Method for Assessing Impacts of Parameter Uncertainty in Sediment Transport Modeling Applications

Morgan D. Ruark[1]; Jeffrey D. Niemann, M.ASCE[2]; Blair P. Greimann, M.ASCE[3]; and Mazdak Arabi, M.ASCE[4]

Abstract: The predictions from a numerical sediment transport model inevitably include uncertainty because of assumptions in the model's mathematical structure, the values of parameters, and various other sources. In this paper, the writers aim to develop a method that quantifies the degree to which parameter values are constrained by calibration data and the impacts of the remaining parameter uncertainty on model forecasts. The method uses a new multiobjective version of generalized likelihood uncertainty estimation. The likelihoods of parameter values are assessed using a function that weights different output variables on the basis of their first-order global sensitivities, which are obtained from the Fourier amplitude sensitivity test. The method is applied to Sedimentation and River Hydraulics—One Dimension (SRH-1D) models of two flume experiments: an erosional case and a depositional case. Overall, the results suggest that the sensitivities of the model outputs to the parameters can be rather different for erosional and depositional cases and that the outputs in the depositional case can be sensitive to more parameters. The results also suggest that the form of the likelihood function can have a significant impact on the assessment of parameter uncertainty and its implications for the uncertainty of model forecasts. **DOI: 10.1061/(ASCE)HY.1943-7900.0000343.**

CE Database subject headings: Sensitivity analysis; Uncertainty principles; Sediment transport; Calibration; Parameters.

Author keywords: Sensitivity analysis; Uncertainty principles; Sediment transport; Calibration; Parameters.

Introduction

The use of numerical sediment transport models has dramatically expanded over the past three decades. One-dimensional sediment transport models in particular are widely used to determine possible impacts of watershed changes, evaluate water supply management, and predict impacts of proposed water resource systems on endangered species. Predictions from sediment transport models always entail uncertainty. Sources of uncertainty include: (1) simplifications in the model's representation of physical processes, (2) unknown initial and/or boundary conditions, (3) errors in the observations that are used to calibrate the model parameters, (4) errors in the values of model parameters, and (5) errors in model forcing (Clyde and George 2004; Gourley and Vieux 2006; Refsgaard et al. 2006; Murray 2007). For one-dimensional sediment transport models, this uncertainty can encompass orders of magnitude in the computed sediment load and the amount of material eroded or deposited at critical locations (Simons et al. 2000; Davies et al. 2002; Eidsvik 2004). Past research has focused on uncertainty arising from sediment transport models or formulas (Davies et al. 2002; Pinto et al. 2006) and the active erosional processes (Daebel and Gujer 2005; Harmel and King 2005; Jepsen 2006; Ziegler 2006) as well as methods to manage uncertainty (Osidele et al. 2003). Less attention has been paid to uncertainty throughout the entire parameter space, or global uncertainty (Chang et al. 1993), and the implications of parameter uncertainty. Models are typically calibrated by adjusting the parameters so that the model outputs reproduce a set of available observations. The performance of the model for the calibration period is usually reported, but little consideration is given to the extent to which the calibration data have constrained the values of the parameters and the behavior of the model for the forecast scenario.

Bayesian methods offer a formal way to assess impacts of parameter uncertainty (or other uncertainties) on model predictions (Clyde and George 2004; Kuczera et al. 2006). Bayesian methods require the modeler to specify a prior joint probability distribution for the uncertain parameters. The prior joint distribution is then combined with observations of model outputs from the calibration period to generate a posterior joint distribution for the parameters (Beven 2000). The updating of the joint distribution is done on the basis of a formal assessment of the likelihood of a set of parameter values given the observed model outputs (Clyde and George 2004). The posterior distribution of the parameter values is then used in the model for the forecast scenario to determine the implied distribution of model outputs. The key advantage of Bayesian methods is that they utilize well-defined theoretical foundations, including a formal likelihood function for updating the joint probability distribution (Clyde and George 2004; Kuczera et al. 2006). Key limitations of Bayesian methods are that they can require inversion of large matrices, which can be a computational burden, and they often employ a variety of simplifying statistical assumptions including normality, independence, and homoscedasticity (Stedinger et al. 2008) that are often violated in sediment transport modeling applications. For example, heteroscedasticity is well documented for discharge hydrographs (Sorooshian and Dracup 1980) and is

[1]Water Resources Engineer, CH2M Hill, 2020 SW 4th Ave., 3rd Floor, Portland, OR 97201-4958.

[2]Associate Professor, Dept. of Civil and Environmental Engineering, Colorado State Univ., Campus Delivery 1372, Fort Collins, CO 80523-1372 (corresponding author). E-mail: jniemann@engr.colostate.edu

[3]Hydraulic Engineer, Sediment and River Hydraulics Group, Bureau of Reclamation, Denver Federal Center, Bldg. 67, Denver, CO 80225.

[4]Assistant Professor, Dept. of Civil and Environmental Engineering, Colorado State Univ., Campus Delivery 1372, Fort Collins, CO 80523-1372.

Note. This manuscript was submitted on September 3, 2009; approved on October 14, 2010; published online on January 13, 2011. Discussion period open until November 1, 2011; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Hydraulic Engineering***, Vol. 137, No. 6, June 1, 2011. ©ASCE, ISSN 0733-9429/2011/6-623–636/$25.00.

likely to occur in sediment transport applications involving unsteady flow.

Generalized likelihood uncertainty estimation (GLUE) offers an alternative method to assess parameter uncertainty (Beven and Binley 1992). The GLUE methodology has been utilized for a variety of modeling applications including rainfall-runoff models (Freer et al. 1996; Blasone et al. 2008), groundwater models (Christensen 2003; Hassan et al. 2008), water quality models (Shirmohammadi et al. 2006), and atmospheric models (Page et al. 2004), but it has received little attention in sediment transport modeling. GLUE follows the Bayesian approach, but it utilizes an informal function to estimate the likelihoods of parameter values given a set of observations. The benefit of the informal likelihood function is that it can be selected on the basis of the model purpose (Mantovan and Todini 2006), and different likelihood functions are known to produce different uncertainty estimates (Freer et al. 1996; Beven 2000). However, Christensen (2003) and Stedinger et al. (2008) have demonstrated that previously used likelihood functions fail to reproduce the known posterior distributions of parameters for simple cases (normally and independently distributed errors). For such cases, these authors identify the appropriate likelihood function, but this function is not easily evaluated within the GLUE framework (Stedinger et al. 2008).

Another challenge in the application of GLUE to sediment transport modeling is the need to evaluate the model performance with respect to multiple objectives or outputs such as sediment size, sediment load, stream velocity, channel geometry, and bed profile. Available methods of computing multiobjective likelihood functions include the use of fuzzy set theory (Beven and Binley 1992; Yang et al. 2004), the successive combination of likelihoods (through multiplication), and the weighted addition of likelihoods (Beven 2000). Such approaches have been addressed elsewhere (Yapo et al. 1998; Mo and Beven 2004; Chahinian and Moussa 2007). One problem in the application of a multiobjective approach is the need to assign weights (or the equivalent) for each model output. Such weights are often set ad hoc on the basis of the model's purpose, but arbitrary weighting can lead to misleading results. In particular, a selected model output may be independent of a particular parameter. Thus, it would be inappropriate to strongly weight the performance with respect to that output variable when evaluating different values of the parameter.

The objective of this paper is to explore the use of a GLUE-based method to assess the implications of parameter uncertainty on the outputs of a one-dimensional sediment transport model. This method is intended to measure how well parameter values have been constrained during model calibration and to determine how the remaining uncertainty in the parameter values affects the model outputs under forecast scenarios. As such, the method is intended to aid model development rather than to provide a formal assessment of uncertainty for project evaluation. The likelihood function used in this paper is developed on the basis of the one described by Christensen (2003) and Stedinger et al. (2008). Global sensitivity analysis (GSA) is employed as a way to weight multiple model outputs. GSA and GLUE have been coupled previously (Ratto et al. 2001), but not for the purpose of weighting multiple outputs. The GSA-GLUE method is applied to Sedimentation and River Hydraulics—One Dimension (SRH-1D) models of two physical experiments (an erosional case and a depositional case) to identify how well the parameters are constrained by the calibration and to partially explore the implications of various assumptions included in the GSA-GLUE methodology.

The outline of the paper is as follows. The next section, Methodology, details how the GSA and GLUE methods are combined to assess the implications of parameter uncertainty. Then, the sediment transport model that is used with the GSA-GLUE method (SRH-1D) is described. Next, the Experiments section summarizes the physical experiments that are simulated with the model. The Results section discusses the main results of the GSA-GLUE approach, and the Analysis section evaluates the main assumptions of the method. Finally, the paper closes by summarizing the conclusions and future directions for research.

Methodology

The GSA-GLUE method developed in this paper includes three main steps. The first step is the GSA. In this step, a sample of parameter sets is generated from a jointly uniform distribution within specified ranges. The model is then run for the calibration period using each parameter set in the sample. On the basis of an analysis of the model results, the sensitivity of each model output to each parameter is estimated. To reduce the number of required simulations, the GSA is performed using the Fourier amplitude sensitivity test (FAST), which generates the parameter sets in a specific way (Cukier et al. 1973; Saltelli et al. 1999), although they are still approximately uniformly distributed. The second step is the application of the GLUE methodology to calculate the likelihood associated with each parameter value, and from those likelihoods, to determine updated likelihood distributions for each parameter. Likelihoods are calculated on the basis of the model's ability to reproduce the observations for the calibration period when each parameter set is used. Because sediment transport models typically produce multiple model outputs of interest (e.g., sediment size, channel profile), the sensitivities calculated in the first step are used to weight the different outputs in the calculation of the likelihoods. This procedure places greater importance on reproducing outputs that are more sensitive to a particular parameter. The third step is to use Latin hypercube sampling (LHS) with the cumulative likelihood distributions of the parameters to generate a new sample of parameter sets (Chang et al. 1993; Hall et al. 2005). The model is run for the forecast period using these parameters sets, and histograms are calculated for the model outputs. These histograms allow an assessment of the implications of the remaining parameter uncertainty on the forecasts of the model. The following subsections describe each of the three steps (GSA, GLUE, and LHS) in greater detail.

Global Sensitivity Analysis

Sensitivity analysis usually aims to quantify how much an output of a model changes when a model parameter (or input) is varied (Saltelli et al. 2008). Whereas local sensitivity analysis evaluates these changes around a set of base values for the parameters, GSA assesses these changes across specified ranges of parameter values. Local analyses usually measure the sensitivity with an index that is related to the partial derivative of the output with respect to the parameter. Unfortunately, such measures are only well defined if the output is linearly dependent on the parameter (Saltelli et al. 2008). In contrast, the GSA described in this paper uses a variance-based measure of sensitivity, which partially overcomes the linearity assumption (Chan et al. 1997). Two measures of sensitivity are calculated. One is the first-order index S_x, which is defined as

$$S_x = \frac{\mathrm{Var}[E(Y|X)]}{\mathrm{Var}(Y)} \tag{1}$$

where $\mathrm{Var}(Y)$ = total variance of the model output Y when all the parameters are varied within their specified ranges; $E(Y|X)$ = expected value of output Y for a particular value of

parameter X; and $\mathrm{Var}[E(Y|X)]$ = variance of $E(Y|X)$ when X is varied over its allowed range. The numerator in Eq. (1) describes the variation in the expected value of Y that occurs within the specified range of X. It aims to characterize how much variation in Y is induced by variations in X. The denominator describes the total variation in the output Y when all parameters are varied.

The second measure is the total-order index S_{Tx}, which can be written as

$$S_{T_x} = 1 - \frac{\mathrm{Var}[E(Y|\tilde{X})]}{\mathrm{Var}(Y)} \tag{2}$$

where $\mathrm{Var}[E(Y|\tilde{X})]$ = variance of the expected value of Y when all inputs except X are held constant. The first-order index evaluates the direct contribution that a parameter makes to the variability of the output. If a model is strictly additive with respect to its parameters, then the first-order indices will sum to one (Saltelli et al. 2008). In more complex models, the effect of a parameter on the model output may be modulated by the other parameter values. The total-order index evaluates the total contribution of a parameter to the output variability when all interactions between parameters are included.

FAST offers an efficient way to estimate these variance-based measures of sensitivity. FAST was initially developed to study first-order effects (Cukier et al. 1973) and was later expanded to include the total-order effects (Saltelli et al. 1999). The computational efficiency of FAST is achieved by varying all parameters of interest simultaneously rather than one by one. The parameters are varied at noninterfering frequencies (Cukier et al. 1973; Schaibly and Shuler 1973) within the ranges that are specified by the modeler. The generated sequence of parameter sets is used in the model to generate an associated sequence of model responses. The model response sequence is then decomposed using a Fourier transform, which determines the variance that is associated with each frequency. By considering certain groups of frequencies, the first-order and total-order sensitivity indices can be calculated for each parameter (Saltelli et al. 1999). The sample size, which is the total number of simulations to be performed, must be specified. The sensitivity estimates from FAST asymptotically converge to the definitions given in Eqs. (1) and (2) as the sample size becomes large.

In the present analysis, FAST is used to calculate the importance of each parameter to variability in each model output. Use of FAST also allows screening of parameters to remove those with little influence on model outputs. In particular, if the first- and total-order sensitivities of all the outputs to a particular parameter are small, then the parameter can be treated as a constant in the analysis to reduce computation time. In addition, the sensitivity indices are used in the likelihood function (described in the following section) to weight the performance of the model in reproducing different model outputs.

Generalized Likelihood Uncertainty Estimation

The GLUE method is used next to determine revised, or posterior, distributions for the parameters. Running the model with each parameter set and comparing its performance to the observed system behavior will obtain information about the likelihood that the parameter set is correct. In typical applications of GLUE, a Monte Carlo sampling of a uniform distribution is used to determine the parameter values that are supplied to the model. However, the samples produced by FAST are also approximately uniform and can be used in GLUE (Ratto et al. 2001).

The likelihood of a parameter value is evaluated on the basis of the model's ability to reproduce observations when that parameter value is used. Many previous papers have used the Nash-Sutcliffe coefficient of efficiency (NSCE) as the basis of the likelihood function (Beldring et al. 2003; Arabi et al. 2007; Engeland et al. 2006; Engeland and Gottschalk 2002; Uhlenbrook and Sieber 2005). NSCE is calculated as

$$\mathrm{NSCE} = 1 - \left[\frac{\sum_{j=1}^{\ell}(O_j - M_j)^2}{\sum_{j=1}^{\ell}(O_j - \bar{O})^2}\right] \tag{3}$$

where O = observed value and M = model's value; $\bar{O}$ = average of the observed values; j = index of locations (or times); and ℓ = total number of locations (or times) where observations are available (Nash and Sutcliffe 1970; Legates and McCabe 1999). NSCE is 1 when the model perfectly reproduces the observations, and it decreases as the model performance deteriorates.

Recent papers (Mantovan and Todini 2006; Stedinger et al. 2008) have argued that arbitrary likelihood functions, such as NSCE, can produce arbitrary results in the GLUE methodology. Stedinger et al. (2008) demonstrated this argument by applying GLUE with a likelihood function developed on the basis of NSCE to a simple case where the appropriate likelihood function is known from basic statistics. The case they considered is linear regression with normal, independently distributed errors with constant variance (Stedinger et al. 2008). In that case, they argued that the appropriate way to calculate the likelihood L for a given parameter set is

$$L = K \exp\left[-\frac{\ell}{2} \cdot \frac{\sum_{j=1}^{\ell}(O_j - M_j)^2}{\sum_{j=1}^{\ell}(O_j - M_j^{\mathrm{MLE}})^2}\right] \tag{4}$$

where M represents the model's value when a particular parameter set is used; M^{MLE} = model's value when the parameters are obtained from the maximum likelihood estimator (MLE), and K = normalization constant that ensures that all the likelihoods sum to 1. This likelihood function has some similarities to NSCE, but it includes two key differences. First, the denominator in Eq. (4) implies that the likelihood is assessed by comparing the performance of a given parameter set with that of the MLE parameter set (Stedinger et al. 2008). Second, the use of ℓ as a coefficient accounts for the number of independent observations that are available to constrain the parameter values (Stedinger et al. 2008). Including ℓ means that poor performance is penalized more when many observations are available. In the present application, a likelihood function is utilized that is similar to the one in Eq. (4), with two key differences. The first difference is that the errors are not expected to be independent between observation locations, in contrast to the assumptions underlying Eq. (4). Thus, the coefficient ℓ is replaced by an effective number of independent locations m. The second difference is the need to account for multiple output variables or objectives in calculating the likelihoods. This issue is confronted by using a weighted sum of likelihoods. In the case where three output or response variables are of interest, the resulting likelihood function is

$$L = K\left\{w_1 \exp\left[-\frac{m}{2} \cdot \frac{\sum_{j=1}^{\ell}(O_{1,j} - M_{1,j})^2}{\sum_{j=1}^{\ell}(O_{1,j} - M_{1,j}^{\mathrm{MLE}})^2}\right] + w_2 \exp\left[-\frac{m}{2} \cdot \frac{\sum_{j=1}^{\ell}(O_{2,j} - M_{2,j})^2}{\sum_{j=1}^{\ell}(O_{2,j} - M_{2,j}^{\mathrm{MLE}})^2}\right] + w_3 \exp\left[-\frac{m}{2} \cdot \frac{\sum_{j=1}^{\ell}(O_{3,j} - M_{3,j})^2}{\sum_{j=1}^{\ell}(O_{3,j} - M_{3,j}^{\mathrm{MLE}})^2}\right]\right\} \tag{5}$$

where the subscripts 1, 2, and 3 distinguish the three response variables and the w's are the individual weights. By using the same m in all three terms in Eq. (5), it is implicitly assumed that the effective number of independent locations is the same for all three variables. The weight for a given output is calculated as the first-order sensitivity of that output to the parameter of interest divided by the sum of the first-order sensitivities of all three outputs to that same parameter. A different likelihood is calculated for each parameter included in a parameter set because the weights depend on the parameter being considered. The weights are calculated using the first-order sensitivities from the GSA, but they could also be calculated using the total-order sensitivities. First-order weights are selected because they could eventually be estimated using methods that are faster than FAST (Saltelli and Bolado 1998; Gatelli et al. 2009), such as random balance designs (Tarantola et al. 2006). The impact of choosing the first-order sensitivity over the total-order sensitivity is evaluated subsequently.

The performance of the MLE is also required to evaluate the likelihood function in Eq. (5). The MLE is not generated as part of the GLUE methodology. As an approximation, it is assumed that the best performing parameter set in the sample is equivalent to the MLE. This assumption is expected to be better when a large number of parameter sets are generated. To select the MLE, performance is judged by finding the minimum of total error ε where

$$\varepsilon = \frac{1}{\sigma_{O_1}^2}\sum_{j=1}^{\ell}(O_{1,j} - M_{1,j})^2 + \frac{1}{\sigma_{O_2}^2}\sum_{j=1}^{\ell}(O_{2,j} - M_{2,j})^2 + \frac{1}{\sigma_{O_3}^2}\sum_{j=1}^{\ell}(O_{3,j} - M_{3,j})^2 \quad (6)$$

and

$$\sigma_{O_1}^2 = \frac{1}{\ell - 1}\sum_{j=1}^{\ell}(O_{1,j} - \bar{O}_1)^2 \quad (7)$$

and

$$\bar{O}_1 = \frac{1}{\ell}\sum_{j=1}^{\ell} O_{1,j} \quad (8)$$

The variances and averages for output variables 2 and 3 are calculated using expressions similar to Eqs. (7) and (8).

The likelihood function in Eq. (5) also requires calculation of the normalization constant, which is found from the constraint that all the likelihoods for a parameter should sum to 1. In practical terms, preliminary likelihoods are calculated by neglecting K in Eq. (5). Then, the sum of these likelihoods is calculated, and the preliminary likelihoods are divided by the sum to determine the final likelihoods. The cumulative likelihood for a selected value of a parameter is determined by summing all the likelihoods associated with values that are smaller than or equal to the selected value. The cumulative likelihoods can be used to generate a posterior cumulative likelihood distribution for each parameter.

The limitations of the GLUE methodology should be emphasized. The likelihood function in Eq. (5) is assumed rather than derived from a particular set of statistical assumptions. In addition, calculation of separate cumulative likelihood distributions for each parameter neglects correlation or dependence between the most likely values of different parameters. Anecdotal results (not shown in this paper) suggest that dependencies might occur, so future research should consider methods to account for such dependencies. These limitations disallow using this methodology to rigorously assess quantitative uncertainty in model responses. However, the methodology is still expected to be useful to roughly assess the extent to which various parameters have been constrained by a calibration exercise and the related objectives given previously.

Latin Hypercube Sampling

The third and final step of the methodology is to use the posterior cumulative distributions of the parameters in the model to simulate the forecast period and to determine the associated distributions for the model outputs. Latin hypercube sampling is used to sample the marginal posterior distribution of each parameter (Chang et al. 1993; Hall et al. 2005). In contrast to Monte Carlo sampling, which generates random values from the distribution, LHS explores the parameter space using regularly spaced percentiles from the distribution. LHS is used in this paper because previous research has shown that smaller sample sizes can be used to characterize a distribution for LHS than can be used for for Monte Carlo simulations (McKay et al. 1979). Even so, the required number of simulations at this stage of the analysis can be rather large if numerous parameters are treated as uncertain. To reduce the number of simulations, the parameters can be screened. Parameters that had little impact on the model results in the calibration period (on the basis of the GSA) can be assigned to the midpoint of the allowable range. The remaining parameters are treated as uncertain and sampled using LHS. In the LHS scheme, the posterior cumulative likelihood function for each parameter is obtained from the GLUE methodology described previously. The cumulative likelihood scale is divided into a selected number of equally sized bins, and the midpoints of those bins are determined. Then, the cumulative likelihood function is used to find the parameter value associated with each midpoint. Because the posterior distributions are typically nonuniform, the parameter values will be irregularly spaced. The values for each parameter are then combined with those for every other parameter so that every combination is included in the sample. The generated parameter sets are then used in the model for the forecast period to determine the associated histograms of model responses.

SRH-1D

The GSA-GLUE methodology described in the previous section is tested using the SRH-1D model, which is currently used by the Bureau of Reclamation to simulate flows and sediment transport in channels and river networks with or without movable boundaries (Huang and Greimann 2007). The model can simulate steady or unsteady flow and can treat cohesive and noncohesive sediment. The model applications considered in this paper use only steady flow and noncohesive sediment.

SRH-1D uses one-dimensional flow calculations, including the standard step energy method for steady gradually varied flow (Huang and Greimann 2007). Between adjacent cross sections (j and $j+1$), the energy equation is written as

$$z_{j+1} + \alpha_{j+1}\frac{v_{j+1}^2}{2g} - z_{j+1} - \alpha_{j+1}\frac{v_{j+1}^2}{2g} - h_f - h_c = 0 \quad (9)$$

where z represents the water surface elevation; α = kinetic energy flux correction coefficient; v = average velocity at the cross section; g = gravitational acceleration; h_f represents friction loss; and h_c represents contraction or expansion losses. Evaluation of the friction loss in Eq. (9) ultimately requires use of Manning's equation and specification of Manning's roughness coefficient n, which is considered a model parameter.

SRH-1D simulates sediment transport using three main elements: sediment routing, bed material mixing, and cohesive sediment consolidation (if cohesive sediment is present). For sediment

routing, SRH-1D can use either unsteady sediment routing or Exner equation routing. Because steady flow is considered in this paper and the bed changes are primarily driven by bed load, the Exner equation is used, and mass conservation can be written as

$$\frac{\partial Q_s}{\partial x} + \varepsilon \frac{\partial A_d}{\partial t} - q_s = 0 \tag{10}$$

where Q_s = volumetric sediment discharge; ε = volume of sediment per unit bed layer volume (related to porosity); A_d = volume of bed sediment per unit length; and q_s = lateral sediment inflow per unit length. The Exner equation is integrated over control volumes associated with cross sections and applied separately for each sediment size fraction. Because the cross sections might be closely spaced in some cases, SRH-1D does not assume that the sediment discharge equals the transport capacity. Rather, it assumes the capacity is reached over some length controlled by the total adaptation length. The total adaptation length L_{tot} is computed on the basis of Greimann et al. (2008) as follows:

$$L_{\text{tot}} = f_s L_b + (1 - f_s)\frac{Q}{\zeta W w_f} \tag{11}$$

where L_b = adaptation length for bed load; ζ = suspended sediment recovery factor; f_s = fraction of suspended load relative to the total load; Q = flow rate; W = channel width; and w_f = fall velocity. The value of L_b is determined as

$$L_b = b_L h \tag{12}$$

where b_L = user-defined bed-load adaptation length parameter; and h = average depth at the cross section. Separate values of ζ are used for deposition and scour:

$$\zeta = \begin{cases} \zeta_d & \text{if deposition} \\ \zeta_s & \text{if erosion} \end{cases} \tag{13}$$

and both values (ζ_d and ζ_s) are considered model parameters. The fraction of suspended load f_s was previously found to be primarily a function of the suspension characteristic Z. Greimann et al. (2008) derived the following empirical function for f_s:

$$f_s = \min(1, 2.5e^{-Z}) \tag{14}$$

where $Z \equiv w_f/(\kappa u_*)$, κ = von Kármán constant (0.4); and u_* = friction velocity.

The transport capacity expression used in this paper is Parker's (1990) gravel equation, which ultimately requires specification of a reference or critical shear stress (θ_r) and a hiding factor (λ), which account for the effect of hiding and sheltering of gravel mixtures. The reference shear stress and hiding factor are treated as model parameters for the purposes of this paper. In Parker's original formulation, $\theta_r = 0.386$, and $\lambda = 0.905$. However, it is probable that no unique set of values applies to all situations, and a wide range of values have been suggested for gravel bed rivers (Buffington and Montgomery 1997).

Bed material mixing is modeled by dividing the bed into an active layer and a series of underlying inactive layers. Erosion and deposition of sediment can only occur from the active layer. Each layer is considered homogeneous within its depth. The active layer thickness is determined as the product of the geometric mean of the largest size class and the active layer thickness multiplier (n_{alt}), which is another model parameter. When erosion occurs, the active layer shifts downward and material from the underlying layers becomes part of the active layer. When deposition occurs, the active layer shifts up and material becomes classified as the top inactive layer. As part of the bed material mixing, the user must specify the weight of bed-load fractions (χ), which is a parameter that controls the weighting of the bed-load grain-size distribution in the transfer of material between the active layer and the underlying layer. If $\chi = 0$, then the grain-size distribution of the sediment that is transferred to the sublayer is equal to the active layer grain-size distribution. If $\chi = 1$, then the grain-size distribution of the sediment that is transferred to the sublayer is equal to the bed-load grain-size distribution.

Table 1. Selected Bounds for the Uniform Distributions Describing the Eight Parameters

Parameter	Minimum value	Maximum value
Critical shear stress (θ_r)	0.01	0.06
Hiding factor (λ)	0	1
Active layer thickness multiplier (n_{alt})	0.1	2
Deposition recovery factor (ζ_d)	0.05	1
Scour recovery factor (ζ_s)	0.05	1
Bed-load adaptation length (b_L)	0	10
Weight of bed-load fractions (ξ)	0	1
Manning's roughness coefficient (n)	0.015	0.065

In the end, eight parameters are treated as uncertain in this analysis: Manning's roughness coefficient (n), critical shear stress (θ_r), hiding factor (λ), deposition recovery factor (ζ_d), scour recovery factor (ζ_d), bed-load adaptation length (b_L), active layer thickness multiplier (n_{alt}), and the weight of bed-load fractions (χ). None of these parameters is easily measurable in the field, and they can vary significantly from case to case. Thus, they are typically calibrated. Table 1 shows the selected minimum and maximum values of each of these parameters used in this analysis. These ranges were chosen because they represent a reasonable range of possible parameter values across various model applications.

SRH-1D produces a large number of outputs, which are available at multiple locations and times for a given simulation. For this study, the model response variables of interest were selected to be the length-averaged median grain size, flow velocity, and bed profile elevation. The length-averaged median grain size $\bar{d}_{50}$ is defined as

$$\bar{d}_{50} \equiv \frac{\sum_{j=1}^{\ell}[(d_{50_j} + d_{50_{j+1}})/2]\Delta L_{j,j+1}}{L_{\text{total}}} \tag{15}$$

where $\Delta L_{j,j+1}$ = length between cross sections j and $j+1$; d_{50_j} = median grain size at cross section j; and L_{total} = total length of the reach (the sum of all ΔLs). Similar expressions are used to calculate the length-averaged flow velocity and bed elevation. In those expressions, the average flow velocity and average bed elevation at each cross section is used in place of the median grain size in Eq. (15). It is assumed that observations are available for these three outputs for a so-called calibration period, and it is assumed that one would want to make predictions for these (unobserved) outputs for a forecast period.

Experiments

The model was applied to two flume experiments. One experiment is an erosional case, and the other is a depositional case. These experiments were chosen because volumetric flow rate, sediment supply, initial bed geometry, and initial bed material are known for both experiments. Thus, there is little uncertainty about the system configuration or the model inputs.

The Ashida and Michiue (1971) experiment was designed to simulate riverbed degradation and scour downstream of a dam. The flume was 0.8 m wide and 20 m long. The experiment used in this paper was called Run 6 by the writers. In this case, the initial bed slope was 0.01 m/m (1%), and a sand-to-gravel particle size distribution was used for the bed material with sizes ranging from 0.2 to 10 mm and an initial median diameter of 1.5 mm. A clear-water discharge of 0.0314 m^3/s was applied for the 10 h experiment.

Unfortunately, the observations that characterize the resulting system behavior are rather limited. The resulting degradation was measured at three locations (7, 10, and 13 m from the downstream end of the flume) at the beginning of the experiment and at hours 1, 2, 4, and 10. Bed gradation was also measured at three locations (1, 10, and 13 m from the downstream end of the flume) at the beginning and end of the experiment. Because of the lack of extensive data, output from a calibrated SRH-1D model was used in place of physical observations when evaluating the parameter uncertainty. This approach implies that any disagreement between model simulations performed as part of the parameter uncertainty analysis and the "observed" values is attributable to errors in the parameter values. This approach also allows for variation in the amount of observations supplied to the method to determine the impact on the results. It is not intended to be the manner in which the method is used in future applications. The model was manually calibrated using comparisons with both the observed bed profile and the observed bed grain-size distribution. The upstream boundary condition was set to zero sediment inflow, and the downstream boundary condition specified the water surface elevation and allowed sediment outflow. Actual observations of the bed profile were used as the initial conditions. Cross-section spacing was 0.5 m, resulting in 41 total cross sections, and grain sizes were broken into nine classifications. Fig. 1(a) compares the bed profile simulated by the calibrated model (using parameter values given in Table 2) with the observations. The calibrated model compares well with the observations through hour 2. After hour 2, the model overestimates the erosion rate. As erosion happens in this experiment, some bedform development occurs, which implies a temporal variation in Manning's n that is not properly captured by the model (the roughness coefficient is a parameter in SRH-1D and is assumed to remain constant during the experiment).

For the analysis of parameter uncertainty, the Run 6 experiment was divided into a calibration period from 0 to 2 h and a forecast period from 2 to 10 h. The forecast period has identical conditions to the calibration period aside from the initial condition. The initial conditions for all forecast simulations in the analysis are the values obtained from the calibrated model (i.e. the "observations").

The Seal et al. (1997) experiment was designed to evaluate downstream fining of poorly sorted sand and gravel in a narrow channel and to simulate deposition and armoring processes. Their experiments consisted of three separate laboratory flume setups (Runs 1, 2, and 3). The flume used in all three experiments was 0.3 m wide and 45 m long with an initial slope of 0.002 m/m (0.2%). A discharge of 0.049 m^3/s was applied at the upstream end of the flume. The durations of the individual setups were 16.83, 32.4, and 65 h, respectively. For each setup, a sand-to-gravel particle-size distribution was used for the sediment feed with sizes from 0.125 to 64 mm. Sediment feed rates for the three experiments varied from 0.05 to 0.19 kg/s. The resulting profile was measured at 18 locations every half hour, hour, and two hours for Runs 1, 2, and 3, respectively. Sediment sizes of the surface were measured at the end of each experiment using standard point counts of 100 grains for 8 to 10 samples over the length of the deposit along the flume. Subsurface sampling was also conducted at the end of each experiment.

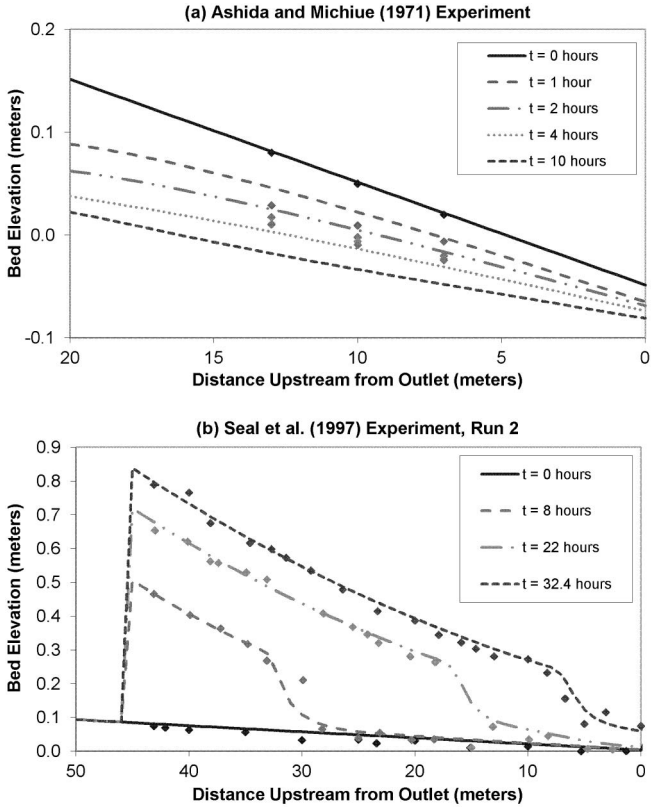

Fig. 1. Observed bed elevations (points) and the bed profiles produced by the calibrated models (lines): (a) Ashida and Michiue (1971); (b) Seal et al. (1997) experiments

For the Seal et al. (1997) experiment, Run 2 was used as the calibration case. To be consistent with the erosional experiment, a calibrated SRH-1D model was developed for Run 2 and used as "observations." The upstream boundary condition is the specified feed rate for sediment inflow, and the downstream boundary conditions are the specified water surface elevation and allowed sediment outflow. Actual observations were used as the initial conditions for the calibration simulation. Cross-section spacing was 1 m for a total of 56 cross sections (this total length is greater than the actual length of the flume). Nine grain-size classifications were used. The calibrated model was developed by comparing the model results with the observed bed profile and sediment-size distributions. Fig. 1(b) compares the bed profile simulated by

Table 2. Parameter Values for the Two Calibrated Models That Are Used to Generate the Output Values That Are Treated as Observations

Parameter	Ashida and Michiue (1971) experiment	Seal et al. (1997) experiment
Critical shear stress (θ_r)	0.0386	0.0386
Hiding factor (λ)	0.905	0.905
Active layer thickness multiplier (n_{alt})	1	1
Deposition recovery factor (ζ_d)	0.25	1
Scour recovery factor (ζ_s)	1	1
Bed-load adaptation length (b_L)	5.0	0.10
Weight of bed-load fractions (ξ)	0	0
Manning's roughness coefficient (n)	0.027	0.022

the calibrated model (using parameter values in Table 2) with the experimental observations. The calibrated model compares well with the observations, with the largest discrepancies occurring near the downstream end of the depositional wedge. Runs 1 and 3 were both used as forecast cases for the analysis of parameter uncertainty. For both forecast cases, the actual observations were used as the initial conditions for the simulations. These runs have the same volumetric flow rate (0.049 m^3/s) as the calibration case (Run 2). However, Runs 1 and 3 have sediment feed rates of 0.09 and 0.05 kg/s and durations of 16.8 and 65 h, respectively.

Results

Global Sensitivity Analysis

The FAST method, as previously described, was applied to the calibration cases of the two physical experiments. The eight parameters identified previously were varied, and the three model outputs were evaluated. Both applications of FAST used sample sizes of 5,000 simulations. Sample sizes down to 968 were found to produce the same qualitative results (i.e., identify the same parameters as being the most important in explaining the variance of each output) in testing with the Ashida and Michiue (1971) experiment. However, the numerical values for the sensitivities change somewhat as the sample size decreases within this range.

Fig. 2(a) plots the estimated contribution of each parameter to the total variance of the three output variables for the Ashida and Michiue (1971) experiment on the basis of the first-order sensitivity indices (sample size 5,000). To produce the partitions shown in a given column, the first-order indices were divided by the sum of the first-order indices and plotted as a percentage. Fig. 2(b) shows the results for a similar computation using the total-order indices. Both Figs. 2(a) and 2(b) suggest that four parameters are primarily responsible for producing variability in the length-averaged median grain size, velocity, and bed profiles. These parameters are the critical shear stress (θ_r), hiding factor (λ), active layer thickness multiplier (n_{alt}), and Manning's n. Only these parameters individually contribute more than 5% of the summed first-order sensitivities for each output. For the d50 output, the hiding factor is the parameter that produces the greatest sensitivity by far. This result should be expected, as hiding factor attempts to account for the differences in the mobility of different size fractions. Thus, it should have a clear impact on sediment-size distribution of the bed. For the velocity output, Manning's n is the parameter that produces the greatest sensitivity. This result reflects the relationship between Manning's n and velocity as stated in Manning's equation. For the bed profile output, critical shear stress is the parameter that produces the most sensitivity. Critical shear stress impacts bed profile through its role in determining the overall erodibility of the bed material. These same relationships hold whether the first-order or the total-order sensitivity is considered. In general, the contributions of the less important parameters are magnified when the total-order indices are considered. This behavior suggests that these parameters are primarily important because they affect the contributions of the more important parameters, such as critical shear stress, hiding factor, and Manning's n.

Figs. 2(c) and 2(d) show the equivalent results for the Seal et al. (1997) experiment. When considering the first-order sensitivities [Fig. 2(c)], four parameters again have contributions larger than 5%: critical shear stress, hiding factor, weight of bed-load fractions, and Manning's n. This list is the same as the erosion case, except that the weight of bed-load fractions (χ) replaces the active layer thickness multiplier. For the d50 output, the hiding factor plays a smaller role for the depositional case than it did for the erosional case, but it is still the parameter that produces the most variance. For the velocity output, Manning's n plays an even larger role for the depositional case than it did for the erosional case. For the bed profile, Manning's n now overtakes the critical shear stress as the

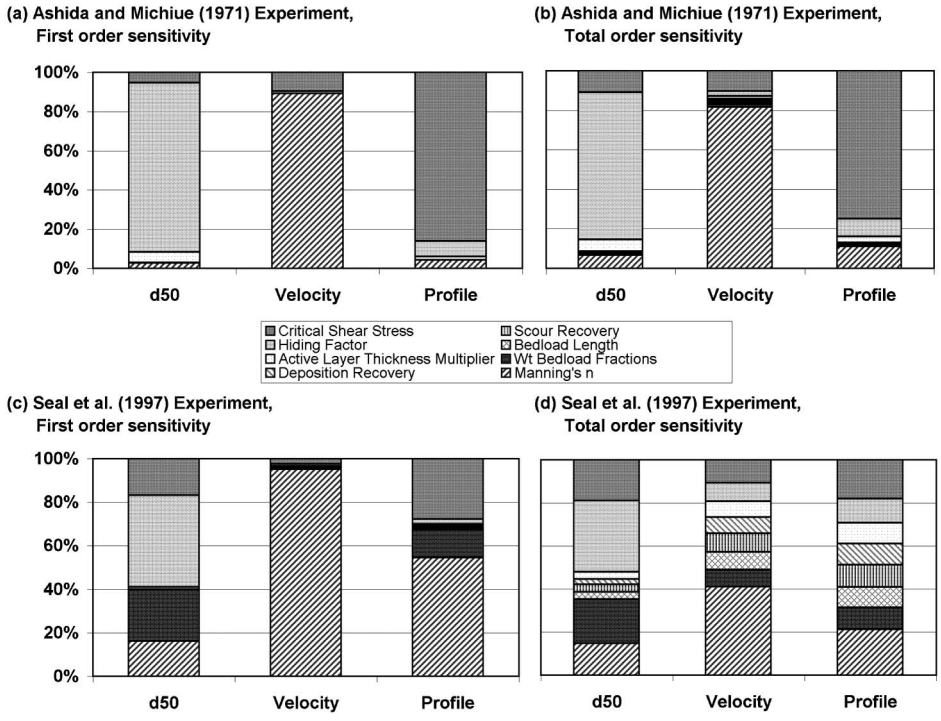

Fig. 2. First-order and total-order sensitivities for the SRH-1D models of the Ashida and Michiue (1971) and Seal et al. (1997) experiments. Each column refers to one of the three length-averaged output variables and each division in the columns refers to one of the eight model parameters. In the columns, the sensitivities to the parameters are stacked in the same order that the parameters appear in the legend

parameter that produces the most variance. The increased importance of Manning's n to the bed profile is expected because the flow velocity plays an important role in deposition.

The total-order sensitivities in Fig. 2(d) show more complex behavior than suggested by the first-order indices. Similar to the results for the erosional case, the less important parameters have a bigger role in the total-order indices than they do in the first-order indices. For the total-order sensitivity, a 5% threshold would identify the same four parameters as most important for the d50 output. However, for the velocity and bed profile outputs, this threshold would identify all parameters as being important. Increasing the threshold to 10% for the velocity and bed profile outputs would identify the same four parameters included in the first-order sensitivity, plus a scour recovery factor for the bed profile output. Comparing the total-order indices from the erosional and deposition cases [i.e., Figs. 2(b) and 2(d)] suggest that the depositional experiment is much more complex than the erosional experiment. For example, most of the bed profile variance comes from the critical shear stress in the erosional case, but for the depositional case, nearly all parameters have roughly comparable influences on the bed profile.

Generalized Likelihood Uncertainty Estimation

After the sensitivity analysis was completed, the GLUE method was used to calculate the posterior likelihood distributions for each parameter as described in the Methodology section. Recall that the likelihood function uses weights determined on the basis of the first-order sensitivity indices. The likelihood function also requires a value for m, the effective number of independent locations. It is assumed that $m = 1$, which is the most conservative value for this variable because it produces the largest estimate of the parameter uncertainty. The effect that m has on the results is evaluated subsequently. The solid lines in Fig. 3 show the posterior cumulative likelihood distributions of critical shear stress, hiding factor, active layer thickness multiplier, and Manning's n for the Ashida and Michiue (1971) experiment. Recall that the GSA identified these parameters as producing the most variance in the model outputs. The dashed lines show the uniform distributions that were assumed prior to simulation of the calibration period and application of the likelihood function. The steep sections in the posterior distributions indicate ranges with higher concentrations of likelihood. Such sections are seen in the distributions for critical shear stress, hiding factor, and Manning's n. In contrast, the distribution for active layer thickness multiplier does not exhibit such a steep section. This result suggests that the active layer thickness multiplier is more poorly constrained by the available observations than the other parameters, likely because no output is highly sensitive to active layer thickness multiplier (Fig. 2). The vertical lines in the figure indicate the true values for each parameter (i.e., the values used in the calibrated model that is used to generate the "observations"). For hiding factor and Manning's n, the true values fall within the region with the highest concentration of likelihood. For critical shear stress, the true parameter value falls in a range that is deemed unlikely by the analysis.

Fig. 4 shows the posterior cumulative likelihood distributions of critical shear stress, hiding factor, weight of bed-load fractions, and Manning's n for the Seal et al. (1997) experiment. These are the parameters found to produce the most variance in the model outputs for this experiment. Once again, steep sections are observed in the cumulative distributions of hiding factor and Manning's n, indicating that the most likely values of these parameters fall within relatively well-defined ranges. The true values for these parameters also fall within the ranges that are considered likely. Critical shear stress and weight of bed-load fractions do not exhibit such large steep sections, suggesting that these parameters are more poorly constrained by the available observations.

Latin Hypercube Sampling

After the posterior distributions were calculated, LHS was used to develop samples from them. For both physical experiments, the sample size was 1,296 parameter sets. Six values were generated for the four parameters that produced the greatest sensitivity in the outputs, whereas a single midpoint value was used for the remaining parameters [similar results were obtained for the Ashida and Michiue (1971) experiment when the number of values was as low as five for the four most important parameters]. Selecting only one value for the parameters that produced relatively little sensitivity in the outputs effectively neglects the uncertainty in the outputs produced by uncertainty in these parameter values. Screening out these parameters ultimately allows a much smaller number of simulations to be conducted for the forecasting period, which reduces computation time.

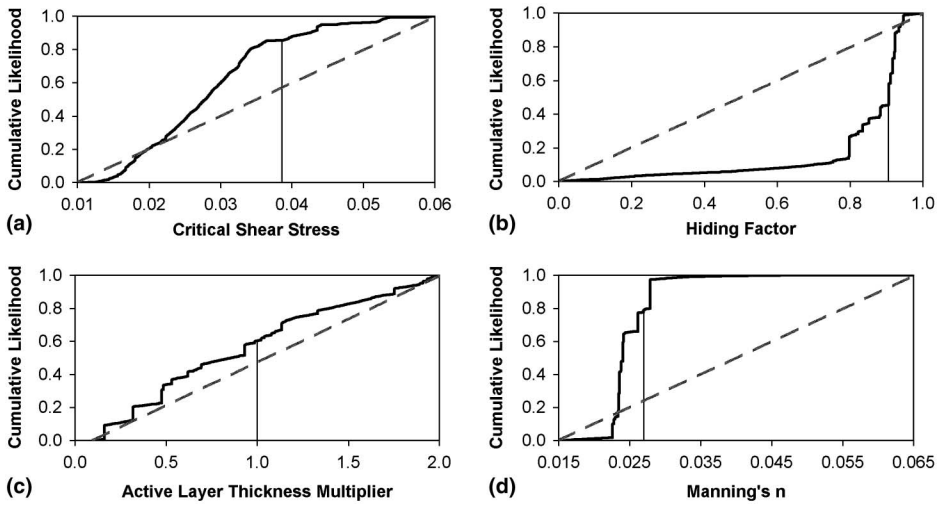

Fig. 3. Posterior cumulative likelihood distributions for: (a) critical shear stress; (b) hiding factor; (c) active layer thickness multiplier; (d) Manning's n for the Ashida and Michiue (1971) experiment. Dashed lines indicate the uniform distribution for each parameter that was assumed prior to simulation of the calibration period; vertical lines indicate the parameter values used in the calibrated model (i.e., the true values)

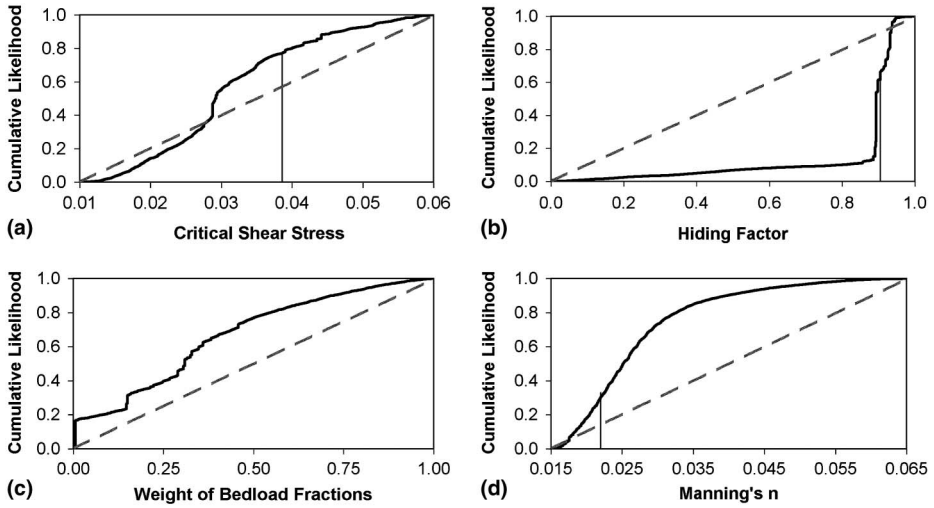

Fig. 4. Posterior cumulative likelihood distributions for: (a) critical shear stress; (b) hiding factor; (c) weight of bed-load fractions; (d) Manning's n for the Seal et al. (1997) experiment. Dashed lines indicate the uniform distribution for each parameter that was assumed prior to simulation of the calibration period; vertical lines indicate the parameter values used in the calibrated model (i.e., the true values). The true value for the weight of bed-load fractions is 0

For the Ashida and Michiue (1971) experiment, six values were generated from the posterior distributions of critical shear stress, hiding factor, active layer thickness multiplier, and Manning's n. Fig. 5 plots the histograms for the length-averaged d50, velocity, and bed profile when the parameter sets generated from LHS are used to simulate the forecast period (shown as solid lines). For comparison, the figure also shows the histograms of these output variables for the calibration period (shown as dashed lines),

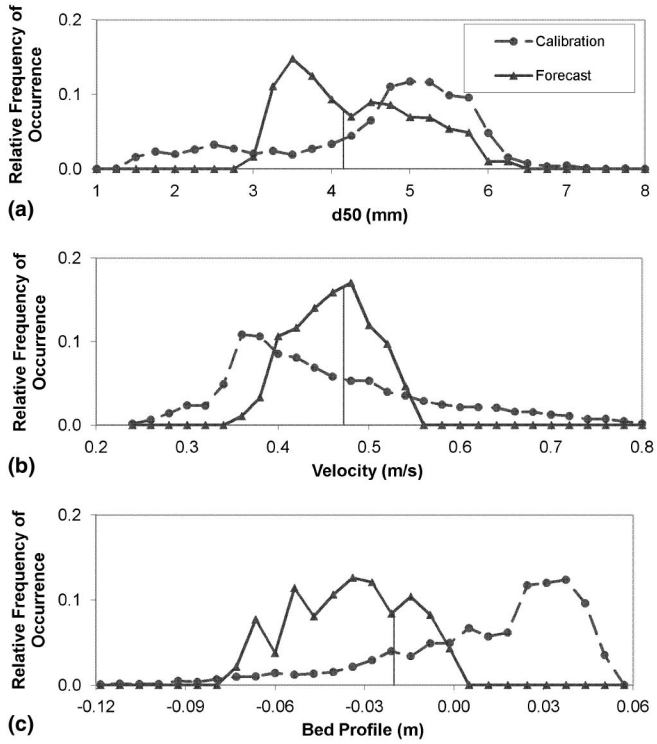

Fig. 5. Histograms of: (a) d50; (b) velocity; (c) bed profile for the calibration and forecast periods for the Ashida and Michiue (1971) experiment. Vertical lines indicate the output values produced by the calibrated model

where the parameters were generated from a uniform distribution via FAST. The vertical lines represent the true values for these outputs on the basis of application of the calibrated model to the forecast period. The histograms of the outputs have changed between the calibration and forecast periods in part because of differences in the initial conditions and elapsed simulation time. In particular, the duration of simulation was only 2 h for the calibration period whereas it was 8 h for the forecast period. Typically, one expects a wider range of output values for a longer simulation (i.e., the forecast period). However, the output histograms also reflect the narrower distributions for the parameters used for the forecast period. The most likely values of d50 range from 3 to 6 mm for the forecast period, and the true value (from the calibrated model) was 4.2 mm. The most likely values for velocity range from 0.35 to 0.54 m/s for the forecast period, and the true value was 0.47 m/s. The most common values of the bed profile range from -0.080 to -0.005 m, and the true value was -0.020 m. Thus, all of the histograms include the true value for the forecast period. In the case of velocity, the actual value is very near the value judged to be most likely from the histogram. Among the three output variables plotted, velocity is particularly interesting because it is not expected to vary between the calibration and forecast periods. Thus, the narrowing of the histogram between the calibration and forecast periods likely reflects the degree to which the parameters were constrained by the observations available for the calibration period.

For the Seal et al. (1997) experiment, six values were generated from the posterior distributions of critical shear stress, hiding factor, weight of bed-load fractions, and Manning's n using the LHS method. All other parameters were fixed at the median value from their respective posterior likelihood distributions. Fig. 6 plots the histograms for the length-averaged d50, velocity, and bed profile for the two Seal et al. (1997) cases that are considered as forecast scenarios (shown as solid lines). The figure also shows the histograms of these output variables for the calibration period (shown as dashed lines), where the parameters were generated from a uniform distribution via FAST. The vertical lines represent the true values for these outputs. Overall, the true values for the forecast cases typically fall within the histograms, although not always at the most likely value. The histograms also indicate a substantial range

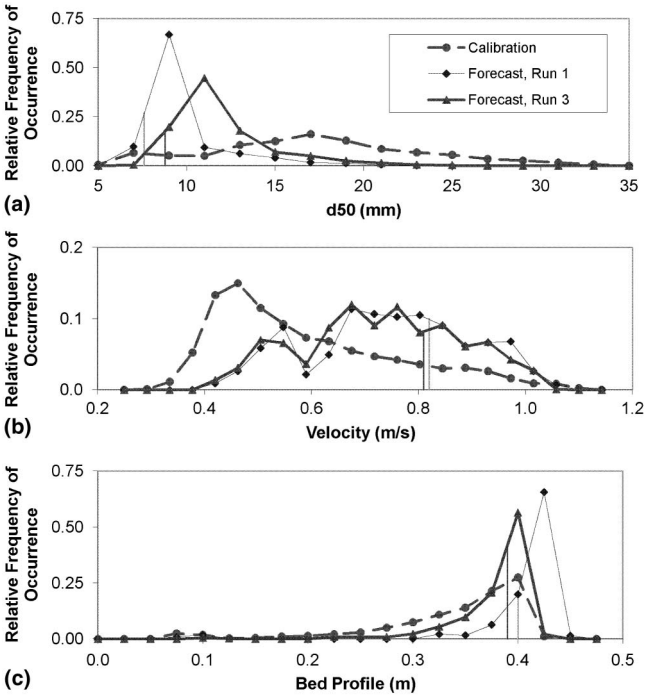

Fig. 6. Histograms of: (a) d50; (b) velocity; (c) bed profile for calibration and forecast cases for Seal et al. (1997) experiment. Vertical lines indicate the output values produced by the calibrated model

of plausible values for all three outputs under the forecast scenarios.

Analysis

The results described in the previous section rely on several decisions made in the application of the GSA-GLUE methodology. These decisions include: (1) the use of the first-order sensitivities rather than the total-order sensitivities to calculate the weights in the likelihood function, (2) the assumed effective number of independent observations m, (3) the mathematical form of the likelihood function, and (4) inclusion of observations of d50, velocity, and bed profile to constrain the parameters. The impact of each of these decisions is examined subsequently using the Ashida and Michiue (1971) experiment.

The impact of using the first-order sensitivities to determine the weights in the likelihood function in Eq. (5) is examined first. Fig. 7 shows the posterior cumulative likelihood distributions for critical shear stress, hiding factor, active layer thickness multiplier, and Manning's n, using both the first-order and total-order sensitivities to determine the weights. The resulting posterior distributions are very similar; the cumulative likelihood distributions for critical shear stress are visually indistinguishable. The same analysis was performed for the Seal et al. (1997) experiment, which produced posterior cumulative likelihood distributions for critical shear stress, hiding factor, weight of bed-load fractions, and Manning's n. Among these parameters, only the cumulative distribution for hiding factor showed a visible difference depending on the weighting used.

Another key assumption is the effective number of independent observations m. Previously, m was assumed to be 1 because of the expected dependence of the errors at different cross sections in a simulation. At this point, the practical effect of m on the results of the analysis is examined. Using the Ashida and Michiue (1971) experiment, posterior cumulative likelihood distributions for critical shear stress and hiding factor were generated using values of m varying from 1 to 20 in Figs. 8(a) and 8(b). As can be seen in these figures, an increase in m creates a posterior distribution with a more erratic shape, in which a very small number of parameter values begin to dominate the distribution. Overall, a larger value of m increases the likelihoods of the parameter values that produce results that are very similar to the observations and penalizes the parameter sets that produce more dissimilar results. The true values of critical shear stress and hiding factor from the calibrated model were 0.0386 and 0.905, respectively (Table 2). Whereas the true value for the hiding factor is located at a jump in the cumulative distribution [Fig. 8(b)], the true value for critical shear stress is located at a relatively flat portion of the distribution, or a region with a lower concentration of likelihood. The hazard of a large value

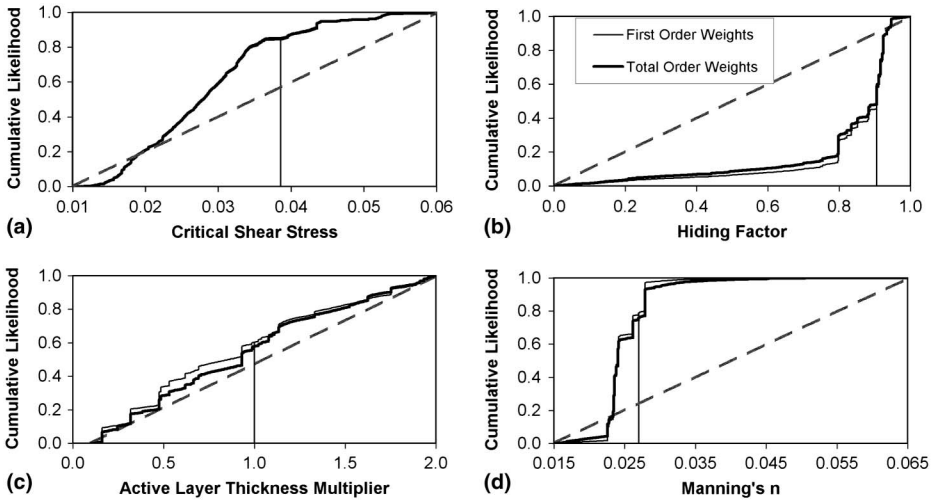

Fig. 7. Comparison of posterior cumulative likelihood distributions for: (a) critical shear stress; (b) hiding factor; (c) active layer thickness multiplier; (d) Manning's n when the likelihood function uses the first-order or total-order sensitivity. Dashed lines indicate the uniform distribution that was assumed for each parameter before the model was applied to the calibration period; vertical lines indicate the parameter values used in the calibrated model. All results are for the Ashida and Michiue (1971) experiment

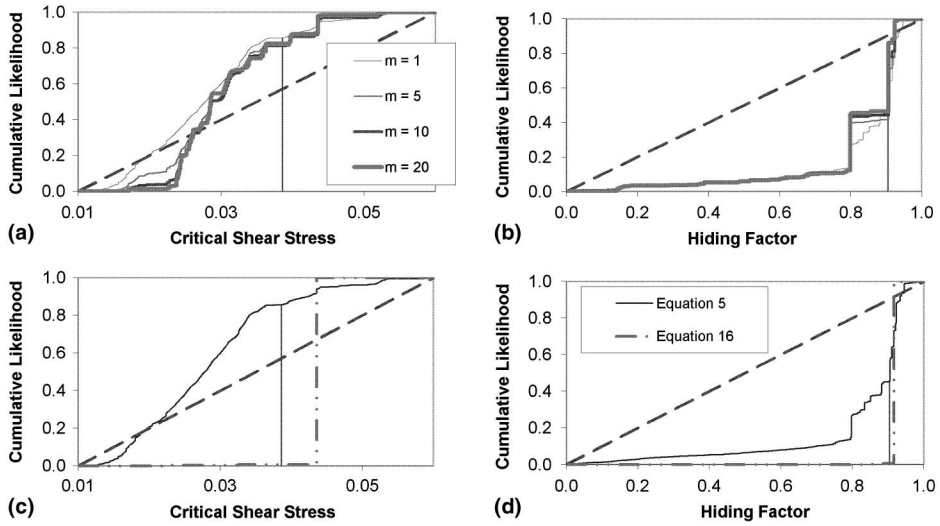

Fig. 8. Impact of the value of m, the effective number of independent observations, on the posterior cumulative likelihood distributions for: (a) critical shear stress; (b) hiding factor, and the impact of the mathematical form of the likelihood function on the posterior cumulative likelihood distributions for: (c) critical shear stress; (d) hiding factor. Dashed diagonal lines indicate the assumed initial uniform distribution for each parameter; vertical lines indicate the parameter values used in the calibrated model. All results are for the Ashida and Michiue (1971) experiment

of m is that the method might overpenalize small disagreements with the observations and miss plausible values of the parameter.

The form of the likelihood function was also assumed in generating the results in the previous section, on the basis of a conceptual extension of the likelihood function presented by Stedinger et al. (2008). Alternative likelihood functions could be devised. For example, one might instead calculate the likelihood by normalizing each output variable and then including it in the function presented by Stedinger et al. (2008). A likelihood function developed on the basis of this approach would be

$$L = \exp\left[\frac{-3m}{2} \cdot \frac{\frac{1}{\sigma^2_{O_1}}\sum_{j=1}^{\ell}(O_{1,j} - M_{1,j})^2 + \frac{1}{\sigma^2_{O_2}}\sum_{j=1}^{\ell}(O_{2,j} - M_{2,j})^2 + \frac{1}{\sigma^2_{O_3}}\sum_{j=1}^{\ell}(O_{3,j} - M_{3,j})^2}{\frac{1}{\sigma^2_{O_1}}\sum_{j=1}^{\ell}(O_{1,j} - M^{\text{MLE}}_{1,j})^2 + \frac{1}{\sigma^2_{O_2}}\sum_{j=1}^{\ell}(O_{2,j} - M^{\text{MLE}}_{2,j})^2 + \frac{1}{\sigma^2_{O_3}}\sum_{j=1}^{\ell}(O_{3,j} - M^{\text{MLE}}_{3,j})^2}\right] \tag{16}$$

where the indices 1, 2, and 3 are the different system outputs; and $\sigma^2_{O_1}$, $\sigma^2_{O_2}$, and $\sigma^2_{O_3}$ are the variances of the observations for each output. The MLE values in this likelihood function are calculated in the manner described previously.

Figs. 8(c) and 8(d) show the posterior cumulative likelihood distributions for critical shear stress and hiding factor when likelihoods are calculated using Eqs. (5) and (16). In both cases, $m = 1$. The likelihood function in Eq. (16) assigns nearly all of

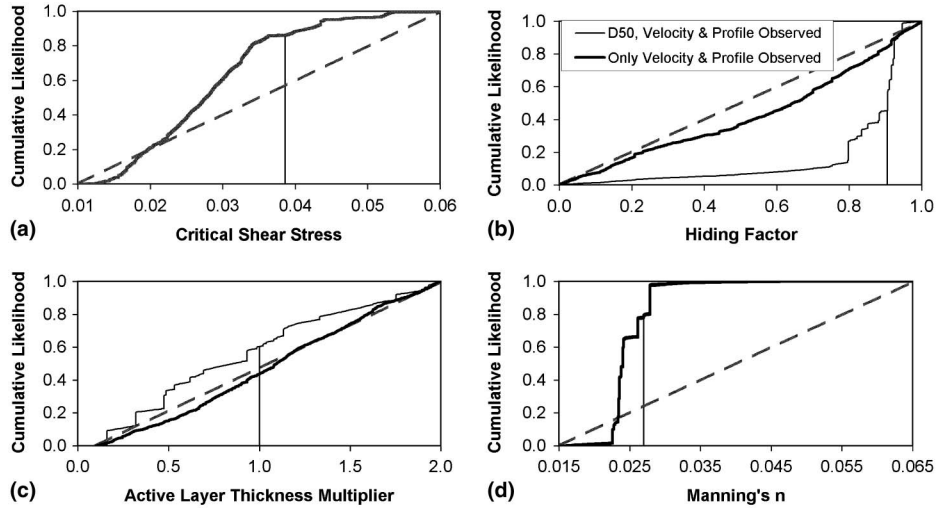

Fig. 9. Comparison of posterior cumulative likelihood distributions for: (a) critical shear stress; (b) hiding factor; (c) active layer thickness multiplier; (d) Manning's n when d50 is observed or unobserved. Dashed lines indicate the uniform distribution that was assumed for each parameter before the model was applied to the calibration period; vertical lines indicate the parameter values used in the calibrated model. All results are for the Ashida and Michiue (1971) experiment

the likelihood to a single parameter value, creating a stair-step cumulative distribution. The parameter set selected by this function is the MLE parameter set. Part of the reason for this result is that the observations used in this analysis are actually model results, so the MLE parameter set is capable of reproducing the results with very little error. Thus, it is judged to have a very high likelihood. However, the parameter values associated with the MLE are not necessarily the true values used to generate the "observations." In particular, notice that the critical shear stress that is identified in Fig. 8(c) is not the true value of the critical shear stress. Overall, these results demonstrate that the form of the likelihood function can have major impact on the results of a GLUE analysis because it contains hidden assumptions about the measurement error and the importance assigned to exactly matching the observations.

To test the impact of the available observations on the results, it is assumed now that no observations were available for d50. In such a case, the likelihood function includes only two outputs rather than three. Fig. 9 compares the posterior cumulative likelihood distributions for critical shear stress, hiding factor, active layer thickness multiplier, and Manning's n, developed using observations for d50, velocity, and bed profile, and developed using only velocity and bed profile. The posterior distributions for critical shear stress and Manning's n [Figs. 9(a) and 9(d)] do not change noticeably. In this model, d50 is not sensitive to critical shear stress or Manning's n, so observations of d50 have little impact on the likelihood distributions for these parameters. The posterior distribution for hiding factor [Fig. 9(b)], however, shows a dramatic change when d50 observations are unavailable, moving closer to a uniform distribution. Because hiding factor is assumed to be uniformly distributed in advance of simulating the calibration period, this implies that the observations from the calibration period are not effective at constraining the value of this parameter. Similarly, the posterior distribution for active layer thickness multiplier [Fig. 9(c)] moves closer to a uniform distribution, implying that the velocity and bed profile observations are of limited effectiveness in constraining this parameter.

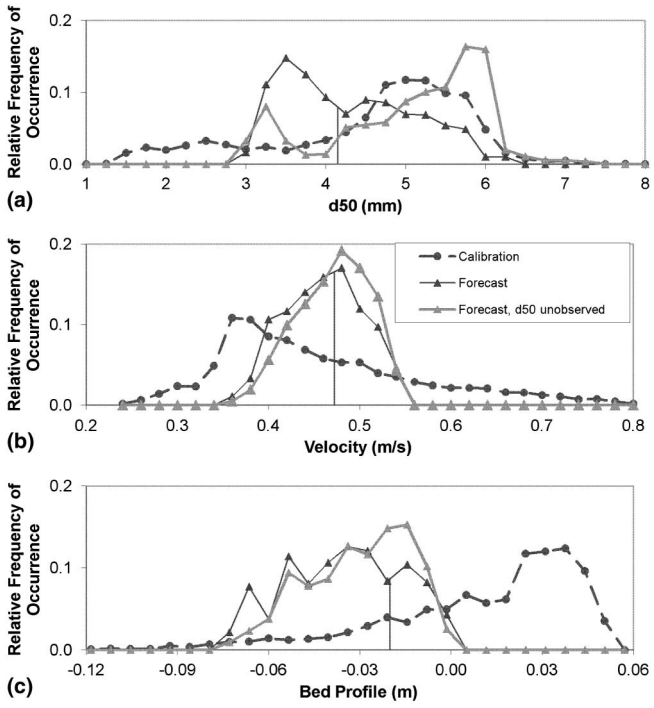

Fig. 10. Histograms of length-averaged (a) d50; (b) velocity; (c) bed profile for the calibration and forecast periods for the Ashida and Michiue (1971) experiment, when d50 is either observed or unobserved during the calibration period. Vertical lines indicate the true values of the output for the forecast period from the calibrated model

The resulting histograms for the simulated d50, velocity, and bed profile outputs for the calibration and forecast periods are shown in Fig. 10. When only velocity and bed profile are observed in the calibration period, the histogram for d50 in the forecast period resembles the histogram from the calibration period, with a most likely value of 5.2 mm. This similarity occurs because hiding factor, which is most important in controlling d50 (see Fig. 3), was poorly constrained by the calibration observations. Active layer thickness multiplier, also known to impact the d50 output, is similarly unconstrained. The histograms of velocity and bed profile after the forecast period are similar irrespective of whether d50 was observed. Because velocity and bed profile observations were available for the calibration period, the most important parameters that impact velocity (Manning's n) and bed profile (critical shear stress) were about equally constrained irrespective of whether d50 was observed. This analysis suggests that it is beneficial during the calibration period to observe any output variable for which forecasts will be required. Such observations help constrain the parameters that impact the same output variable. In the circumstance where direct observations of the desired output are not possible, the GSA provides a tool to determine whether other more measurable outputs depend on the same parameters.

Conclusions

In this paper, a new method was developed to assess the degree to which parameter values are constrained by calibration data and the impact of remaining parameter uncertainty on sediment transport model forecasts. The method begins by assuming the parameters are uniformly distributed within specified bounds and then updates these distributions by comparing the results of simulations run on the basis of these parameter values against observations for a calibration period. The distributions are then updated using a likelihood function that extends the one proposed by Stedinger et al. (2008) to include multiple output variables. In the likelihood function, the output variables are weighted using first-order global sensitivities, which are calculated using FAST. The updated distributions of the parameters are then sampled using LHS to produce histograms of model outputs for the forecast period. The main conclusions from the application of this method are as follows:

1. The sensitivities of length-averaged median grain size, flow velocity, and bed profile to the model parameters can be quite different for erosion and deposition cases. In the erosional experiment by Ashida and Michiue (1971), median grain size is most dependent on hiding factor, velocity is most dependent on Manning's n, and bed profile is most dependent on critical shear stress. For the depositional experiment by Seal et al. (1997), median grain size is most dependent on hiding factor and weight of bed-load fractions, velocity is most dependent on Manning's n, and bed profile is most dependent on critical shear stress and Manning's n. Also, the outputs for the depositional case tend to be sensitive to more parameters than the outputs for the erosional case. For example, for the erosion case considered in this paper, median grain size is most sensitive to hiding factor and relatively insensitive to the other parameters. For the depositional case considered, median grain size is sensitive to critical shear stress, hiding factor, weight of bed-load fractions, and Manning's n.
2. The analysis of global sensitivities suggests the importance of calibrating against observations of variables that will be

included in the forecast. For example, if the forecast includes median grain size, then the model should be calibrated using observations of median grain size. This approach assists the calibration method in constraining the parameters most important to the variables included in the forecast. If the variables included in the forecast cannot be observed directly during the calibration period, then the global sensitivities can be used to identify alternate output variables that depend strongly on the same parameters. These alternate variables could then be used to constrain these parameters, reducing the uncertainty in the forecast.

3. On the basis of the evaluation of the impact of parameter uncertainty presented, weighting the different output variables on the basis of the first-order sensitivity in the likelihood function appears to be an adequate substitute for use of the total-order sensitivity. This approximation is potentially beneficial because faster methods are available to estimate the first-order sensitivity than the total-order sensitivity. Further testing is needed to identify model structures and applications where the total-order sensitivity might produce substantially different results.
4. By using two mathematical forms of the likelihood function, it was observed that the choice of the likelihood function can produce widely differing estimates of the parameter uncertainty remaining after calibration, and thus the uncertainty in the model forecasts attributable to parameter uncertainty. Similarly, the choice of the variable m, the effective number of independent observations, has a significant impact on the results. These issues are related to implicit assumptions about measurement and other errors in the analysis. Further research is needed to determine the applicability of the likelihood function. It is recommended that this research begins by applying the methodology to cases where the error and model structures are simple and the likelihood function is known from basic statistics and then transitions toward more complex but interesting sediment transport cases.

Overall, the research described in this paper should be expanded to consider other cases to establish the generality of the results. Additional cases might include more flume-scale experiments, such as deposition in wide and sandy channels (Toro-Escobar et al. 2000) and erosion in alluvial channels. They should also include river-scale models, where a sufficient set of field observations exists. Testing could also consider additional output variables such as channel width, flow depth, d16, d84, and sediment load. Other sediment transport equations such as Meyer-Peter Muller (1948), Laursen (1958), and Ackers-White (1973) could be examined to see how the relationships between model parameters and outputs change.

Acknowledgments

The authors thank the U.S. Bureau of Reclamation for their financial support of this project and two anonymous reviewers for their help in improving this paper.

References

Ackers, P., and White, W. R. (1973). "Sediment transport: New approach and analysis." *J. Hydraul. Div.*, 99(11), 2041–2060.

Arabi, M., Govindaraju, R. S., Engel, B., and Hantush, M. (2007). "Multiobjective sensitivity analysis of sediment and nitrogen processes with a watershed model." *Water Resour. Res.*, 43(6), 1–11.

Ashida, K., and Michiue, M. (1971). "An investigation of river bed degradation downstream of a dam." *14th Int. Association for Hydraulic Research Congress*, Paris, 247–256.

Beldring, S., Engeland, K., Roald, L. A., Saelthun, N. R., and Vokso, A. (2003). "Estimation of parameters in a distributed precipitation-runoff model for Norway." *Hydrol. Earth Syst. Sci.*, 7(3), 304–316.

Beven, K. J. (2000). *Rainfall-runoff modelling: The primer*, Wiley, West Sussex, England.

Beven, K., and Binley, A. (1992). "The future of distributed models: Model calibration and uncertainty prediction." *Hydrol. Processes*, 6(3), 279–298.

Blasone, R.-S., Madsen, H., and Rosbjerg, D. (2008). "Uncertainty assessment of integrated distributed hydrological models using GLUE with Markov chain Monte Carlo sampling." *J. Hydrol. (Amsterdam)*, 353(1–2), 18–32.

Buffington, J. M., and Montgomery, D. R. (1997). "A systematic analysis of eight decades of incipient motion studies, with special reference to gravel-bedded rivers." *Water Resour. Res.*, 33(8), 1993–2029.

Chahinian, N., and Moussa, R. (2007). "Comparison of different multiobjective calibration criteria of a conceptual rainfall-runoff model of flood events." *Hydrol. Earth Syst. Sci. Discuss.*, 4(3), 1031–1067.

Chan, K., Saltelli, A., and Tarantola, S. (1997). "Sensitivity analysis of model output: Variance-based methods make the difference." *Winter Simulation Conf.*, S. Andradottir, K. J. Healy, D. H. Withers, and B. L. Nelson, eds., Association for Computing Machinery, New York, 261–268.

Chang, C.-H., Yang, J.-C., and Tun, Y.-K. (1993). "Sensitivity and uncertainty analysis of a sediment transport model: A global approach." *Stoch. Hydrol. Hydraul.*, 7(4), 299–314.

Christensen, S. (2003). "A synthetic groundwater modelling study of the accuracy of GLUE uncertainty intervals." *Nord. Hydrol.*, 35(1), 45–59.

Clyde, M., and George, E. I. (2004). "Model uncertainty." *Stat. Sci.*, 19(1), 81–94.

Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G., and Schaibly, J. H. (1973). "Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I. Theory." *J. Chem. Phys.*, 59(8), 3873–3878.

Daebel, H., and Gujer, W. (2005). "Uncertainty in predicting riverbed erosion caused by urban stormwater discharge." *Water Sci. Technol.: Water Supply*, 52(5), 77–85.

Davies, A. G., van Rijn, L. C., Damgaard, J. S., van de Graaff, J., and Ribberink, J. S. (2002). "Intercomparison of research and practical sand transport models." *Coastal Eng.*, 46(1), 1–23.

Eidsvik, K. J. (2004). "Some contributions to the uncertainty of sediment transport predictions." *Cont. Shelf Res.*, 24(6), 739–754.

Engeland, K., Braud, I., Gottschalk, L., and Leblois, E. (2006). "Multiobjective regional modelling." *J. Hydrol. (Amsterdam)*, 327(3–4), 337–351.

Engeland, K., and Gottschalk, L. (2002). "Bayesian estimation of parameters in a regional hydrological model." *Hydrol. Earth Syst. Sci.*, 6(5), 883–898.

Freer, J., Beven, K., and Ambroise, B. (1996). "Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach." *Water Resour. Res.*, 32(7), 2161–2173.

Gatelli, D., Kucherenko, S., Ratto, M., and Tarantola, S. (2009). "Calculating first-order sensitivity measures: A benchmark of some recent methodologies." *Reliab. Eng. Syst. Saf.*, 94(7), 1212–1219.

Gourley, J. J., and Vieux, B. E. (2006). "A method for identifying sources of model uncertainty in rainfall-runoff simulations." *J. Hydrol. (Amsterdam)*, 327(1–2), 68–80.

Greimann, B., Lai, L., and Huang, J. (2008). "Two-dimensional total sediment load model equations." *J. Hydraul. Eng.*, 134(8), 1142–1146.

Hall, J. W., Tarantola, S., Bates, P. D., and Horritt, M. S. (2005). "Distributed sensitivity analysis of flood inundation model calibration." *J. Hydraul. Eng.*, 131(2), 117–126.

Harmel, R. D., and King, K. W. (2005). "Uncertainty in measured sediment and nutrient flux in runoff from small agricultural watersheds." *Trans. ASAE*, 48(5), 1713–1721.

Hassan, A. E., Bekhit, H. M., and Chapman, J. B. (2008). "Uncertainty assessment of a stochastic groundwater flow model using GLUE analysis." *J. Hydrol. (Amsterdam)*, 362(1-2), 89–109.

Huang, J. V., and Greimann, B. (2007). *User's Manual for SRH-1D V2.1*, Bureau of Reclamation, U.S. Department of the Interior, Denver.
Jepsen, R. A. (2006). "Uncertainty in experimental techniques for measuring sediment erodibility." *Integr. Environ. Assess. Manage.*, 2(1), 39–43.
Kuczera, G., Kavetski, D., Franks, S., and Thyer, M. (2006). "Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters." *J. Hydrol. (Amsterdam)*, 331(1-2), 161–177.
Laursen, E. M. (1958). "The total sediment load of streams." *J. Hydraul. Div.*, 84(1), 1531–1536.
Legates, D. R., and McCabe, G. J. Jr. (1999). "Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation." *Water Resour. Res.*, 35(1), 233–241.
Mantovan, P., and Todini, E. (2006). "Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology." *J. Hydrol. (Amsterdam)*, 330(1–2), 368–381.
McKay, M. D., Beckman, R. J., and Conover, W. J. (1979). "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code." *Technometrics*, 21(2), 239–245.
Meyer-Peter, E., and Müller, R. (1948). "Formula for bed-load transport." *Proc. Int. Association for Hydraulic Research, 2nd Meeting*, Stockholm, Sweden.
Mo, X., and Beven, K. (2004). "Multi-objective parameter conditioning of a three-source wheat canopy model." *Agric. For. Meteorol.*, 122(1–2), 39–63.
Murray, A. B. (2007). "Reducing model complexity for explanation and prediction." *Geomorphology*, 90(3–4), 178–191.
Nash, J. E., and Sutcliffe, J. V. (1970). "River flow forecasting through conceptual models. Part I—A discussion of principles." *J. Hydrol. (Amsterdam)*, 10(3), 282–290.
Osidele, O. O., Zeng, W., and Beck, M. B. (2003). "Coping with uncertainty: A case study in sediment transport and nutrient load analysis." *J. Water Resour. Plng. and Mgmt.*, 129(4), 345–355.
Page, T., Whyatt, J. D., Beven, K. J., and Metcalfe, S. E. (2004). "Uncertainty in modelled estimates of acid deposition across Wales: A GLUE approach." *Atmos. Environ.*, 38(14), 2079–2090.
Parker, G. (1990). "Surface based bedload transport relationship for gravel rivers." *J. Hydraul. Res.*, 28(4), 417–436.
Pinto, L., Fortunato, A. B., and Freire, P. (2006). "Sensitivity analysis of non-cohesive sediment transport formulae." *Cont. Shelf Res.*, 26(15), 1826–1839.
Ratto, M., Tarantola, S., and Saltelli, A. (2001). "Sensitivity analysis in model calibration: GSA-GLUE approach." *Comput. Phys. Commun.*, 136(3), 212–224.
Refsgaard, J. C., van der Sluijs, J. P., Brown, J., and van der Keur, P. (2006). "A framework for dealing with uncertainty due to model structure error." *Adv. Water Resour.*, 29(11), 1586–1597.
Saltelli, A., et al. (2008). *Global sensitivity analysis: The primer*, Wiley, West Sussex, England.
Saltelli, A., and Bolado, R. (1998). "An alternative way to compute Fourier amplitude sensitivity test (FAST)." *Comput. Stat. Data Anal.*, 26(4), 445–460.
Saltelli, A., Tarantola, S., and Chan, K. P.-S. (1999). "A quantitative model-independent method for global sensitivity analysis of model output." *Technometrics*, 41(1), 39–56.
Schaibly, J. H., and Shuler, K. E. (1973). "Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. II. Applications." *J. Chem. Phys.*, 59(8), 3879–3888.
Seal, R., Paola, C., Parker, G., Southard, J. B., and Wilcock, P. R. (1997). "Experiments on Downstream fining of gravel: I. Narrow-channel runs." *J. Hydraul. Eng.*, 123(10), 874–884.
Shirmohammadi, A., et al. (2006). "Uncertainty in TMDL models." *Trans. ASABE*, 49(4), 1033–1049.
Simons, R. K., Canali, G. E., Anderson-Newton, G. T., and Cotton, G. K. (2000). "Sediment transport modeling: Calibration, verification, and evaluation." *Soil Sediment Contam.*, 9(3), 261–289.
Sorooshian, S., and Dracup, J. A. (1980). "Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Corrolated and heteroscedastic error cases." *Water Resour. Res.*, 16(2), 430–442.
Stedinger, J. R., Vogel, R. M., Lee, S. U., and Batchelder, R. (2008). "Appraisal of the generalized likelihood uncertainty estimation (GLUE) method." *Water Resour. Res.*, 44, W00B06.
Tarantola, S., Gatelli, D., and Mara, T. A. (2006). "Random balance designs for the estimation of first order global sensitivity indices." *Reliab. Eng. Syst. Saf.*, 91(6), 717–727.
Toro-Escobar, C. M., Paola, C., Parker, G., Wilcock, P. R., and Southard, J. B. (2000). "Experiments on downstream fining of gravel. II: Wide and sandy runs." *J. Hydraul. Eng.*, 126(3), 198–208.
Uhlenbrook, S., and Sieber, A. (2005). "On the value of experimental data to reduce the prediction uncertainty of a process-oriented catchment model." *Environ. Model. Software*, 20(1), 19–32.
Yang, T.-C., Yu, P.-S., Kuo, C.-M., and Wang, Y.-C. (2004). "Application of fuzzy multiobjective function on storm-event rainfall-runoff model calibration." *J. Hydrol. Eng.*, 9(5), 440–445.
Yapo, P. O., Gupta, H. V., and Sorooshian, S. (1998). "Multi-objective global optimization for hydrologic models." *J. Hydrol. (Amsterdam)*, 204(1–4), 83–97.
Ziegler, C. K. (2006). "Using mathematical models to assess sediment stability." *Integr. Environ. Assess. Manage.*, 2(1), 44–50.