poorly by the BMA model. Moreover, probabilistic information such as CIs cannot be obtained for the other variables. This limitation is critical 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 [Huang and Greimann 2010; Russel et al. 2010; Ahn et al. 2013; Ahn and Yang 2015]. Several functions have been suggested to compute likelihoods for multiple objectives or variables [Beven and Binley 1992; Yapo et al. 1998; Mo and Beven 2004; Chahinian and Moussa 2007; Beven 2011]. For example, van Griensven and Meixner [2007]

proposed the Global Optimization Criterion to compute likelihoods from multiple variables by weighting them based on the number of observations divided by the minimum of the sums of the squared residuals. 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 for BMA because they compute the likelihoods based on residuals, which are the differences between model outputs and observations, while BMA likelihoods are calculated using the conditional probability density at the observed value.

Another limitation in the application of BMA to sediment transport models is BMA's assumption that the same standard deviations apply to a variable at all locations and times (homoscedasticity). While this assumption is expected to hold for a variable like bed elevation, which is often defined relative to an arbitrary datum, the uncertainty in a variable like sediment transport rate is expected to depend on the variable's magnitude. In such a case, if the standard deviations are estimated from a calibration period with small values, BMA might underestimate the predictive uncertainty for a forecast scenario with large values. A number of methods have been proposed to consider heteroscedasticity in BMA, particularly for hydrologic modeling, 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]. However, those methods typically focus on calibration datasets that include a large number of times, which is unusual for sediment transport models of natural rivers.

The objective of this research is to develop and test a multivariate version of BMA to assess the uncertainty associated with the selection and application of a transport equation in one-dimensional sediment transport models. A likelihood function is suggested to consider multiple

output variables of interest, and the standard deviations used in BMA are allowed to vary with the magnitude of the variables if needed. To test the multivariate BMA, two published flume experiments including a depositional case [Seal et al. 1997] and an erosional case [Pender et al. 2001] with non-cohesive material are used. Four sediment transport equations included in the Sedimentation and River Hydraulics – One Dimension (SRH-1D) [Huang and Greimann 2013] model are used to simulate the flume cases, and each equation is treated as a separate model. The multivariate BMA forecasts are then compared to the forecasts from individual models and univariate BMA based on the accuracy of their predictions and the coverage of the observations by their CIs.

4.2 Methodology

4.2.1 Existing BMA Model

BMA [Hoeting et al. 1999] defines the uncertainty in the model prediction for variable Δ using a probability density function (PDF) $p(\Delta|\mathbf{D},\mathbf{y})$, which is the posterior distribution of Δ given a set of competing models $\mathbf{D} = \{D_1, D_2, \ldots, D_A\}$ and calibration dataset $\mathbf{y}$. This PDF can be written:

$$p\left(\Delta\middle|\mathbf{D},\mathbf{y}\right)=\sum_{a=1}^{A} p\left(\Delta\middle|D_a,\mathbf{y}\right)p\left(D_a\middle|\mathbf{y}\right) \tag{34}$$

where A is the number of competing models, $p(\Delta|D_a,\mathbf{y})$ is the posterior distribution from model a, and $p(D_a|\mathbf{y})$ is the posterior probability that reflects how well model a fits the dataset $\mathbf{y}$. Because the posterior probabilities of the competing models always sum to one, they can be considered as weights. Thus, one can view the BMA PDF as a weighted average of the posterior distributions for the individual models.

BMA has been applied to dynamic models by assuming that: (1) the forecast M_a is the most likely result from model a, and (2) the uncertainty associated with each model can be represented using a normal distribution that is centered on the prediction M_a [Raftery et al. 2005]. To obtain the forecasts, the parameters of each model are calibrated using the dataset **y**. From these assumptions, the posterior distribution of Δ given a set of competing model predictions **M** = {M_1, … , M_A} is:

$$p\left(\Delta\middle|\mathbf{M}\right)=\sum_{a=1}^{A}w_a p\left(\Delta\middle|M_a,\sigma_a\right) \tag{35}$$

where w_a is model a's weight and $p(\Delta|M_a,\sigma_a)$ for model a is found from a normal distribution with mean M_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 likelihood of the observed values over all locations and times in the calibration dataset **y**. Using the log-likelihood L_{BMA} for algebraic simplicity and numerical stability [Raftery et al. 2005; Vrugt et al. 2008] and assuming that the residuals of each model are independent, one can write:

$$L_{\text{BMA}}=\sum_{i=1}^{N}\log\left[\sum_{a=1}^{A}w_a g\left(y_i\middle|M_{ai},\sigma_a\right)\right] \tag{36}$$

where $g(y_i|M_{ai},\sigma_a)$ is the probability density of the observation y_i given model a's prediction M_{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. This method is easy to implement and ensures that the model weights w_a are always positive and add up to one [Givens and Hoeting 2012].

BMA then applies the 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 period [Raftery et al. 2005; Vrugt et al. 2008; Sabatine et al. 2015]. The mean of each model's normal distribution changes as the models make their forecasts, but the other quantities remain fixed. A deterministic BMA prediction can be obtained from the expectation of the BMA PDF in Eq. (35), which is the same as the weighted average of the predictions from the competing models [Raftrey et al. 2005]:

$$E\left(\Delta\middle|\mathbf{M}\right)=\sum_{a=1}^{A}w_aM_a \tag{37}$$

The uncertainty contained in the BMA prediction can be described using the CI, which quantifies the predictive uncertainty including both the uncertainty due to the model selection and the uncertainty associated with each model. The CI is calculated numerically from the BMA PDF in Eq. (35) for a selected level of probability (e.g., 90%).

4.2.2 Multivariate BMA Model

The multivariate BMA model uses a modified likelihood function that includes multiple variables within a single formula. This method applies two primary assumptions: (1) the likelihood of a model being correct is based on all variables of interest, which implies each model has same weight w_a for all variables, and (2) the uncertainty associated with each model forecast depends on the variable being considered, which implies that each variable k can have a different standard deviation σ_{ak}. The multivariate log-likelihood L_{mBMA} is calculated as:

$$L_{\text{mBMA}}=\sum_{k=1}^{K}\left\{\frac{1}{N_k}\sum_{i=1}^{N_k}\log\left[\sum_{a=1}^{A}w_ag\left(y'_{ik}\middle|M'_{aik},\sigma'_{ak}\right)\right]\right\} \tag{38}$$

where K is the number of variables considered, N_k is the number of observations for variable k in the calibration dataset, and $g(\cdot)$ represents the probability density of reproducing the normalized observation y'_{ik} given model a's normalized prediction M'_{aik} with standard deviation σ'_{ak} for the normalized variable k at a measurement i. In Eq. (38), the likelihood of each variable k is divided by N_k 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 adopted as follows:

$$y'_{ik} = \frac{y_{ik}}{y_{k,50}} \tag{39}$$

$$M'_{aik} = \frac{M_{aik}}{y_{k,50}} \tag{40}$$

where y_{ik} is the observation, M_{aik} is model a's prediction for variable k at a measurement i, and $y_{k,50}$ is the median of the observed variable k. 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'_{ak} has a normalized scale, it needs to be transformed to the original scale as:

$$\sigma_{ak} = \sigma'_{ak} y_{k,50} \tag{41}$$

Other normalization methods were considered (e.g., normalizing by the mean or the interquartile range) and found to produce similar results for the experiments that follow. The w_a and σ_{ak} values from the EM algorithm are then applied to Eq. (35) to generate the BMA PDF separately for each variable.

The multivariate BMA model can allow the uncertainty in a model prediction to vary with the magnitude of the prediction. For user-selected variables, this modification is accomplished by assuming that the coefficient of variation (CV) rather than the standard deviation is constant for all locations and times. The CV is defined as the ratio of the standard deviation to the mean. This approach follows the idea of a linear dependence of the standard deviations on the mean [Vrugt and Robinson 2007]. For variables where the CV is assumed constant, the standard deviation for the given variable and model can vary depending on the model prediction:

$$\sigma'_{aik} = \mathrm{CV}_{ak} M'_{aik} \tag{42}$$

where CV_{ak} is model a's CV for variable k, and σ'_{aik} is model a's standard deviation for the normalized variable k at a measurement i. For the mixed case that applies the constant standard deviation for one variable and a constant CV for another, the log-likelihood is:

$$L_{\mathrm{mBMA}} = \frac{1}{N_1}\sum_{i=1}^{N_1}\log\left[\sum_{a=1}^{A} w_a g\left(y'_{i1}\middle|M'_{ai1},\sigma'_{a1}\right)\right] + \frac{1}{N_2}\sum_{i=1}^{N_2}\log\left[\sum_{a=1}^{A} w_a g\left(y'_{i2}\middle|M'_{ai2},\mathrm{CV}_{a2}\right)\right] \tag{43}$$

where the subscripts 1 and 2 are the indices for each variable. After obtaining the best model weights and the CVs, the BMA PDF for this variable can be generated as:

$$p\left(\Delta\middle|\mathbf{M},\mathbf{y}\right) = \sum_{a=1}^{A} w_a p\left(\Delta\middle|M_a,\mathrm{CV}_a\right) \tag{44}$$

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. Flume experiments rather than natural rivers were chosen because the channel geometry, flow rate, sediment supply, and bed materials are well-documented, which reduces the uncertainty from sources other than the model itself. The cases also allow separate consideration of scour and sedimentation processes.

The Seal et al. [1997] experiment was designed to investigate downstream fining and sediment sorting during aggradation in narrow channels. The flume had a rectangular shape, was 0.3 m wide and 45 m long, and had an initial slope of 0.002. No initial bed material was used and water discharge was steady at 0.049 m^3/s for 64 hours. At the upstream end, sediment was supplied at 0.047 kg/s during the experiment, and the material was a mixture of sand and gravel ranging from 0.125 mm to 65 mm with a D50 of 5 mm. Bed elevations were measured at 18 locations along the flume length every 4-5 hours, and D50 of the deposited material was measured along the flume length at five different times. Both bed elevations and D50 values are considered variables of interest for the BMA modeling, and the dataset is divided into a calibration period (hours 0 to 32) and forecast period (hours 32 to 64) (Table 7).

The Pender et al. [2001] experiment was designed to simulate bed degradation and investigate changes in transport rate. The flume had a trapezoidal shape with 45° side slopes, was 0.8 m wide at its base and 20 m long, and had an initial slope of 0.0026. The initial bed material was a mixture of sand and gravel ranging from 0.25 mm to 22.63 mm. Water discharge was steady at

Table 7 Number of available observations from the calibration and forecast periods of two experiments.

Variable	Seal et al. [1997]		Pender et al. [2001]	
	Calibration	Forecast	Calibration	Forecast
Bed Elevation	126	87	168	203
D50 Size	46	18	-	-
Transport Rate	-	-	17	17

0.117 m^3/s for 84.6 hours, and no sediment was supplied. Bed elevations were measured using a 2-3 hour time increment at 21-42 locations, and bed load transport rate was measured 5 m from the downstream end during the experiment. Both bed elevation and sediment transport rate are considered as variables of interest, 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 7). The calibration period is shorter than the forecast period because bed degradation is much quicker early in the experiment.

4.3.2 Sediment Transport Equations

SRH-1D 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 [Huang and Greimann 2013]. Given the flume experiment configurations, the model is applied for steady flow, non-cohesive sediment, and a fixed-width channel. Four equations in SRH-1D are treated as separate sediment transport models for BMA modeling in this study, and the following paragraphs describe the mathematical details of those equations for computing sediment transport capacity.

The Parker [1990] equation was developed to compute bed load transport capacity as:

$$q_{b,t} = \frac{p_t \left(\tau_g / \rho\right)^{1.5}}{g\left[\left(\rho_s / \rho\right) - 1\right]} F\left(\varphi_t\right) \tag{45}$$

where $q_{b,t}$ is volumetric bed load transport rate per unit width, p_t is the fraction of material in grain size class t, τ_g is grain shear stress, ρ is density of water, ρ_s is density of the sediment, and the empirical function $F(\varphi_t)$ was determined from field experiments with mixed size gravel (2 mm to 64 mm), where:

$$F\left(\varphi_t\right) = \begin{cases} 11.93\left(1 - 0.853/\varphi_t\right)^{4.5} & \varphi_t > 1.59 \\ 0.002183 \exp\left[14.2\left(\varphi_t - 1\right) - 9.28\left(\varphi_t - 1\right)^2\right] & 1 < \varphi_t \le 1.59 \\ 0.002183 \varphi_t^{14.2} & \varphi_t \le 1 \end{cases} \tag{46}$$

where φ_t is computed as:

$$\varphi_t = \frac{\theta_t}{\xi_t \theta_r} \tag{47}$$

and θ_t is the Shields parameter, θ_r is non-dimensional reference shear stress, and ξ_t is the hiding and exposure function. The Shields parameter is computed as:

$$\theta_t = \frac{\tau_g}{\left(\rho_s - \rho\right) g d_t} \tag{48}$$

where d_t is the median grain diameter of class t. The function ξ_t 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:

$$\xi_t = \left(d_t / d_{50}\right)^{-\lambda} \tag{49}$$

where λ is a hiding and exposure coefficient. The Parker equation assumes that the θ_r and λ are constant with flow and the grain size distribution.

Wilcock and Crowe (W&C) [2003] formulated an expression for the bed load transport capacity following the formation of Eq. (45) and defined the function $F(\varphi_t)$ using a mixture of sand and gravel (0.5 mm to 64 mm) as:

$$F\left(\varphi_t\right)=\begin{cases}14\left(1-0.894/\sqrt{\varphi_t}\right)^{4.5} & \varphi_t \geq 1.35 \\ 0.002002\varphi_t^{7.5} & \varphi_t < 1.35\end{cases} \tag{50}$$

The W&C equation allows the non-dimensional reference shear stress θ_r to depend on the fraction of sand within the bed as:

$$\theta_r = \theta_{r0} + 0.015\left[\exp\left(-20F_s\right)\right] \tag{51}$$

where θ_{r0} is a datum reference shear stress, and F_s is the fraction of bed material sand sized or less. The datum reference shear stress is the lowest value θ_r can have, and θ_r approaches this value as F_s becomes large. The hiding and exposure function is defined using the geometric mean particle diameter d_m as:

$$\xi_t = \left(d_t/d_m\right)^{-\lambda} \tag{52}$$

and the coefficient λ was determined as:

$$\lambda = 1-\left(1-\lambda_0\right)\left[1+\exp\left(1.5-d_t/d_m\right)\right]^{-1} \tag{53}$$

where λ_0 is a datum hiding and exposure coefficient, and the value of λ in Eq. (53) approaches λ_0 for large d_t/d_m and approaches 0.88 for small d_t/d_m [Wilcock and Crowe 2003].

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

$$q_{b,t} = 3.97 g^{0.5} \left[\left(\rho_s / \rho \right) - 1 \right]^{0.5} d_t^{1.5} \left[\frac{RS_f}{\left[\left(\rho_s / \rho \right) - 1 \right] d_t} - 0.0495 \right]^{1.5} \tag{54}$$

where R is hydraulic radius, and S_f is energy slope.

Unlike the three bed load equations, the Wu [2000] equation was designed to compute the transport capacities of bed load and suspended load separately and combine them to obtain total bed material load. The equations of transport capacity were developed for the bed load $q_{b,t}$ using a mixture of sand and cobbles (0.062 mm to 128 mm) and for the suspended load $q_{s,t}$ using a mixture of silt and sand (0.01 mm to 2 mm) as follows:

$$q_{b,t} = 0.0053 p_t \sqrt{g \left[\left(\rho_s / \rho \right) - 1 \right] d_t^3} \left[\left(\frac{0.05 {d_{50}}^{1/6}}{n} \right)^{1.5} \frac{\tau_b}{\tau_{c,t}} - 1 \right]^{2.2} \tag{55}$$

$$q_{s,t} = 0.0000262 p_t \sqrt{g \left[\left(\rho_s / \rho \right) - 1 \right] d_t^3} \left[\frac{U}{w_{f,t}} \left(\frac{\tau_b}{\tau_{c,t}} - 1 \right) \right]^{1.74} \tag{56}$$

where τ_b is bed shear stress, $w_{f,t}$ is the fall velocity of particles in class t, and $\tau_{c,t}$ is a dimensional critical shear stress (see Section 2.4.1 and Huang and Greimann [2013] for details).

4.3.3 BMA Modeling

For BMA modeling, the parameters of each model were estimated using an optimization algorithm and the dataset from the calibration period (Table 8). The optimization was performed using Multivariate Shuffled Complex Evolution Metropolis – Uncertainty Analysis (MSU)

Table 8 Optimized eight parameters used for each model's simulation of two experiments.

	Seal et al. [1997]				Pender et al. [2001]			
Parameter	Parker	W&C	MPM	Wu	Parker	W&C	MPM	Wu
Roughness coefficient	0.023	0.024	0.025	0.023	0.015	0.016	0.015	0.015
Reference shear stress	0.066	0.052	0.050	0.046	0.055	0.017	0.050	0.038
Hiding and exposure coefficient	0.958	0.868	0	0.797	0.037	0.287	0	0.003
Active layer thickness multiplier	15.270	11.236	4.934	3.145	3.151	0.110	1.494	2.930
Deposition recovery factor	0.873	0.688	0.680	0.902	0.314	0.180	0.131	0.772
Scour recovery factor	0.077	0.463	0.123	0.519	0.241	0.204	0.334	0.443
Bed load adaptation length multiplier	1.155	3.169	9.594	3.353	21.245	3.682	23.283	21.736
Weight of bed load fractions	0.661	0.474	0.996	0.151	0.687	0.116	0.572	0.655

[Sabatine et al. 2015], which is able to compute the likelihood of parameter sets using multiple model output variables. The roughness coefficient n has units of $s \cdot m^{-1/3}$, and all the other parameters are non-dimensional. The reference shear stress θ_r and the hiding and exposure coefficient λ are calibrated in the Parker and Wu models, whereas the datum values of those parameters θ_{r0} and λ_0 are calibrated in the W&C model. Unlike the other models, the MPM model in SRH-1D uses fixed values of 0.0495 and 0 for θ_r and λ, respectively [Wong and Parker 2006], which may limit its ability to reproduce the results of the flume experiments. Readers are referred to Huang and Greimann [2013] for mathematical descriptions of the other parameters used in SRH-1D simulations.

Fig. 20 shows the observations and best outputs from the individual models for the calibration period of each case. To quantitatively compare the performance of the models, Table 9 also shows the Nash-Sutcliffe coefficient of efficiency (NSCE) for each model and variable of

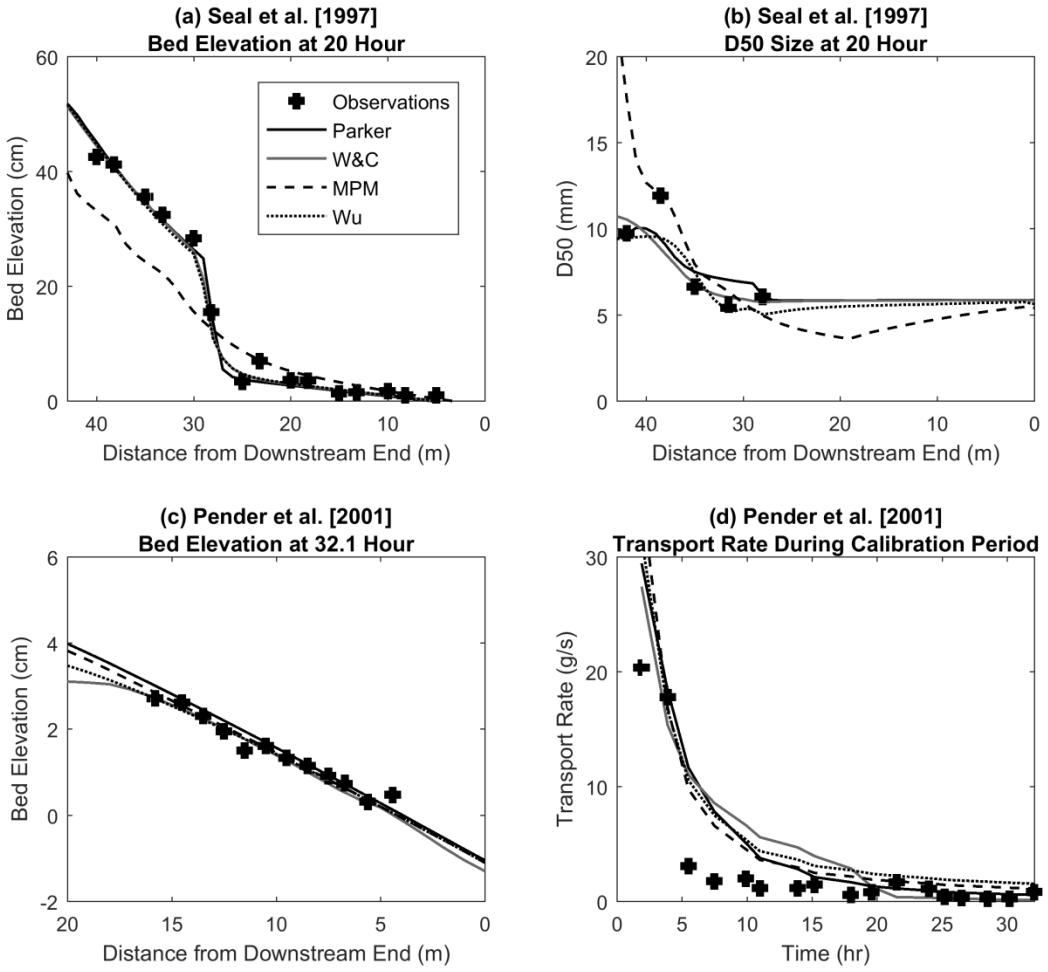

Fig. 20 Observations and individual model outputs for: (a) bed elevation and (b) D50 at 20 hour of the Seal et al. [1997] experiment; (c) bed elevation at 32.1 hour and (d) sediment transport rate during the calibration period of the Pender et al. [2001] experiment.

Table 9 Nash-Sutcliffe coefficient of efficiency (NSCE) values for individual models and BMA models for the calibration (Calib.) and forecast (Fore.) periods of two experiments. Highest NSCE value for each variable is shown in bold face.

	Seal et al. [1997]				Pender et al. [2001]			
	Bed Elevation		D50		Bed Elevation		Transport Rate	
Model	Calib.	Fore.	Calib.	Fore.	Calib.	Fore.	Calib.	Fore.
Parker	**0.99**	0.84	0.36	-0.99	0.92	0.92	**0.63**	-0.53
W&C	**0.99**	0.98	0.42	-0.72	0.88	0.72	0.60	-1.37
MPM	0.84	0.53	0.22	-14.89	**0.93**	0.92	0.45	-9.49
Wu	0.98	0.98	0.47	**-0.23**	0.91	0.84	0.54	-34.45
Bed Elevation BMA	**0.99**	0.98	0.42	-0.70	**0.93**	0.92	0.46	-9.20
D50 BMA	0.97	0.96	**0.53**	-2.54	-	-	-	-
Transport Rate BMA	-	-	-	-	0.92	0.92	0.60	-0.80
Multivariate BMA	**0.99**	**0.99**	0.43	-0.57	0.92	**0.93**	**0.63**	**-0.24**

interest for the calibration period of both flume experiments. NSCE can range from $-\infty$ to 1 and is 1 when the model reproduces the observations perfectly. Note that BMA does not use NSCE in its evaluation of the models.

For the Seal et al. [1997] experiment, the Parker, W&C, and Wu models produce similar bed elevations and D50 sizes, and they reproduce the observations better than the MPM model (Fig. 20). However, the Wu model performs the best for D50 based on the NSCE values (Table 9). For the Pender et al. [2001] experiment, all the models provide similar bed elevations, and all the simulations show notable differences from the observations for sediment transport rate during the

first 15 hours. Based on the NSCE values (Table 9), the MPM model has the best performance for bed elevation while the Parker model has the best performance for transport rate.

Using the outputs from the individual models, univariate BMA models were developed separately for each variable in Fig. 20. In addition, the proposed multivariate BMA model was developed using both variables of interest for both experiments. For the Seal et al. [1997] case, all the BMA models use constant standard deviations for both variables under the assumption that the uncertainties in the model predictions of bed elevation and D50 do not depend on the magnitudes of the variables. For the Pender et al. [2001] case, the univariate BMA model uses constant standard deviations again, which is consistent with the pre-existing BMA method. However, the multivariate BMA model uses constant standard deviations for bed elevation and constant CVs for sediment transport rate. This approach assumes that the uncertainty in sediment transport rate predictions depends on the magnitude of the transport rate. These assumptions will be tested later in this paper.

4.4 Results and Analysis

4.4.1 Model Weights and Standard Deviations

The weights of the four transport models were determined using BMA (based on their ability to reproduce the observations for the calibration period) (Fig. 21). For the Seal et al. [1997] experiment, W&C is weighted heaviest by the bed elevation BMA model, while Wu has the greatest weight for the D50 BMA model. The weights for the multivariate BMA model are similar to those of the bed elevation BMA model, but they emphasize Wu more due to its success in reproducing the D50 profile (recall Table 9). MPM has a significant weight in the D50 BMA model because a couple of data points are captured by this model (see Fig. 20). However, the

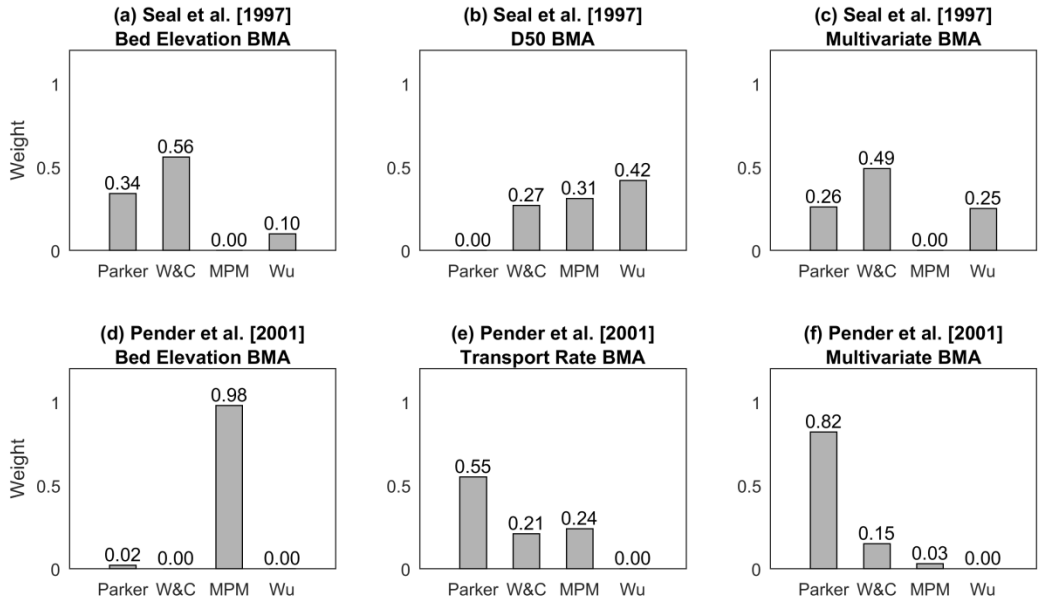

Fig. 21 Weights of individual models determined by each BMA model for two experiment cases.

multivariate BMA assigns zero weight to MPM because it produces poor bed elevation profiles. MPM's weaker performance for this experiment might occur because the range of sediment material sizes is wider than those originally used for MPM model development. For the Pender et al. [2001] experiment, the bed elevation BMA model assigns MPM a weight near one. All four equations produce similar bed elevation results (Fig. 20c), but MPM is consistently closer to the observations and thus preferred (recall Table 9). The sediment transport rate BMA model assigns significant weights to Parker, W&C, and MPM because no single equation consistently outperforms the others. The multivariate BMA model also selects Parker, W&C, and MPM, but it emphasizes Parker more. Although Parker was assigned little weight for the bed elevation BMA model, it produces the best results when both bed elevation and transport rate are considered (Table 9).

The standard deviation for each model is usually larger when the model's weight is smaller, which implies that the uncertainty in the predictions from a model are larger if the model is less likely to be correct. For the Seal et al. [1997] case, the multivariate BMA is able to specify the standard deviations for both bed elevation and D50, and those standard deviations are similar to the values from the univariate BMA models (which only produce standard deviations for the variable they consider). For the Pender et al. [2001] case, the multivariate BMA determines standard deviations for bed elevation, but the standard deviations are not constant for sediment transport rate (the CVs are constant). Like the standard deviations, the CVs tend to be larger when the model weights are smaller. For example, the CV is largest for Wu, which is assigned a weight near zero.

4.4.2 BMA Predictions and Uncertainty

The predictions and CIs from the BMA models for selected times during the forecast periods are shown in Fig. 22 and Fig. 23 for the Seal et al. [1997] and Pender et al. [2001] experiments, respectively (the results are similar for other times during the forecast periods). The predictions are the means of the BMA PDFs as described earlier. The authors also used the modes of the BMA PDFs (i.e. the most likely values) as the prediction but found similar results. For the Seal et al. [1997] case, the bed elevation BMA model produces more accurate forecasts of bed elevation than the D50 BMA model for the time that is shown. The D50 BMA model generates lower bed elevations because the models that it emphasizes underestimate the amount of deposited material. The D50 BMA model also does not forecast D50 as well as the bed elevation BMA model for the time that is shown. For the Pender et al. [2001] case (Fig. 23), the bed elevation BMA model again predicts bed elevation well, but it predicts the transport rate poorly. The transport rate BMA model predicts both bed elevation and transport rate reasonably well.

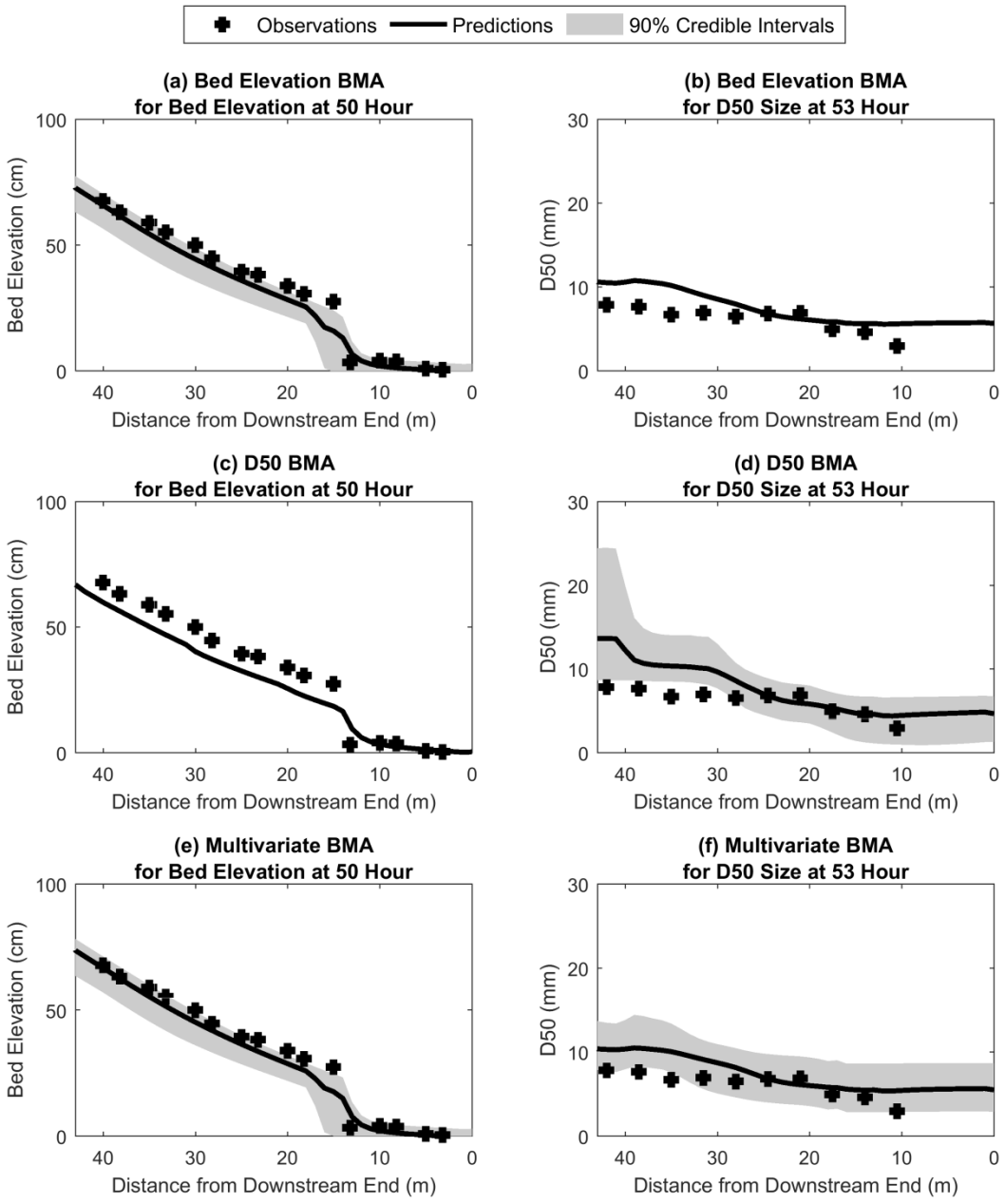

Fig. 22 Observations, predictions, and 90% credible intervals from each BMA model for bed elevation at 50 hour (a, c, e) and D50 size at 53 hour (b, d, f) in the forecast period of the Seal et al. [1997] experiment.

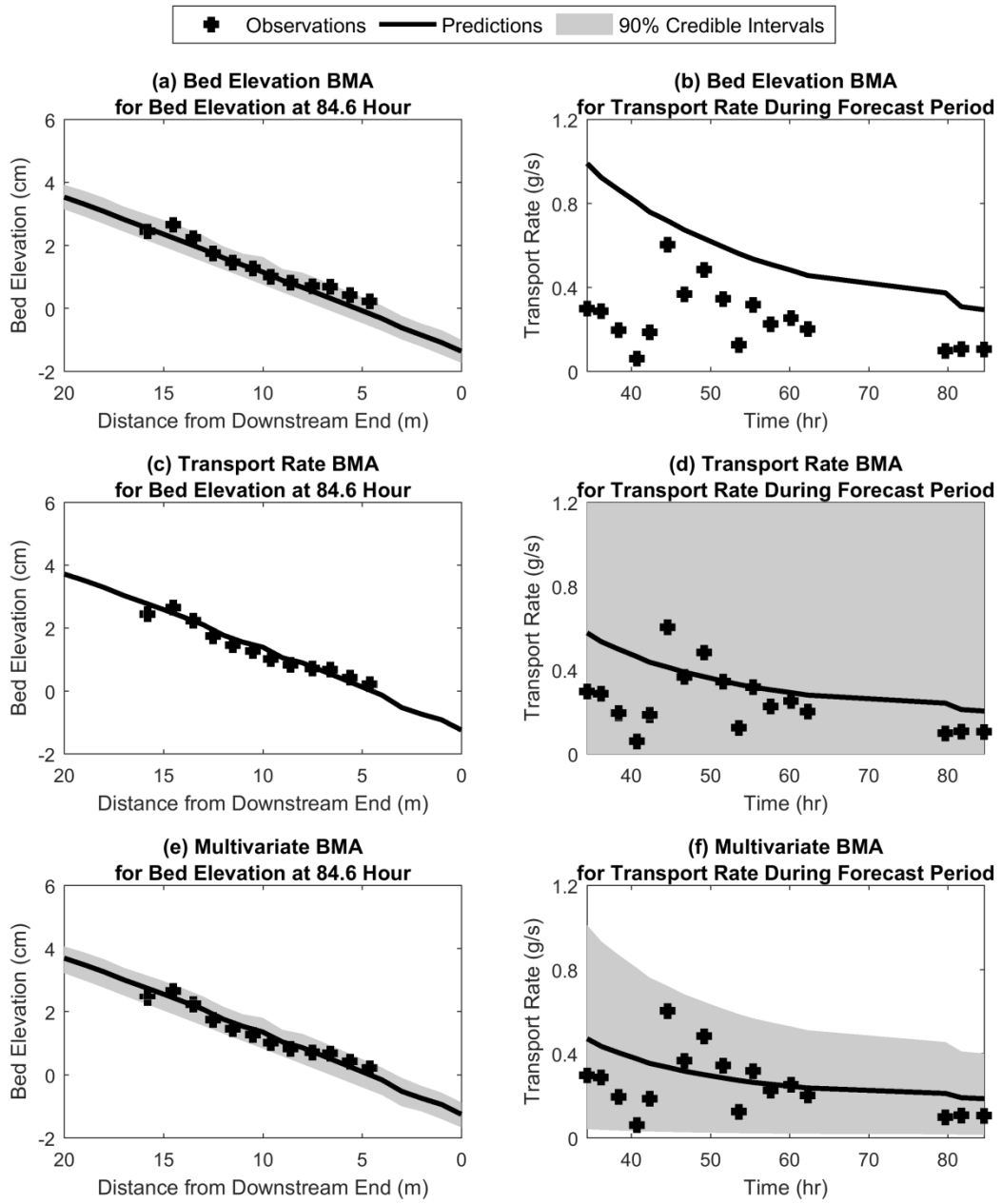

Fig. 23 Observations, predictions, and 90% credible intervals from each BMA model for bed elevation at 84.6 hour (a, c, e) and sediment transport rate during the forecast period (b, d, f) of the Pender et al. [2001] experiment. The credible interval in (d) extends approximately from -5 g/s to 6 g/s, which is much wider than the vertical axis range.

For both flume experiments, the multivariate BMA model performs well for both variables. The multivariate BMA model also produces CIs for both variables, and in most cases, those CIs resemble those from the univariate BMA models. However, the multivariate BMA model produces more realistic CIs for the sediment transport rate in the Pender et al. [2001] case. In particular, the CI from the transport rate BMA model is very wide (beyond the axis limits in the figure) and even includes negative values. In contrast, the CI for the multivariate BMA model is much smaller and remains positive at all times.

The overall performance of the BMA model predictions can be compared using NSCE (Table 9). For the Seal et al. [1997] case, the multivariate BMA model produces the highest NSCE values for bed elevation in both the calibration and forecast periods and intermediate performance for D50 in both periods. The Parker model also has a high NSCE value for bed elevation in the calibration period but does not perform as well during the forecast period. For D50, the D50 BMA model has the highest NSCE in the calibration period but the lowest value in the forecast period. This result occurs because the MPM model captures some D50 values well during the calibration period and thus is assigned a substantial weight (Fig. 21b). However, it produces the worst bed elevation profiles during the calibration period (Fig. 20b). This poor performance indicates a flaw in the model of the sediment mechanics, and it ultimately produces errors in the D50 values for the forecast period. This flaw might be caused by the MPM assumption that fixes the hiding and exposure coefficient λ at 0 [Wong and Parker 2006] (Table 8), which allows too much transport of small particles and overestimates downstream fining. In contrast, the multivariate BMA model ignores MPM by considering both bed elevation and D50, and it produces higher NSCE values for the forecast period compared to either univariate BMA model. For the Pender et al. [2001] case, all three BMA models provide better forecasts for bed

elevation than the individual models. The bed elevation BMA model has a slightly higher NSCE for bed elevation than the multivariate BMA model for the calibration period, but the multivariate BMA model has a higher NSCE for the forecast period. Furthermore, the bed elevation BMA model has a large negative NSCE for transport rate. The multivariate BMA model has the highest NSCE for transport rate in both the calibration and forecast periods.

The percentage of observations covered by the 90% CI for each BMA model is shown in Table 10. If model uncertainty is the only significant source of error and the simplifications used in the BMA method are appropriate, then the 90% CIs should cover approximately 90% of the observations. For the Seal et al. [1997], the CIs for all the BMA models have approximately 90% coverage for the calibration period. For the forecast period, the CIs for all the BMA models cover less than 90% of the data, 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 by deviations from the BMA assumptions. For the Pender et al. [2001] case, the CIs for all the BMA models cover more than 90% of the observations for the forecast period, which suggests that the assumptions used in the BMA models do not hold exactly. The CI for the transport rate BMA model covers 100% of the observations because it is very wide (Fig. 23). The CI for the multivariate BMA model also covers 100% of the observations, but it is much more realistic as shown in Fig. 23. Recall that these two models differ not only in the variables that they consider but also in whether they assume that the standard deviations for transport rate remain constant. Overall, the CIs from the multivariate BMA model cover closer to 90% of the observations than the CIs from the univariate BMA models.

Table 10 Percentage of observations covered by the 90% credible intervals from BMA models of the calibration (Calib.) and forecast (Fore.) periods of two experiments. The percentage that is closest to 90% is in bold face for each variable.

	Seal et al. [1997]				Pender et al. [2001]			
	Bed Elevation		D50		Bed Elevation		Transport Rate	
Model	Calib.	Fore.	Calib.	Fore.	Calib.	Fore.	Calib.	Fore.
Bed Elevation BMA	**89.7**	77.0	-	-	**88.7**	87.2	-	-
D50 BMA	-	-	84.8	66.7	-	-	-	-
Transport Rate BMA	-	-	-	-	-	-	**82.4**	100
Multivariate BMA	88.9	**82.8**	**89.1**	**83.3**	88.1	**92.6**	100	100

4.4.3 Evaluation of Assumptions

A key assumption of the multivariate BMA method is that the standard deviation for some variables cannot be considered constant. To test this assumption, the available data from the calibration period of each experiment was divided into ranges. Each range (i.e. subset of data) was used to develop a different univariate BMA model for that variable, and the BMA standard deviations for each model were then be plotted as a function of the average value of the data in the subset (Fig. 24). For the Seal et al. [1997] case, the standard deviations are approximately constant for all ranges of data for both bed elevation and D50, suggesting that they will also remain constant for the forecast period. For the Pender et al. [2001] case, the standard deviations are again approximately constant for the ranges of bed elevation, but they clearly increase with transport rate. Linear regressions of the plotted data also support these visual interpretations. The coefficients of determination (r^2 values) are all below 0.01 except when the standard deviation is regressed against transport rate (which produces an r^2 value of 0.69). This analysis

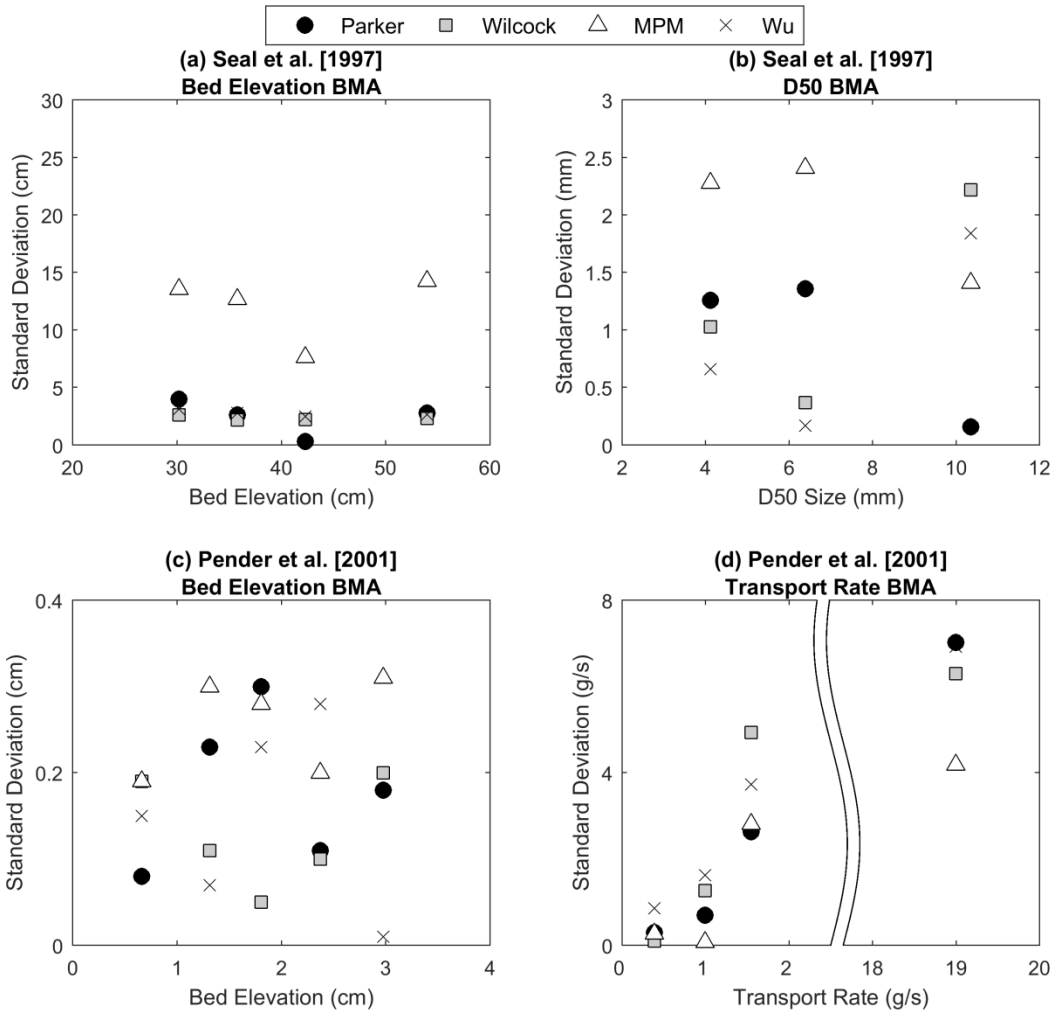

Fig. 24 Standard deviations obtained for each univariate BMA model when subsets of the calibration data are used for BMA model development. In each plot, the x-coordinate is the average value of subset of data used.

supports using constant standard deviations for all variables except for transport rate. Similar analyses can also be used in future applications to determine whether the standard deviations or CVs of different variables should be considered constant.

The impact of using constant CVs for transport rate in the multivariate BMA model is illustrated in Fig. 25. This figure shows predictions and CIs for transport rate during the calibration period when two multivariate BMA models are used. The first assumes constant CVs for transport rate while the second assumes constant standard deviations. When constant standard deviations are used, the 90% CI has a nearly constant width over time, and it covers only 82% of calibration data. The CI also has a much wider range than the observations after 10 hr. Moreover, the lower bound of the CI goes below zero, which is not realistic. On the other hand, when constant CVs are used, the CI narrows with time and still covers all the observations. It also spans only positive values of transport rate. Overall, allowing the standard deviations to vary according to the scale of predicted values improves the uncertainty estimates for the predictions of transport rate.

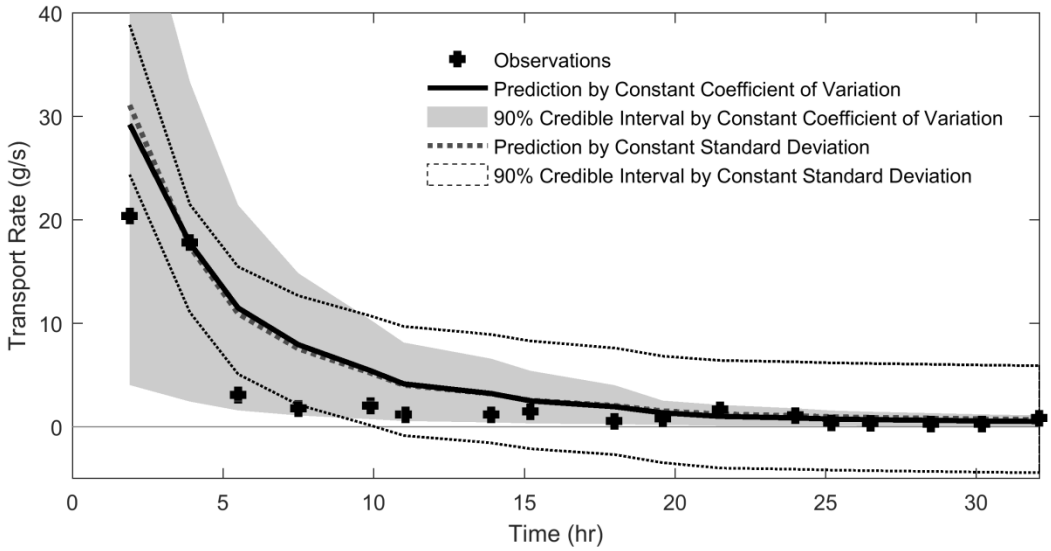

Fig. 25 Observations, predictions, and 90% credible intervals from multivariate BMA models by applying constant coefficients of variation and constant standard deviations for sediment transport rate in the calibration period of Pender et al. [2001] experiment.

Another key assumption is the use of normal distributions to describe the uncertainty associated with each model. Gamma distributions were also implemented. This approach has similar performance in the accuracy of the predictions and the coverage of observations compared to a BMA model using normal distributions. Furthermore, the CIs are always non-negative because the gamma distribution does not allow values to be negative. However, two weaknesses were also found. First, the shape and scale parameters, which define the Gamma distribution, are not able to be calibrated using the EM method, so a more complex optimization algorithm is required. Second, the predictions from the individual models are used for the means of the gamma distributions. However, because the gamma distribution is typically not symmetrical, the individual model predictions do not fall at the modes (most likely values) of the gamma distributions.

4.5 Conclusions

1. Unlike the univariate BMA models, the multivariate BMA model suggested here is able to generate probabilistic predictions for multiple variables during the calibration and forecast periods. The products of multivariate BMA include a single weight for each competing model and either a standard deviation or CV for each model and variable considered. Those quantities are sufficient to develop CIs for all considered variables.

2. For the cases considered, the multivariate BMA usually produces more accurate predictions for the variables than the individual sediment transport models or the univariate BMA models. The improved performance is more notable during the forecast period than the calibration period. In some calibration cases, an individual transport model or a univariate BMA model produces good performance for one variable but poor performance for another

variable. In such a case, that model can ultimately produce poor forecasts for both variables. In contrast, multivariate BMA selects the model weights based on multiple variables, which helps it avoid models that produce good calibration results for the wrong reasons and thus would produce unreliable forecasts.

3. For the cases considered, the CIs from multivariate BMA are also more accurate than the CIs from the univariate BMA models. The most notable improvement was observed in the CI for transport rate for the Pender et al. [2001] case. In multivariate BMA, the CVs for this variable were assumed to be constant instead of the standard deviations. This modification substantially reduces the CI for multivariate BMA and makes it much more consistent with the variability of the observations.

4. Variables whose standard deviations vary can be identified by dividing the calibration data into ranges and developing BMA models for those ranges. The standard deviations from the BMA models can then be plotted against the mean value of the data in each range. Variables whose standard deviations vary with the magnitude of the observation can potentially be modeled using constant coefficients of variation.

Several notable avenues are available for future research. First, the proposed method can be applied to natural rivers where erosion and sedimentation processes are more complex and interact. This application might also allow consideration of other available sediment transport equations such as Ackers and White [1973], Brownlie [1981], modified Laursen's formula [Madden 1993], and Yang [1996]. Second, consideration should be given to allowing the model weights to vary in space and/or time because the most likely transport model can potentially change as the system configuration changes. Third, the magnitude of uncertainty might also

increase with the duration of the forecast period. If the standard deviation for each model's prediction were to vary with the simulation time, it might be possible to reduce the underestimates of predictive uncertainty, which were observed in the forecast period of the Seal et al. [1997] case. Fourth, the relationship between the weights for the transport equations and the hydraulic conditions in the river should be explored further. BMA assumes that the weights identified from the calibration period also apply to the forecast period, but the appropriate weights are likely to change if the hydraulic conditions change. Fifth, the transport equation is not the only source of uncertainty in sediment transport modeling. Uncertainty might originate from the model parameters, the data used for model forcing variables, the channel geometry information, and the observations used for model calibration [Ruark et al. 2011]. Methods should be developed to thoroughly consider the combination of these uncertainties.

CHAPTER 5
CONCLUSIONS

5.1 Summary

This dissertation presented studies on the uncertainty associated with sediment transport modeling. Specifically, this research focused on the effects of poorly specified model parameter values, measurement errors included in model input data, and imperfect transport equations used in the model structure. In the context of Bayesian inference, this study aimed to find formal ways to evaluate and quantify how those uncertain model elements affect the predictions from sediment transport models. Each chapter suggested a new method to overcome the limitations of current methods. Three new methodologies were developed based on the application to a one-dimensional sediment transport model SRH-1D, but they can also be coupled to any types of other numerical models. The following paragraphs summarize the conclusions from the three objectives in this research.

The goal of Chapter 2 was to develop and test an efficient method to evaluate the parameter uncertainty and its impacts on prediction uncertainty. The Evolving Latin Hypercube (ELH) method constructs the posterior sample by replicating parameter sets rather than generating similar parameter sets to improve the efficiency of uncertainty estimation. This new method then updates the posterior sample by adding parameter sets generated from the parameter space, which can be specified by considering previous simulation results. The performance of ELH was evaluated by comparing to the existing methods Generalized Likelihood Uncertainty Estimation (GLUE) [Beven and Binley 1992] and Shuffled Complex Evolution Metropolis - Uncertainty Analysis (SCEM-UA) [Vrugt et al. 2003], through application to synthetic posterior parameter probability density functions (PDF) and a sediment transport model. First, ELH was

able to estimate the joint posterior PDF of the parameters and the prediction uncertainty using fewer model simulations than GLUE or SCEM-UA. The reduction in required simulations was greatest for cases where the pre-specified parameter ranges are much wider than the main body of the joint posterior PDF. Second, ELH could estimate the parameter variance better than SCEM-UA when the posterior parameter distribution has a large variance. For example, SCEM-UA underestimated the large posterior variances for a multivariate Gaussian target distribution, while ELH accurately estimated such variances. In addition, SCEM-UA also provided smaller IQR ratios than both ELH and GLUE for the sediment transport model parameters that were poorly constrained by the data. Third, SCEM-UA and ELH could produce similar estimates for the prediction uncertainty that originates from uncertain parameters, and these results were more accurate than GLUE. For the sediment transport model case, the mean predictions from all three methods indicated almost the same Nash–Sutcliffe coefficient of efficiency (NSCE) values. However, the continuous ranked probability score (CRPS) values suggested that the probabilistic forecasts from GLUE performed worse than the other two methods and provided unrealistic credible intervals (CI), likely because it neglects parameter correlation. Fourth, the replication process of ELH caused more erratic histograms of the posterior parameter sample compared to SCEM-UA. In particular, ELH showed higher errors in the estimated PDFs and CDFs for the synthetic cases with the multivariate Gaussian PDF and the banana-shaped PDF. Nonetheless, both SCEM-UA and ELH represented the overall aspects of the target posterior PDFs well for all synthetic cases considered.

Chapter 3 described the ways to include the uncertainties in various input data when estimating the uncertainty in the predictions from a sediment transport model. For each input variable, the uncertainty was defined using a Gaussian distribution, and the mean and standard deviation of

that distribution were estimated within the parameter uncertainty estimation framework. The study then evaluated the contribution of input uncertainty to the overall prediction uncertainty. From the case studies that simulate a natural river with the diverse assumptions of input uncertainty, the following conclusions were drawn. First, including input uncertainty required many more model simulations than considering parameter uncertainty alone. The large increase in simulations was partly due to difficulty inferring the posterior distributions for inputs that have little impact on the model results. However, the required number of simulations can be reduced by only considering the most important sources of input uncertainty. For example, the required number of simulations increased only by a factor of 2 when only the single most important uncertainty source (benchmark elevations) was included in the analysis. Furthermore, the estimated prediction uncertainty for this simplified case was similar to the case that considered all the uncertain inputs. Second, including input uncertainty increases the flexibility of the simulations to match observations, which might cause over-calibration, but the results indicate that it did not reduce the reliability of the mean model predictions for the forecast period. For the cases considered, the performance of the forecasts increases slightly when input uncertainty is included. Third, considering input uncertainty can modify the estimated uncertainty of the parameter values. For example, the median estimate of Manning's roughness was around 0.039 for all cases considered, but the ratio of the posterior and prior interquartile ranges (IQRs) for that parameter was 33% when input uncertainty was neglected and 43% when all types of input uncertainty were included. This increase occurred because variations in the inputs allowed wider ranges of parameter values to produce good model performance. Fourth, including input uncertainty can produce much larger estimates of the prediction uncertainty, which are also more accurate. When all the uncertain inputs were included, the prediction uncertainty increased by

72% and 199% for the calibration and forecast periods, respectively, compared to the case without input uncertainty. In addition, the range of predictions covered only 46% of the deposition volume observations in the forecast period when only parameter uncertainty was considered, but the range covered 94% of the observations when all the uncertain inputs were included in the predictions.

Chapter 4 presented a multivariate BMA method that can consider multiple model output variables when assessing the uncertainty related to the selection and application of a transport equation. The method combines the predictions from a set of competing equations by a weighted-average where the weights reflect the uncertainty due to the equation selection. This new method also allows the uncertainty associated with each equation to vary with the magnitude of the variables if needed. Specifically, the equation uncertainties can be modeled using constant coefficients of variation (CV) or constant standard deviations (SD). By this approach, the multivariate BMA provides not only a weight for each competing equation but also either a SD or CV for each equation and variable considered. Those quantities are sufficient to generate probabilistic predictions for all considered variables. Thus, the multivariate BMA model can produce probabilistic forecasts for multiple variables, whereas the existing univariate BMA models are limited to single variable forecasts. From the case studies conducted in this research, the following conclusions were made. First, the multivariate BMA usually provided more accurate predictions for the considered variables than the individual sediment transport models or the univariate BMA models. The improved performance was more evident during the forecast period than the calibration period. For example, an individual transport model or a univariate BMA model showed good calibration for one variable but poor calibration for another variable. In such a case, that model can ultimately lead to poor forecasts for both variables. On

the other hand, the multivariate BMA determines the model weights based on multiple variables, which helps it avoid models that produce good calibration results for the wrong reasons and thus would produce unreliable forecasts. Second, the CIs of the predictions from the multivariate BMA model are also more accurate than the CIs from the univariate BMA models. Specifically, the multivariate BMA model applied the constant CVs to sediment transport rate predictions, which produced a reduced CI that was much more consistent with the variability of the observations.

The main significance of this research is that the methodologies developed here can reduce the gap between the sediment transport modeling discipline and the Bayesian inference paradigm. As discussed in Chapter 1 and the introduction sections of Chapters 2 to 4, existing Bayesian methods have significant limitations with respect to hydraulic and sediment transport models because: (1) they often require too many model simulations and (2) past research has not sufficiently addressed ways to treat diverse uncertain input (and/or output) variables. In terms of the computational time, the ELH method from Chapter 2 showed significant reductions in the number of required simulations when estimating the parameter uncertainty compared to exiting methods. Chapter 3 showed that only considering the most important sources of input uncertainty can retain the efficiency while providing the prediction uncertainty estimates similar to the case that considered all the uncertain inputs. In addition, the BMA approach developed in Chapter 4 only requires a single calibrated simulation for each of the competing equations (or models), which requires relatively little computation time. Besides the efficiency improvements, this study also demonstrated the feasibility of an input error model that was developed for hydrologic modeling to address various input uncertainties in sediment transport modeling (Chapter 3). The multivariate BMA method proposed in Chapter 4 enables implementation of

model structure uncertainty analysis to consider multiple model output variables. In addition, this method provided a way to identify and consider the heteroscedasticity of variable's uncertainty when generating the probabilistic predictions. All these efforts can ultimately enhance the applicability of Bayesian inference to the field of hydraulic and sediment transport modeling, which will bring innovative improvements to interpret river system behavior. The probabilistic predictions for the changes in channel bed elevations, sediment size distributions, and volume of eroded or deposited sediments will be useful for identifying potential problems such as river bank erosion, structure failure by bed degradation, or decrease in dam capacity due to reservoir sedimentation. Formal and reliable estimates from Bayesian uncertainty analysis for those problems will be helpful for seeking sustainable solutions to protect water resources and infrastructure systems. Furthermore, it is expected to lessen conservatism in decision-making and reduce the conflicts between discipline-specific objectives.

5.2 Future Avenues

While this research suggested novel methods for using Bayesian inference in sediment transport modeling studies, the line of research should be expanded along several avenues in the future:

1. Errors in the calibration dataset should also be considered as important uncertainty sources. Model calibration is implemented to reduce the discrepancy between observations and model responses by assuming that the observed values are accurate. Unfortunately, like input uncertainty, the calibration data also inherently possess uncertainty due to their measurement errors, which can necessarily affect the uncertainty estimation. For the considered cases, the CIs from both the existing and new methods often covered fewer observations than expected, which indicates an underestimation of uncertainty. Throughout the dissertation, it was

suggested that this underestimation likely occurred because uncertainty in the observations used for calibration was neglected. Influence of inaccurate or uncertain calibration data on the calibration results has been noted in various model applications in hydrology and hydraulics [Khu et al. 2008; McMillan et al. 2012; Williams et al. 2013; Le Coz et al. 2014]. It is difficult to evaluate that uncertainty using a Bayesian approach because there is no way to know what the true value is. Thus, it is impossible to reliably compute likelihoods [Khadam and Kaluarachchi 2004; Ajami et al. 2007]. For this reason, assumptions that make the calibration process more complex need to be made to properly include the calibration data uncertainty. In addition, the estimation of overall uncertainty will require more simulations as the degrees of freedom increase.

2. Flow rate data used for forecast simulations requires additional consideration of the underlying uncertainty. The current research examined the uncertainties in various input data by assuming that the input uncertainty arises from the measurement errors. Specifically, the input uncertainty identified from the calibration period was applied to the input data of the forecast period simulations. This approach is effective for most input data, which represent either the initial conditions of the forecast simulation or the boundary conditions that do not vary along the simulation, because those inputs can be determined using the information available at the moment of simulation setup. However, a time series of flow rate is supposed to represent a future scenario of the flow in a natural river, which is not observed yet. Hydrologists who investigate climate change, rainfall, and basin discharge using various hydrologic models can estimate and provide the streamflow information in the forecast period of interest. That information also inherently possesses uncertainty, so its influence should be included when assessing the prediction uncertainty. The uncertainty estimations

for such flow predictions have been implemented using Bayesian methods in numerous hydrologic modeling studies [Kavetski et al. 2003; Reggiani et al. 2009; Bouda et al. 2011; Zhao et al. 2015]. Thus, a similar approach could be implemented for sediment transport modeling, and the new method could be linked to the framework suggested in this research.

3. All the methodologies suggested in this research should be integrated to form a single process of uncertainty estimation. For example, ELH currently considers only parameter uncertainty and could be extended to include input uncertainty. The BMA method could then be modified to consider a sample of predictions from that extended ELH method, instead of using only a single best prediction for each transport equation. The integrated framework could produce more precise probabilistic predictions by thoroughly aggregating all uncertainties as well as distinguishing the influence of each uncertain element on the overall prediction uncertainty. In addition, this framework should retain flexibility, so it can replace any of the current individual approaches with different methods if needed.

4. Besides the development of new methods, the impact of quality and/or quantity of calibration data on the performance of the uncertainty estimates is also worth investigating. Past research using hydrologic models has found that it is not the size of the calibration dataset that matters but the variability of the observed values in the dataset [Kuczera 1982; Sorooshian et al. 1983; Gupta and Sorooshian 1985; Yapo et al. 1996]. Specifically, Sadegh and Vrugt [2013] suggested that both wet and dry periods are required to make sure that all the different components of the watershed model work so that many parameters as possible can be estimated from the calibration data. Using a Bayesian uncertainty analysis, Vrugt et al. [2002] also demonstrated that only a few streamflow data measurements are necessary to reliably calibrate a conceptual hydrologic model and the remaining data contain redundant

information. For the Tachia River case used in this research, the size of the calibration dataset was not so large, but it contained measurements from both scour and sedimentation at different locations. By virtue of that variability, the posterior distributions of the parameters that have large influences on the model results were well identified by the calibration data, and the mean predictions also showed good performance for both the calibration and forecast periods. In Chapter 4, the multivariate BMA method considers the correlations between the observed values (e.g., adjacent bed elevations) in the calibration data, and the results indicated that removing redundancy originating from the correlations did not make the uncertainty estimates worse. The effectiveness of calibration data can be further examined using additional cases where datasets provide different variability and/or redundancy so that general guidelines for sediment transport model applications can be established. Such efforts would enable a modeler to acquire the calibration data more efficiently while maintaining the reliability of the uncertainty estimates.

5. The MATLAB codes that were written for implementing all the methodologies developed in this research will be released with a user's manual to the public. Those materials will be produced in order to help practitioners who are unfamiliar with either specialized statistical theories or MATLAB programing techniques to conduct uncertainty analysis for their model applications.

REFERENCES

Abbaspour, K. C., J. Yang, I. Maximov, R. Siber, K. Bogner, J. Mieleitner, J. Zobrist, and R. Srinivasan (2007), Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT, *Journal of hydrology*, 333, 413-430, doi:10.1016/j.jhydrol.2006.09.014.

Ackers, P., and W. R. White (1973), Sediment transport: new approach and analysis, *Journal of the Hydraulics Division*, 99(11), 2041-2060.

Ahn, J., 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.

Ahn, J., C. T. Yang, D. B. Pridal, and J. I. Remus (2013), Numerical modeling of sediment flushing from Lewis and Clark Lake, *International Journal of Sediment Research*, 28(2), 182-193.

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.

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.

Anderson, N. H. (1965), Averaging versus adding as a stimulus-combination rule in impression formation, *Journal of Experimental Psychology*, 70(4), 394.

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, doi:10.1007/s00477-014-0958-4.

Bagnold, R. A., and O. Barndorff-Nielsen (1980), The pattern of natural size distributions, *Sedimentology*, 27(2), 199-207.

Bates, J. M., and C. W. J. Granger (1969), The combination of forecasts, *Operational Research*, 20(4), 451-468.

Bertin, X., A. B. Fortunato, and A. Oliveira (2007), Sensitivity analysis of a morphodynamic modeling system applied to a Portuguese tidal inlet, In: *Proceedings of 5th IAHR Symposium on River, Coastal, and Estuarine Morphodynamics*, 11-77.

Beven, K. J. (2011), *Rainfall-Runoff Modelling: The Primer*, John Wiley & Sons, West Sussex, UK.

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, doi:10.1016/S0022-1694(01)00421-8.

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, doi:10.1016/j.jhydrol.2008.02.007.

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, doi:10.1016/j.jhydrol.2007.12.026.

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, doi:10.1002/hyp.331.

Bouda, M., A. N. Rousseau, B. Konan, P. Gagnon, and S. J. Gumiere (2011), Bayesian uncertainty analysis of the distributed hydrological model HYDROTEL, *Journal of Hydrologic Engineering*, 17(9), 1021-1032.

Box, G. E. P., and G. C. Tiao (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Boston, MA.

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, doi:10.1002/1096-9837(200008)25:8<825::AID-ESP101>3.0.CO;2-3.

Brown, T. A. (1974), *Admissible Scoring Systems for Continuous Distributions*, Technical Report P-5235, Rand Corporation, Santa Monica, CA.

Brownlie, W. R. (1981), *Prediction of Flow Depth and Sediment Discharge in Open Channels*, Ph.D. dissertation, California Institute of Technology, Pasadena, CA.

Brunner G. W. (2016), *HEC-RAS River Analysis System User's Manual*, Hydrologic Engineering Center, U.S. Army Corps of Engineers, Davis, CA.

Buckland, S. T., K. P. Burnham, and N. H. Augustin (1997), Model selection: an integral part of inference, *Biometrics*, 53(2), 603-618.

Bunte, K., and S. R. Abt (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.

Burnham, K. P., and D. R. Anderson (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd Edition, Springer Science and Business Media.

Camenen B., and P. Larroudé (2003), Comparison of sediment transport formulae for the coastal environment, *Coastal Engineering*, 48(2), 111-132.

Capaldi, A., S. Behrend, B. Berman, J. Smith, J. Wright, and A. L. Lloyd (2012), Parameter estimation and uncertainty quantification for an epidemic model, *Mathematical Biosciences and Engineering*, 9(3), 553-576.

Carpenter, T. M., and K. P. Georgakakos (2006), Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales, *Journal of Hydrology*, 329(1), 174-185.

Chahinian, N., and R. Moussa (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.

Chang, C. H., J. C. Yang, and Y. K. Tung (1993), Sensitivity and uncertainty analysis of a sediment transport model: A global approach, *Stochastic Hydrology and Hydraulics*, 7(4), 299–314.

Chapman, W. L., W. J. Welch, K. P. Bowman, J. Sacks, and J. E. Walsh (1994), Arctic sea ice variability: Model sensitivities and a multidecadal simulation, *Journal of Geophysical Research*, 99, 919–935.

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 & Software*, 80, 159-174, doi:10.1016/j.envsoft.2016.02.026.

Chow, V. T. (1959), *Open Channel Hydraulics*, McGraw-Hill, New York, NY.

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, FL.

Clément, P., and H. Piégay (2003), Statistics and fluvial geomorphology. In: *Tools in Fluvial Geomorphology*, edited by G. M. Kondolf, and H. Piégay, pp. 597-630, John Wiley & Sons, Chichester UK.

Clyde, M. A. (1999), Bayesian model averaging and model search strategies, *Bayesian Statistics 6*, 157-185.

Clyde, M., and E. I. George (2004), Model uncertainty, *Statistical Science*, 19(1), 81-94.

Corazza, M. Z., T. Abrao, F. G. Lepri, S. M. Gimenez, E. Oliveira, and M. J. Santos (2012), Monte Carlo method applied to modeling copper transport in river sediments, *Stochastic Environmental Research and Risk Assessment*, 26(8), 1063-1079.

Cressie, N., C. A. Calder, J. S. Clark, J. M. Ver Hoef, and C. K. Wikle (2009), Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling, *Ecological Applications*, 19(3), 553-570.

Cui, Y., and G. Parker (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, A. G., L. C. van Rijn, J. S. Damgaard, J. van de Graaff, and J. S. Ribberink (2002), Intercomparison of research and practical sand transport models, *Coastal Engineering*, 46(1), 1-23.

De Risi, R., K. Goda, N. Mori, and T. Yasuda (2017), Bayesian tsunami fragility modeling considering input data uncertainty, *Stochastic Environmental Research and Risk Assessment*, 31(5), 1253-1269.

Dehghani, M., B. Saghafian, F. Nasiri-Saleh, A. Farokhnia, and R. Noori (2014), Uncertainty analysis of streamflow drought forecast using artificial neural networks and Monte-Carlo simulation, *International Journal of Climatology*, 34(4), 1169-1180, doi:10.1002/joc.3754.

Deltares (2014), *Delft3D: Simulation of Multi-dimensional Hydrodynamic Flows and Transport Phenomena, Including Sediments - User Manual*, Deltares, Delft, Netherlands.

Dempster, A. P., N. M. Laird, and D. B. Rubin (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 S. Sanner (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.

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.

Einstein, H. A. (1950), *The Bed-load Function for Sediment Transportation in Open Channel Flows*(Technical Bulletin No. 1026), U.S. Department of Agriculture, Soil Conservation Service, Washington, DC.

Engelund, F., and E. Hansen (1972), *A Monograph on Sediment Transport in Alluvial Streams*, Teknish Forlag, Technical Press, Norway.

Ficklin, D. L., Y. Luo, and M. Zhang (2013), Climate change sensitivity assessment of streamflow and agricultural pollutant transport in California's Central Valley using Latin hypercube sampling, *Hydrological Processes*, 27(18), 2666-2675, doi:10.1002/hyp.9386.

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, doi:10.1016/j.jhydrol.2003.12.037.

Frigge, M., D. C. Hoaglin, and B. Iglewicz (1989), Some implementations of the boxplot, *The American Statistician*, 43(1), 50-54.

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.

Gaeuman, D., E. D. Andrews, A. Krause, and W. Smith (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., and D. B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Statistical Science*, 7(4), 457-472.

Geman, S., and D. Geman (1984), Stochastic relaxation, Gibbs distribution and Bayesian restoration of images, *IEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Geweke, J. (1991), Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, *Research Department Staff Report*, 148, Federal Reserve Bank of Minneapolis, Minneapolis, MN.

Givens, G. H., and J. A. Hoeting (2012), *Computational Statistic*, 2nd Edition. John Wiley & Sons, Hoboken, NJ.

Gneiting, T., A. E. Raftery, F. Balabdaoui, and A. Westveld (2004), Verifying probabilistic forecasts: Calibration and sharpness, *17th Conference on Probability and Statistics in the Atmospheric Sciences*, American Meteorological Society, Seattle, WA.

Google Maps (2014), Tachia River, Taiwan, retrieved from https://www.google.com/maps/search/tachia+river+taiwan/@24.2916791,120.6174143,17625m/data=!3m1!1e3

Granger, C. W., and R. Ramanathan (1984), Improved methods of combining forecasts, *Journal of Forecasting*, 3(2), 197-204.

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, FL.

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.

Gupta, V. K., and S. Sorooshian (1985), The relationship between data and the precision of parameter estimates of hydrological models, *Journal of Hydrology*, 81, 57–77.

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, doi:10.1007/s11222-006-9438-0.

Hansen, B. E. (2007), Least squares model averaging, *Econometrica*, 75(4), 1175-1189.

Hassan, A. E., H. M. Bekhit, and J. B. Chapman (2008), Uncertainty assessment of a stochastic groundwater flow model using GLUE analysis, *Journal of Hydrology*, 362(1-2), 89-109, doi:10.1016/j.jhydrol.2008.08.017.

Hastings, W. K. (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 57(1), 97-109.

Hersbach, H. (2000), Decomposition of the continuous ranked probability score for ensemble prediction systems, *Weather and Forecasting*, 15(5), 559-570.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999), Bayesian model averaging: a tutorial, *Statistical Science*, 14(4), 382-401.

Huang, J. V., and B. P. Greimann (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 B. P. Greimann (2013), *User's Manual for SRH-1D 3.0*, Bureau of Reclamation, U.S. Department of the Interior, Denver, CO.

Huard, D., and A. Mailhot (2006), A Bayesian perspective on input uncertainty in model calibration: Application to hydrological model 'abc', *Water Resources Research*, 42, W07416, doi:10.1029/2005WR004661.

Humphrey, G. B., M. S. Gibbs, G. C. Dandy, and H. R. Maier (2016), A hybrid approach to monthly streamflow forecasting: integrating hydrological model outputs into a Bayesian artificial neural network, *Journal of Hydrology*, 540, 623-640, doi:10.1016/j.jhydrol.2016.06.026.

Jia, Y., and S. S. Y. Wang (2001), *CCHE2D: Two-dimensional Hydrodynamic and Sediment Transport Model for Unsteady Open Channel Flows over Loose Bed(NCCHE-TR-2001-2)*, National Center for Computational Hydroscience and Engineering, University of Mississippi.

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, doi:10.1002/qre.1044.

Jones, D. R., M. Schonlau, and W. J. Welch (1998), Efficient global optimization of expensive black-box functions, *Journal of Global Optimization*, 13, 455-492, doi:10.1023/A:1008306431147.

Jung, J. Y., J. D. Niemann, and B. P. Greimann (2017), Modeling input errors to improve uncertainty estimates for one-dimensional sediment transport models, *Stochastic Environmental Research and Risk Assessment*, 1-16, doi:10.1007/s00477-017-1495-8.

Jung, J. Y., J. D. Niemann, and B. P. Greimann (2018), Combining predictions and assessing uncertainty from sediment transport equations using multivariate Bayesian model averaging,

Journal of Hydraulic Engineering, 144(4), 04018008, doi:10.1061/(ASCE)HY.1943-7900.0001436.

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.

Kavetski, D., S. W. Franks, and G. Kuczera (2003), Confronting input uncertainty in environmental modeling, *In: Calibration of Watershed Models, Water Science and Applications Series Volume 6*, edited by Q. Duan et al., pp. 49-68, AGU, Washington, DC.

Khadam, I. M., and J. J. Kaluarachchi (2004), Use of soft information to describe the relative uncertainty of calibration data in hydrologic models, *Water Resources Research*, 40, W11505, doi:10.1029/2003WR002939.

Khu, S. T., H. Madsen, and F. Di Pierro (2008), Incorporating multiple observations for distributed hydrologic model calibration: An approach using a multi-objective evolutionary algorithm and clustering, *Advances in Water Resources*, 31(10), 1387-1398.

Kleiber, W., A. E. Raftery, and T. Gneiting (2011), Geostatistical model averaging for locally calibrated probabilistic quantitative precipitation forecasting, *Journal of the American Statistical Association*, 106(496), 1291-1303.

Kuczera, G. (1982), On the relationship between the reliability of parameter estimates and hydrologic time series used in calibration, *Water Resources Research*, 18, 146–154.

Kuczera, G., D. Kavetski, S. Franks, and M. Thyer (2006), Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, *Journal of Hydrology*, 331(1-2), 161-177, doi:10.1016/j.jhydrol.2006.05.010.

Lai, Y. (2016), *SRH-2D: Theory and User's Manual, Bureau of Reclamation*, U.S. Department of the Interior, Denver, CO.

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, CO.

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.

Laursen, E. M. (1958), The total sediment load of streams, *Journal of the Hydraulics Division*, 84(1), 1531-1536.

Le Coz, J., B. Renard, L. Bonnifait, F. Branger, and R. Le Boursicaud (2014), Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: A Bayesian approach, *Journal of Hydrology*, 509, 573-587.

Li, L., and C. Y. Xu (2014), The comparison of sensitivity analysis of hydrological uncertainty estimates by GLUE and Bayesian method under the impact of precipitation errors, *Stochastic Environmental Research and Risk Assessment*, 28(3), 491-504.

Li, P., and Q. D. Vu (2013), Identification of parameter correlations for parameter estimation in dynamic biological models, *BMC Systems Biology*, 7(1), 91, doi:10.1186/1752-0509-7-91.

Limeneros, J. T. (1970), *Determination of the Manning Coefficient From Measured Bed Roughness in Natural Channels*, Geological Survey, U.S. Department of the Interior, Washington DC.

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, doi:10.1198/TECH.2009.08040.

López, R., and J. Barragán (2008), Equivalent roughness of gravel-bed rivers, *Journal of Hydraulic Engineering*, 134(6), 847-851.

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.

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.

McLean, D. G. (1985), Sensitivity analysis of bedload equations, *Proceedings of Canadian Society for Civil Engineering Annual Conference 1985*, 1-15.

McMillan, H., T. Krueger, and J. Freer (2012), Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality, *Hydrological Processes*, 26(26), 4078-4111.

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.

Meyer-Peter, E., and R. Müller (1948), Formula for bed-load transport, *Proceedings of 2nd International Association for Hydraulic Research*, 39-64.

Miller, P. S., and R. C. Lacy (2003), Metamodels as a tool for risk assessment, *In: Experiments in Consilience: Integrating Social and Scientific Responses to Save Endangered Species*, edited by F. R. Westley, and P. S. Miller, pp. 333-351, Island Press, Washington, DC.

Mo, X., and K. Beven (2004), Multi-objective parameter conditioning of a three-source wheat canopy model, *Agricultural and Forest Meteorology*, 122(1), 39-63.

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.

Nelson, J. M. (2016), *FaSTMECH Solver Manual*, International River Interface Cooperative (iRIC).

Nelson, W. B. (1982), *Applied Life Data Analysis*, John Wiley & Sons, New York, NY.

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, doi:10.1016/j.jhydrol.2016.06.022.

Nourani, V., S. Mousavi, D. Dabrowska, and F. Sadikoglu (2017), Conjunction of radial basis function interpolator and artificial intelligence models for time-space modeling of contaminant transport in porous media, *Journal of Hydrology*, 548, 569-587, doi:10.1016/j.jhydrol.2017.03.036.

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, NY.

Owen-Joyce, S. J., and L. H. Raymond (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, CO.

Parker, G. (1990), Surface-based bedload transport relation for gravel rivers, *Journal of Hydraulic Research*, 28(4), 417-436.

Parrish, M. A., H. Moradkhani, and C. M. DeChant (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., T. B. Hoey, C. Fuller, and I. K. Mcewan (2001), Selective bedload transport during the degradation of a well sorted graded sediment bed, *Journal of Hydraulic Research*, 39(3), 269-277.

Pinto, L., A. B. Fortunato, and P. Freire (2006), Sensitivity analysis of non-cohesive sediment transport formulae, *Continental Shelf Research*, 26(15), 1826-1839.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski (2005), Using Bayesian model averaging to calibrate forecast ensembles, *Monthly Weather Review*, 133(5), 1155-1174.

Reggiani, P., M. Renner, A. H. Weerts, and P. A. H. J. M. Van Gelder (2009), Uncertainty assessment via Bayesian revision of ensemble streamflow predictions in the operational river Rhine forecasting system, *Water Resources Research*, 45(2), doi: 10.1029/2007WR006758.

Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks (2010), Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resources Research*, 46(5), doi: 10.1029/2009WR008328.

Ribberink, J. S. (1998), Bed-load transport for steady flows and unsteady oscillatory flows, *Coastal Engineering*, 34(1), 59-82.

Robert, C. P., G. Casella, and G. Casella (2010), *Introducing Monte Carlo Methods with r*, Springer, New York, NY.

Rottner, J. (1959), A formula for bed load transportation, *La Houille Blanche*, 14(3), 285-307.

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), 1-7.

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, and 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, 4831-4850, doi:10.5194/hess-17-4831-2013.

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.

Saltelli, A., T. H. Andres, T. Homma (1995), Sensitivity analysis of model output - Performance of the iterated fractional factorial design method, *Computational Statistics & Data Analysis*, 20, 387-407.

Sauer, V. B., and R. W. Meyer (1992), *Determination of Error in Individual Discharge Measurements*, Geological Survey, U.S. Department of the Interior, Denver, CO.

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.

Seal, R., C. Paola, G. Parker, J. B. Southard, and P. R. Wilcock (1997), Experiments on downstream fining of gravel: I. Narrow-channel runs, *Journal of Hydraulic Engineering*, 123(10), 874-884.

Shen, Z. Y., L. Chen, and T. Chen (2012), Analysis of parameter uncertainty in hydrological and sediment modeling using GLUE method: a case study of SWAT model applied to Three Gorges Reservoir Region, China, *Hydrology and Earth System Sciences*, 16(1), 121, doi:10.5194/hess-16-121-2012.

Shimizu, Y., and H. Takebayashi (2014), *Nays2DH Solver Manual*, International River Interface Cooperative (iRIC).

Simões, F. J. (2009), SToRM: A numerical model for environmental surface flows, *33rd IAHR Congress: Water Engineering for a Sustainable Environment*, Vancouver, Canada.

Sloughter, J. M., T. Gneiting, and A. E. Raftery (2010), Probabilistic wind speed forecasting using ensembles and Bayesian model averaging, *Journal of the American Statistical Association*, 105(489), 25-35.

Smith, R., C. Tebaldi, D. Nychka, and L. Mearns (2009), Bayesian modeling of uncertainty in ensembles of climate models, *Journal of American Statistical Association*, 104 (485), 97-116, doi:10.1198/jasa.2009.0007.

Sorooshian, S., V. K. Gupta, and J. L. Fulton (1983), Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall-runoff models: influence of calibration data variability and length on model credibility, *Water Resources Research*, 19, 251–259.

Steffler, P., and J. Blackburn (2002), *River2D: Two-dimensional Depth Averaged Model of River Hydrodynamics and Fish Habitat*, University of Alberta.

Stein, M. (1987), Large sample properties of simulations using Latin hypercube sampling, *Technometrics*, 29(2), 143-151.

Stone, N. (2011), *Gaussian Process Emulators for Uncertainty Analysis in Groundwater Flow*, Doctoral dissertation, University of Nottingham, Nottingham, UK.

Sun, S., and J. L. Bertrand-Krajewski (2013), Separately accounting for uncertainties in rainfall and runoff: Calibration of event-based conceptual hydrological models in small urban

catchments using Bayesian method, *Water Resources Research*, 49, 5381–5394, doi:10.1002/wrcr.20444.

Teng, J., A. J. Jakeman, J. Vaze, B. F. Croke, D. Dutta, and S. Kim (2017), Flood inundation modelling: A review of methods, recent advances and uncertainty analysis, *Environmental Modelling & Software*, 90, 201-216, doi:10.1016/j.envsoft.2017.01.006.

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, doi:10.1007/s11222-006-8769-1.

Thiemann, M., M. Trosset, H. Gupta, and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrologic models, *Water Resources Research*, 37(10), 2521- 2535, doi:10.1029/2000WR900405.

Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, *Water Resources Research*, 45, W00B14, doi:10.1029/2008WR006825.

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.

Tramblay, Y., V. Thiemig, A. Dezetter, and L. Hanich (2016), Evaluation of satellite-based rainfall products for hydrological modelling in Morocco, *Hydrological Sciences Journal*, 61(14), 2509-2519, doi:10.1080/02626667.2016.1154149.

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 & Software*, 20(1), 19-32, doi:10.1016/j.envsoft.2003.12.006.

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, doi:10.2166/hydro.2007.104.

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, doi:10.1016/j.jhydrol.2005.09.008.

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., A. E. Raftery, and S. Richardson (2001), Variable selection and Bayesian model averaging in case-control studies, *Statistics in Medicine*, 20(21), 3215-3230.

Vieira, D. A., and W. Wu (2002), *One-dimensional Channel Network Model CCHE1D Version 3.0 – User's Manual*, Technical Report No. NCCHE-TR-2002-2, National Center for Computational Hydroscience and Engineering, University of Mississippi.

Vrugt, J. A., and B. A. Robinson (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., 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, doi:10.5194/hess-15-3701-2011.

Vrugt, J. A., C. G. H. Diks, and M. P. Clark (2008), Ensemble Bayesian model averaging using Markov chain Monte Carlo sampling, *Environmental Fluid Mechanics*, 8, 579-595, doi:10.1007/s10652-008-9106-3.

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, doi:10.1515/IJNSNS.2009.10.3.273.

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.

Vrugt, J. A., W. Bouten, H. V. Gupta, and S. Sorooshian (2002), Toward improved identifiability of hydrologic model parameters: The information content of experimental data, *Water Resources Research*, 38, 1312, doi: 10.1029/2001WR001118.

Wallingford, H. R. (1990), *Sediment Transport, the Ackers and White Theory Revised*, Report SR237, HR Wallingford, England.

Wikle, C. K., and M. B. Hooten (2010), A general science-based framework for nonlinear spatio-temporal dynamical models, *Test*, 19 (3), 417-451, doi:10.1007/s11749-010-0209-z.

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 J. C. Crowe (2003), Surface-based transport model for mixed-size sediment, *Journal of Hydraulic Engineering*, 129(2), 120-128.

Williams, R. D., J. Brasington, M. Hicks, R. Measures, C. D. Rennie, and D. Vericat (2013), Hydraulic validation of two-dimensional simulations of braided river flow with spatially continuous ADCP data, *Water Resources Research*, 49(9), 5183-5205.

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.

Wong, M., and G. Parker (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, 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, doi:10.1061/(ASCE)0733-9429(2009)135:1(22).

Wu, W., G. C. Dandy, and H. R. Maier (2014), Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling, *Environmental Modelling & Software*, 54, 108-127, doi:10.1016/j.envsoft.2013.12.016.

Wu, W., S. S. Wang, and Y. Jia (2000), Nonuniform sediment transport in alluvial rivers, *Journal of Hydraulic Research*, 38(6), 427-434, doi:10.1080/00221680009498296.

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., H. V. Gupta, and S. Sorooshian (1996), Calibration of conceptual rainfall-runoff models: Sensitivity to calibration data, *Journal of Hydrology*, 181, 23–48.

Yapo, P. O., H. V. Gupta, and S. Sorooshian (1998), Multi-objective global optimization for hydrologic models, *Journal of Hydrology*, 204(1), 83-97.

Ye, M., S. P. Neuman, and P. D. Meyer (2004), Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff, *Water Resources Research*, 40, W05113, doi:10.1029/2003WR002557.

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, doi:10.1016/S0048-9697(99)00282-X.

Zhao, T., Q. J. Wang, J. C. Bennett, D. E. Robertson, Q. Shao, and J. Zhao (2015), Quantifying predictive uncertainty of streamflow forecasts based on a Bayesian joint probability model, *Journal of Hydrology*, 528, 329-340.

Zheng, Y., and A. A. Keller (2007), Uncertainty assessment in watershed-scale water quality modeling and management: 1. Framework and application of generalized likelihood uncertainty estimation (GLUE) approach, *Water Resources Research*, 43, W08407, doi:10.1029/2006WR005345.