THESIS

A METHOD FOR ASSESSING IMPACTS OF PARAMETER UNCERTAINTY IN SEDIMENT TRANSPORT MODELING APPLICATIONS

Submitted by

Morgan D. Ruark

Department of Civil and Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2009

COLORADO STATE UNIVERSITY

July 13, 2009

WE HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER OUR SUPERVISION BY MORGAN D. RUARK ENTITLED A METHOD FOR ASSESSING IMPACTS OF PARAMETER UNCERTAINTY IN SEDIMENT TRANSPORT MODELING APPLICATIONS BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE.

Committee on Graduate Work

Stephanie Kampf, PhD

Blair Greimann, PhD

Advisor: Jeffrey D. Niemann, PhD

Department Head: Luis Garcia, PhD

ABSTRACT OF THESIS

A METHOD FOR ASSESSING IMPACTS OF PARAMETER UNCERTAINTY IN SEDIMENT TRANSPORT MODELING APPLICATIONS

Numerical sediment transport models are widely used to evaluate impacts of water management activities on endangered species, identify appropriate strategies for dam removal, and other projects. The SRH-1D (Sedimentation and River Hydraulics - One Dimension) numerical model, formerly known as GSTARS, is used by the U.S. Bureau of Reclamation for many such evaluations. The predictions from models such as SRH-1D include uncertainty due to errors in the model's mathematical structure, uncertainty in parameter values, and other sources. Quantifying this uncertainty and its origins could provide guidance for more efficient data collection and model calibration and could ultimately reduce project design requirements. In this research, we seek to evaluate impacts that parameter uncertainty has on the uncertainty in model forecasts. This assessment is made using a new multi-objective version of Generalized Likelihood Uncertainty Estimation (GLUE). In this approach, the likelihood of parameter values is assessed using a function that weights different output variables using their first order global sensitivities obtained from Fourier Amplitude Sensitivity Test (FAST). The method is applied to SRH-1D models of two flume experiments: an erosional case described by Ashida and Michiue (1971) and a depositional case described by Seal et al. (1997). Eight parameters (critical shear stress, hiding factor, active layer thickness multiplier, recovery factor for deposition, recovery factor for scour, bedload adaptation length, weight of bedload fractions, and Manning's roughness) are initially considered uncertain in the analysis, and the sensitivities and uncertainties of output variables describing median grain size, flow velocity, and bed profile elevation are analyzed. Overall, the results suggest that the sensitivities of the model outputs can be rather different for erosional and depositional cases and that the outputs in the depositional case are typically 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.

Morgan D. Ruark
Civil and Environmental Engineering Department
Colorado State University
Fort Collins, CO 80523
Fall 2009

ACKNOWLEDGEMENTS

The financial support of the U.S. Bureau of Reclamation is gratefully acknowledged. Dr. Jeff Niemann offered guidance and insight throughout this research. Dr. Blair Greimann offered assistance with sediment transport modeling, and application of the SRH-1D model. Dr. Stephanie Kampf was kind enough to step in as a committee member at short notice. Dr. Mazdak Arabi assisted with questions about sensitivity analysis and uncertainty estimation.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1 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 identify sediment equilibrium conditions (Huang, Greimann and Yang, 2003), assess historical conditions to determine possible impacts of watershed changes (Holmquist-Johnson, 2004), evaluate water supply management (Greimann et al., 2006), manage reservoirs (Greimann and Huang, 2006), and predict impacts of proposed water resource systems on endangered species (Holmquist-Johnson, 2004).

Predictions from sediment transport models always entail uncertainty. Sources of uncertainty include: (a) errors or simplifications in the mathematical structure of the model and the model's representation of processes, (b) errors in the initial and boundary conditions, (c) errors in the observations used to calibrate the model parameters, (d) errors in the values of model parameters, and (e) errors in model inputs or 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 amount of material eroded or deposited at critical locations (Simons et al., 2000; Davies et al., 2002; Eidsvik, 2003). Past research has focused on uncertainty arising from sediment transport models or formulae (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 and Yang, 1993) and the implications of parameter uncertainty.

Currently, one of the methods used to assess parameter uncertainty in sediment transport models of river systems to simply identify a worst-case scenario and develop a numerical model for this condition. While this method has the advantage of requiring minimal model simulation, no formal criteria are available to determine the worst-case scenario and no likelihood of occurrence is associated with this scenario. When applied to the design of proposed projects, an overly conservative evaluation could result in over-designed and needlessly expensive structures. Similarly, excessively weak evaluation could result in the failure of projects to meet their objectives and possibly even public hazards.

Bayesian methods offer a formal method 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 a calibration period to generate a posterior joint distribution for the parameters (Beven, 2000). The updating of the joint distribution is based on 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 a well-defined theoretical foundation 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. Changing variance, or heteroscedasticity, can be seen in hydrologic modeling applications, such as discharge hydrographs (Sorooshian and Dracup, 1980). This suggests that some sediment transport modeling applications, especially those involving streamflow, may violate the homoscedastic assumption.

The generalized likelihood uncertainty estimation (GLUE) method 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 and Beven, 1996; Campling et al. 2002; Blasone et al., 2007), groundwater models (Christensen, 2003; Hassan et al., 2008), water quality models (Shirmohammadi et al., 2006), and atmospheric models (Page et al., 2004), but to our knowledge, it has not been applied to sediment transport models. GLUE follows the Bayesian approach, but it utilizes an informal function to estimate the likelihood of parameter values given a set of observations. The benefit of the informal likelihood function is that it can be selected based on the model purpose (Mantovan and Todini, 2006), and different likelihood functions are known to produce different uncertainty estimates (Freer and Beven, 1996; Beven, 2000). However, a series of papers (Christensen, 2004; 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 multiple objectives or outputs such as sediment size, sediment load, stream velocity, channel geometry, and bed profile. Available methods of computing multi-objective 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 weighted addition of likelihoods (Beven, 2000). Such approaches have been addressed elsewhere (Yapo et al., 1998; Mo and Beven, 2004; Chanhinian and Moussa, 2007). For this series of experiments, a weighted likelihood function has been selected. The use of such a weighted likelihood function requires selection of individual weights for each function. In the field of multi-objective optimization, such weights are often set ad hoc, but in uncertainty evaluations, a more rigorous (or at least standardized) approach would be beneficial.

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 and specifically to consider the issues of the likelihood function and the weighting of multiple modeling objectives. The likelihood function used in this paper is based on the one described by Christensen (2004) and Stedinger et al. (2008). Additionally, global sensitivity analysis (GSA) is employed as a way to consider 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 SRH-1D models of two physical experiments (an erosional case and a depositional case); both to identify the parameters that lead to the most uncertainty in the model forecasts, and to partially explore the implications of various assumptions in the GSA-GLUE methodology.

The outline of the paper is as follows: (1) the next section, Methodology, details how the GSA and GLUE-based methods are combined to assess the implications of parameter uncertainty, (2) then, the Sedimentation and River Hydraulics – One Dimension (SRH-1D) model is described in the following section, (3) next, the Experiments section summarizes the physical experiments used to test the method, (4) then the Results section discusses the main results of the GSA-GLUE approach, (5) next, the Analysis section evaluates key assumptions of the method, and (6) finally, the paper closes by summarizing the conclusions and future directions for research.

2 Methodology

In this analysis, we assume that the model will be used to simulate two periods of time (or, equivalently, two scenarios), which we call the calibration and forecast periods. Traditionally, the calibration period is used to determine single values for each of the model parameters, which allow the model to efficiently reproduce observed system responses (i.e. model outputs) for that period. Here, the calibration period will be used to determine distributions of parameter values. The forecast period is the unobserved scenario or period where the model is used to forecast the system behavior. In our analysis, the distributions of parameter values will be used to obtain distributions of the model responses for that period.

In practical terms, 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. Based on 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 places specific constraints on the generation of the parameters sets, although they are still approximately uniformly distributed. This sampling procedure allows calculation of the sensitivities using a Fourier analysis of the model responses (see below). Note that parameters that have little effect on any model output can be fixed at this point and excluded from further consideration. The second step is the application of the GLUE methodology to calculate the likelihoods associated with each parameter set, and from those likelihoods to determine updated likelihood distributions for each parameter. Likelihoods are calculated based on the model's ability to reproduce the observed system response when given parameter sets are used. Because sediment transport models typically produce multiple model outputs of interest (e.g., sediment size, channel profile, etc.), the sensitivities calculated in the first step are used to weight different outputs in the calculation of 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 likelihood distributions of the parameters to generate a new sample of parameter sets. The model is run for the forecast period using these parameters sets, and quantiles are calculated for the model outputs. These quantiles allow an assessment of the implications of parameter uncertainty on the forecasts of the model. The following subsections below describe each of the three steps in greater detail.

2.1 GSA

Sensitivity analysis usually aims to quantify how much the outputs of a model change when a model parameter (or input) is varied (Saltelli et al., 2008). While a local sensitivity analysis evaluates these changes around 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 (Saltelli et al., 2008). In contrast, GSA most simply uses standardized regression coefficients as measures of global sensitivity; however, this approach assumes linearity of each output with respect to each parameter (Saltelli et al., 2008). A superior method is to use a variance-based measure of sensitivity, which partially overcomes the linearity assumption (Chan et al., 1997). Here, we consider two variance-based measures of sensitivity. One is the first order index S_x, which is defined as:

$$S_x = \frac{\operatorname{var}\left[E\left(Y \middle| X\right)\right]}{\operatorname{var}(Y)} \tag{1}$$

where $\operatorname{var}(Y)$ is the total variance of the model output Y when all the parameters are varied within their specified ranges, $E\left(Y \middle| X\right)$ is the expected value of output Y for a particular value of parameter X, and $\operatorname{var}\left[E\left(Y \middle| X\right)\right]$ is the variance of $E\left(Y \middle| X\right)$ when X is varied over its allowed range. The second measure is the total order index S_{Tx}, which can be written as:

$$S_{T_x} = 1 - \frac{\operatorname{var}\left[E\left(Y \middle| \widetilde{X}\right)\right]}{\operatorname{var}(Y)} \tag{2}$$

where $\operatorname{var}\left[E\left(Y \middle| \widetilde{X}\right)\right]$ is the 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 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. Further details about these sensitivity measures and their properties can be in Saltelli et al. (1999).

FAST offers an efficient way to estimate these variance-based measures of sensitivity. FAST was initially developed to study first order effects in coupled reaction systems in chemical models (Cukier et al., 1973) and was later expanded to include the total order effects (Saltelli et al., 1999). In FAST, the efficiency is achieved by varying all parameters of interest simultaneously rather than perturbing the parameters one-by-one. The parameters are varied at non-interfering frequencies (Cukier et al., 1973; Shaibly and Shuler, 1973) within the ranges that are specified by the modeler. The generated sequence of parameter sets is then used in the model to generate an associated sequence of model responses. The model response sequence is 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, the total number of simulations performed, must be specified. The estimates of the sensitivity indices from FAST asymptotically converge to the definitions given in Equations (1) and (2) as the sample size becomes large.

In the present analysis, FAST is used to calculate the importance of each parameter to each model output. Use of FAST also allows screening of parameters, in order to remove those with little influence on model outputs. In particular, if all the first and total order sensitivities of 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 below to weight the performance of the model in reproducing different model outputs.

2.2 GLUE

The GLUE method is next used to determine revised, or posterior, distributions for the parameters. By running the model with each parameter set and comparing its performance to the observed system behavior, we have generated 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 for the model. However, the samples produced by the FAST method are also approximately uniform and can be used in the GLUE method (Ratto et al., 2001).

GLUE evaluates the likelihood of each parameter set based on the model's ability to reproduce observations when that parameter set is used. Many previous papers have used the Nash-Sutcliffe Coefficient of Efficiency or NSCE as the basis of the likelihood function (Beldring et al., 2003; Arabi et al., 2007; Engelund et al., 2006; Engelund and Gottschalk, 2002; Uhlenbrook and Sieber, 2003). NSCE is calculated as:

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

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

Recent papers (Mantovan and Todini, 2006; Stedinger et al. 2008) have shown that arbitrary likelihood functions, such as NSCE, can produce arbitrary results in the GLUE methodology. Stedinger et al. (2008) demonstrated this by applying GLUE, with a likelihood function based on NSCE, to a simple case where the appropriate likelihood function is known for basic statistics. The case they considered is linear regression with assumed 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:

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

where M represents the model's value when a particular parameter set is used, M^{MLE} is the model's value when the parameters are obtained from the Maximum Likelihood Estimator (MLE), and K is a normalization constant that ensures that all the likelihoods sum to one. This likelihood function has some similarities to NSCE, but it includes two key differences. First, the denominator determines the likelihood by comparing the performance of a given parameter set to that of the MLE (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 likelihood (Stedinger et al., 2008). These changes allow for (1) comparison of the individual model to the best parameter set, as identified by the MLE, and (2) weighting of the likelihood function based on the number of independent observations available.

In the present application, a likelihood function is utilized that is similar to the one in Equation (4), with two key differences. The first difference is that the errors are not expected to be independent between observation locations, so 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 using a weighted sum of likelihoods. In the case where three output or response variables are available, the resulting likelihood function is:

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

where the subscripts 1, 2, and 3 distinguish the three response variables and the *w*'s are the individual weights. The weights are calculated using the first order sensitivities from the GSA, but could also be calculated using the total order sensitivities (the impact of this selection is evaluated later). First order weights were selected because they may be estimated using methods which are faster than FAST (Gatelli et al., 2008), such as Monte-Carlo based methods (Homma and Saltelli, 1996) and Random Balance Designs (Tarantola et al., 2006).

Specifically, when calculating the likelihood for a given parameter, the weights are the first order sensitivities of the three observed outputs to that parameter, divided by the sum of the first order sensitivities of the observed output variables. Note that a given parameter set will have a different likelihood for each parameter. To evaluate the likelihood function in Equation (5), the performance of the MLE is required. In general, the MLE is not generated from the GLUE methodology, so it is assumed that the performance of the MLE is the same as the best performing parameter set in the sample. For simplicity, performance is judged by finding the minimum of total error ε where:

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

and

$$\sigma_{O_1}^2 = \frac{1}{l-1}\sum_{j=1}^{l}(O_{1,j} - \overline{O_1})^2 \tag{7}$$

and $$\overline{O_1} = \frac{1}{l}\sum_{j=1}^{l} O_{1,j} \ . \tag{8}$$

The variances and averages for output variables 2 and 3 would be calculated using expressions equivalent to Equations (7) and (8). The normalization constant K is found from the constraint that all the likelihoods for a parameter should sum to one. In practical terms, preliminary likelihoods are calculated by neglecting K in Equation (5). Then, the sum of these likelihoods is calculated, and each preliminary likelihood is divided by the sum to determine the final likelihoods. The cumulative posterior distribution for each parameter can be calculated by summing all the likelihoods associated with values of the parameter.

The limitations of this methodology should be stressed. The likelihood function in Equation (5) is proposed rather than derived from a particular set of statistical assumptions. In addition, the method neglects correlation or dependence between the most likely values of different parameters. Only the marginal posterior distributions for the parameters are produced by this method.

2.3 LHS

The third and final step of the methodology is to use the posterior distributions of the parameters in the model to simulate the forecast period and to determine the associated distributions for the model outputs. LHS is used to sample the marginal posterior distribution of each parameter (Hall et al., 2005; Chang et al. 2005). In contrast to Monte Carlo sampling, which generates random values from the distribution, LHS attempts to explore the parameter space representatively, selecting parameter values at regularly-spaced percentiles. LHS is used because previous research has shown that smaller sample sizes can be used to characterize a distribution for LHS than 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 have little impact on the model results, based on the GSA conducted for the calibration period, 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 earlier. 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 non-uniform, 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 sample of parameter sets is then used in the model for the forecast period to determine the associated distribution of model responses.

3 SRH-1D

The methodology for assessing parameter uncertainty is tested using the SRH-1D model. SRH-1D is an outgrowth of the Generalized Stream Tube model for Alluvial River Simulation (GSTARS) and 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, 2006). The model can simulate steady or unsteady flow and can treat cohesive and non-cohesive sediment. It has been used in the assessment of the impacts of proposed projects (Greimann et al., 2006), evaluation of existing structures including dams and off-take structures (Greimann and Huang, 2006), and evaluation of historical conditions of riverine sediment supply (Huang and Bauer, 2005; Huang et al., 2003)).

The experiments considered in this paper use only steady flow and non-cohesive sediment. SRH-1D uses one-dimensional flow calculations, including the standard step energy method for steady gradually varied flow (Huang and Greimann, 2007). The hydraulic component determines flow depths based on volumetric flows, cross-sectional geometry, Manning's equation, hydraulic gradient, and other energy losses. Between adjacent cross-sections (j and j+1), the energy equation is written:

$$Z_{j+1} + \beta_{j+1}\frac{v_{j+1}^2}{2g} - Z_j - \beta_j\frac{v_j^2}{2g} - h_f - h_c = 0 \qquad (10)$$

where Z represents the water surface elevation, β is a velocity distribution coefficient, v is the average velocity at the cross-section, g is gravitational acceleration, h_f represents friction loss, and h_c represents contraction or expansion losses. Evaluation of the friction loss in Equation (10) ultimately requires use of Manning's equation and specification of Manning's roughness coefficient n.

SRH-1D also 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 the Exner equation routing. Because steady flow is considered here, the Exner equation is used and mass conservation can be written:

$$\frac{\partial Q_s}{\partial x} + \varepsilon \frac{\partial A_d}{\partial t} - q_s = 0 \tag{11}$$

where Q_s is volumetric sediment discharge, ε is volume of sediment per unit bed layer volume (related to porosity), A_d is volume of bed sediment per unit length, and q_s is 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 fractions. Lateral inflows are specified by the user and are zero in the present application. Because the cross-sections may 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 an equalization length. Evaluation of the equalization length requires specification of a bedload adaption length parameter as well as separate deposition and scour recovery factors. The transport capacity expression used here is Parker's sand and gravel equation, which ultimately requires specification of a reference, or critical, shear stress and a hiding factor, which accounts for differences in critical shear stresses for particles of different sizes.

Bed material mixing is modeled by dividing the bed into a thin 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 using the geometric mean of the largest size class and a user-specified proportionality constant. 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 is now classified as the top inactive layer. As part of the bed material mixing, the user must specify a bedload weighting, which controls the importance of bed load in the transfer of material between the active layer and the underlying layer.

In the end, eight parameters are treated as uncertain in this analysis: Manning's n, critical shear stress, hiding factor, deposition recovery factor, scour recovery factor, bedload adaptation length, active layer thickness multiplication factor, and the weighting of bedload fractions for transfer from surface to subsurface. None of these parameters are measurable in the field, and they can vary significantly from case to case. Thus, they are typically specified by the modeler. 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.

Table 1. Selected bounds for the uniform distributions describing the eight parameters.

Parameter	Minimum value	Maximum value
Critical Shear Stress	0.01	0.06
Hiding Factor	0	1
Active Layer Thickness Multiplier	0.1	2
Deposition Recovery Factor	0.05	1
Scour Recovery Factor	0.05	1
Bedload Adaptation Length	0	10
Weight of Bedload Fractions	0	1
Manning's n	0.015	0.065

SRH-1D produces a large number of outputs including: mass balance, sediment load, sediment sizes, bed profile, flow velocity, and sediment concentrations. These outputs are available at multiple locations and times for a given simulation. For our key model response variables, we selected the length-averaged median sediment size, flow velocity, and bed profile. For sediment size, the variable is defined as:

$$\overline{d_{50}} = \frac{\sum_{j=1}^{l} \left(\frac{d_{50_j} + d_{50_{j+1}}}{2} \Delta L_{j,j+1} \right)}{L_{total}} \tag{12}$$

where $\Delta L_{j,\,j+1}$ is the length between cross-sections j and $j + 1$, d_{50_j} is the median grain size at cross-section j, and L_{total} is the total length of the reach (the sum of all ΔL's). Similarly, the length-averaged flow velocity is defined as:

$$\bar{v} = \frac{\sum_{j=1}^{l} \left(\frac{v_j + v_{j+1}}{2} \Delta L_{j,j+1} \right)}{L_{total}} \tag{13}$$

where v_j is the average flow velocity at cross-section j. Finally, the length-averaged bed elevation is defined as:

$$\overline{P} = \frac{\sum_{j=1}^{l} \left(\frac{P_j + P_{j+1}}{2} \Delta L_{j,j+1} \right)}{L_{total}} \tag{14}$$

where P_j is the average bed elevation of the channel at cross-section j.

4 Experiments

The model was applied to two flume experiments. One experiment is an erosional case and the other is depositional. These experiments were chosen in part due to their well-documented conditions. In particular, 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. Another reason that these experiments were selected is that calibrated SRH-1D models were already available for both cases.

The Ashida and Michiue (1971) experiment was designed to simulate river bed 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 authors. In this case, the initial bed slope was 0.01 m/m (1%), a sand-to-gravel particle size distribution was used for the bed material with sizes ranging from 0.2 mm to 10 mm and an initial median diameter of 1.5 mm. A clear-water discharge of 0.0314 m^3/s was applied at the upstream end of the flume for the ten hour 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 meters 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 meters from the downstream end of the flume) at the beginning and end of the experiment. Due to the lack of an extensive set of physical data, output from a calibrated SRH-1D model was used in place of physical observations when evaluating the parameter uncertainty. This approach means that any disagreement between model simulations performed as part of the analysis of parameter uncertainty and the “observed” values is due to errors in the parameter values. This approach also allows us to vary the amount of observations supplied to the method and to determine which observations are most valuable in reducing parameter uncertainty. This opportunity is exploited later in the Analysis section.

Figure 1a shows the comparison of values from the calibrated model, as seen in

Table 2, and the experimental observations. The upstream boundary condition was set to no sediment inflow, and the downstream boundary condition was set to allow sediment outflow. Actual observations were used as the initial conditions for the calibration simulation. Cross section spacing was every 0.5 meters, with 41 total cross sections. Grain sizes were broken into 9 classifications. The model was manually calibrated using both comparisons with the observed bed profile and the observed bed grain size distribution. The calibrated model compares well to these observations through hour 2. After hour 2, the calibrated model appears to be overestimating the erosion that occurs. It is suspected that as erosion happens in this experiment, armoring of the bed occurs, which is not properly captured by the calibrated model.

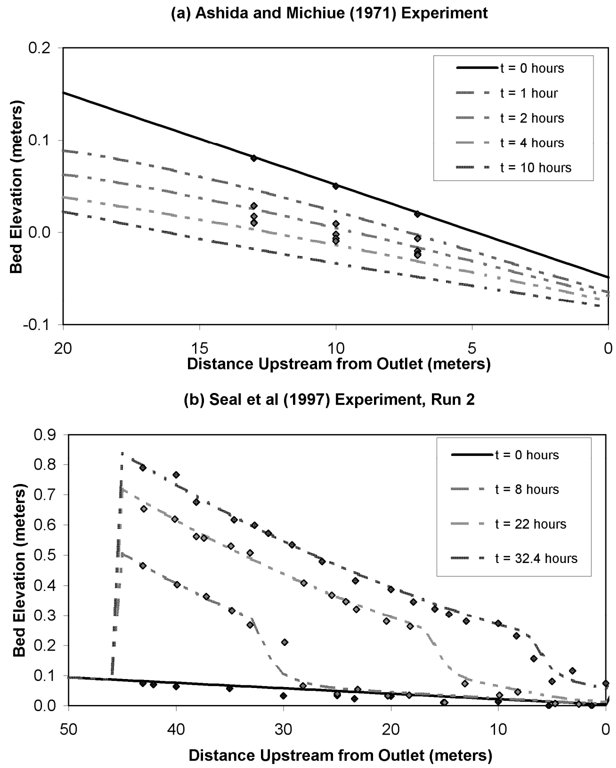

Figure 1. Observed bed elevations (points) and the bed profiles produced by the calibrated models (lines) for the (a) Ashida and Michiue (1971) and (b) Seal et al. (1997) experiments.

Table 2. Parameter values for two calibrated models.

Parameter	Ashida and Michiue (1971)	Seal et al (1997)
Critical Shear Stress	0.0386	0.0386
Hiding Factor	0.905	0.905
Active Layer Thickness Multiplier	1	1
Deposition Recovery Factor	0.25	1
Scour Recovery Factor	1	1
Bedload Adaptation Length	5.0	0.10
Weight of Bedload Fractions	0	0
Manning’s n	0.027	0.022

For the analysis of parameter uncertainty, the Run 6 experiment was divided into a calibration period from 0 to 2 hours and a forecast period from 2 hours to 10 hours.

Results of the calibrated model from 0 to 2 hours are treated as observations to which the simulations using other parameter values are compared. After simulation of the calibration period, the bed elevation and grain size composition of the bed are set to the profile of the calibrated model. The forecast period from 2 to 10 hours is then simulated. For the forecast cases, the observations from the calibrated model were used as the initial conditions for the simulations. The forecast period has identical conditions to the calibration period aside from the initial condition.

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). All three flume experiments were 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 hours, 32.4 hours, and 65 hours. 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 0.19 to 0.05 kg/s. The resulting profile was regularly measured (every half hour, every hour, and every 2 hours for Runs 1, 2, and 3, respectively) at 18 locations for the duration of the experiments. 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.

To be consistent with the erosional experiment, a calibrated SRH-1D model was developed for the runs and used in place of the physical data as "observations." The upstream boundary conditions specified the feed rate for sediment inflow, and the downstream boundary conditions allowed sediment outflow. Actual observations were used as the initial conditions for the calibration simulation. Cross section spacing was every 1 meter, for a total of 56 cross sections (for model

continuity, this total length is greater than the actual length of the flume). Nine grain size classifications were used. This model was calibrated using comparisons with the observed bed profile and the sediment size distributions. For this experiment, Run 2 was used as the calibration case. The duration of Run 2 is 32.4 hours with a sediment feed rate of 0.09 kg/s. Figure 1b shows a comparison of the calibrated model values, as seen in

Table 2, and the experimental observations. The calibrated model compares well to the observations with the largest discrepancies in occurring near the downstream end of the depositional wedge. Run 1 and Run 3 were both used as forecast cases. 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, Run 1 and 3 have sediment feed rates of 0.09 kg/s and 0.05 kg/s and durations of 16.8 and 65 hours, respectively.

5 Results

5.1 GSA

The FAST method, as previously described, was applied to the calibration cases of the two physical experiments. The eight parameters identified earlier were varied, and the three model outputs were evaluated. Both applications of FAST used sample sizes of 5000 simulations (smaller sample sizes are discussed later in this paper).

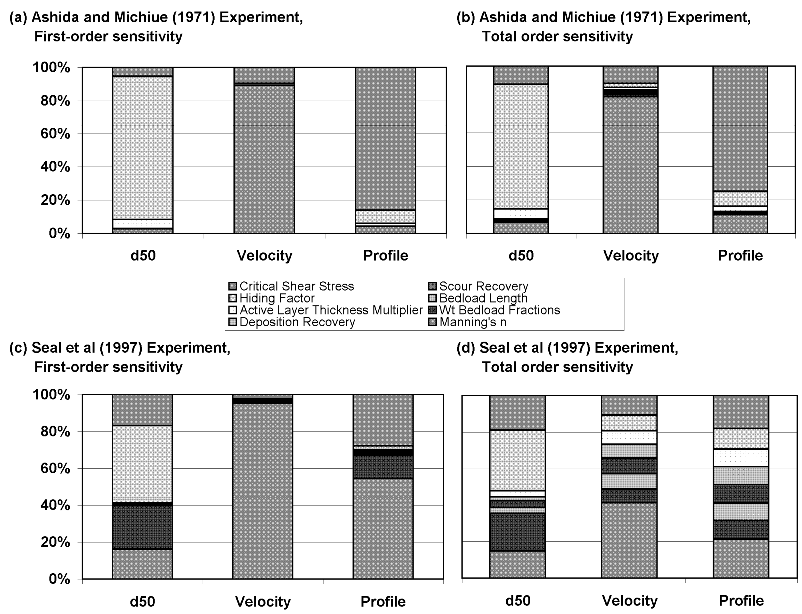

Figure 2. First order and total order sensitivity indices for SRH-1D models of Ashida and Michiue (1971) and Seal et al. (1997) experiments. Sensitivities are stacked in each column in the same order they are listed in the legend.

Figure 2a plots the estimated contribution of each parameter to the total variance of the three output variables for the Ashida and Michiue (1971) experiment based on the first order sensitivity indices. 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. Recall these indices should sum to 1 only if the model is additive with respect to the parameters (Saltelli et al., 2008). Figure 2b shows the results for a similar computation using the total order indices. Notice also that the legend lists the parameters in the order that they are stacked in each column. Both Figure 2a and Figure 2b suggest that four parameters are particularly important in producing variability in the length-averaged median grain size, velocity, and bed profiles. These parameters are the critical shear stress, hiding factor, active layer thickness multiplier, and Manning's n. In fact, only these parameters have total sensitivity values above 5%. For the d50 output, hiding factor is the parameter that produces the greatest sensitivity by far. This result should be expected because hiding factor attempts to account for the difference in the mobility of different size fractions. Thus, it should have a clear impact on sediment size distribution of the bed. For velocity, 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 erodability of the bed material. It is interesting to note that these same relationships hold whether the first order or the total order sensitivity is considered. In general, the contribution of the less important parameters is magnified when the total order indices are considered. This 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. Hence, their contribution to the variance is larger when parameter interactions are included in the analysis.

Figure 2c and Figure 2d show the equivalent results for the Seal et al. (1997) experiment. When considering the first order sensitivities (Figure 2c), four parameters again have contributions larger than 5%: critical shear stress, hiding factor, weight of bedload fractions, and Manning's n. This list is similar to the erosion case, with the weight of bedload fractions replacing the active layer thickness multiplier. For d50, 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 producing the most variance. For velocity, 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 parameter that producing the most variance. The increased importance of Manning's n in determining variability of the bed profile is expected because the flow velocity plays an important role in deposition.

The total order sensitivities in Figure 2d 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 scour recovery factor for the bed profile output. Comparing the total order indices from the erosional and deposition cases (i.e. Figure 2b and Figure 2d) suggest that the depositional process is much more complex than the erosional process. For example, most of the bed profile variance comes from the critical shear stress in the erosional case, but for the depositional case, all parameters have roughly comparable influences on the bed profile.

5.2 GLUE

After the sensitivity analysis was completed for the two experiments, 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 based on the first order sensitivity indices, which are plotted in Figure 2a and Figure 2c. The final weights are tabulated in Table 3. The likelihood function also requires a value for *m*, the effective number of independent locations. Here, *m* is 1, the most conservative value for this variable, since it will produce the largest estimate in parameter value uncertainty. The effect that *m* has on the results is evaluated later.

Table 3. Weights used in the evaluation of the likelihood function in Equation 5 when median grain size, flow velocity, and bed profile are observed.

Parameter	Weight of		
	d50 output	Velocity output	Bed profile output
Ashida and Michiue (1971) experiment, First order sensitivity weights			
Critical shear stress	0.027	0.098	0.875
Hiding factor	0.826	0.022	0.153
Active layer thickness multiplier	0.589	0.025	0.386
Manning's n	0.014	0.942	0.044
Ashida and Michiue (1971) experiment, Total order sensitivity weights			
Critical shear stress	0.061	0.114	0.825
Hiding factor	0.766	0.050	0.184
Active layer thickness multiplier	0.419	0.173	0.408
Manning's n	0.035	0.850	0.115
Seal et al. (1997) experiment, First order sensitivity weights			
Critical shear stress	0.211	0.064	0.725
Hiding factor	0.857	0.043	0.100
Weight of bedload fractions	0.458	0.025	0.517
Manning's n	0.049	0.604	0.347

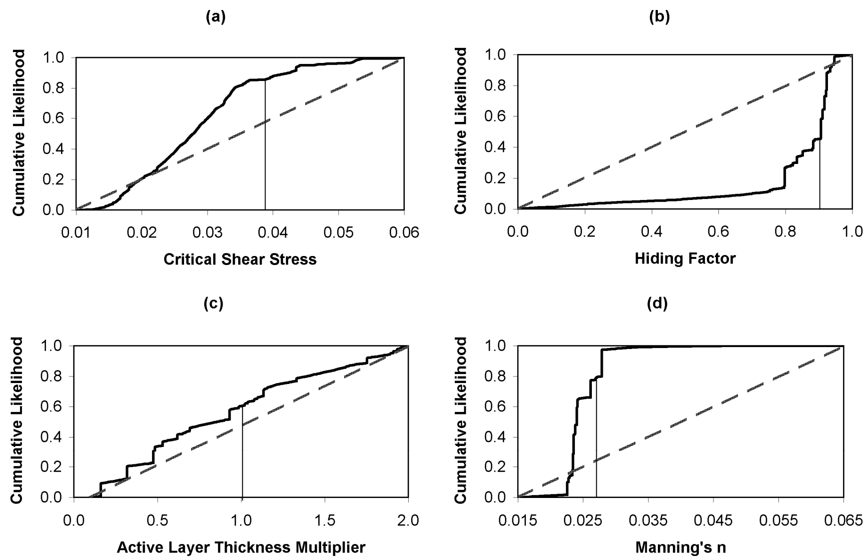

Figure 3. Posterior cumulative likelihood distributions for (a) critical shear stress, (b) hiding factor, (c) active layer thickness multiplier, and (d) Manning's n for the Ashida and Michiue (1971) experiment. Dashed lines indicate the assumed initial uniform distribution for each parameter. Vertical lines indicate the parameter value used in the calibrated model.

The solid lines in Figure 3 show the marginal 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 initial uniform distributions. In the interests of brevity, the posterior distributions of the remaining parameters have not been shown. The steep sections in these 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 demonstrate such a steep section. This result suggests that the active layer thickness parameter is more poorly constrained by the available observations than the other parameters.

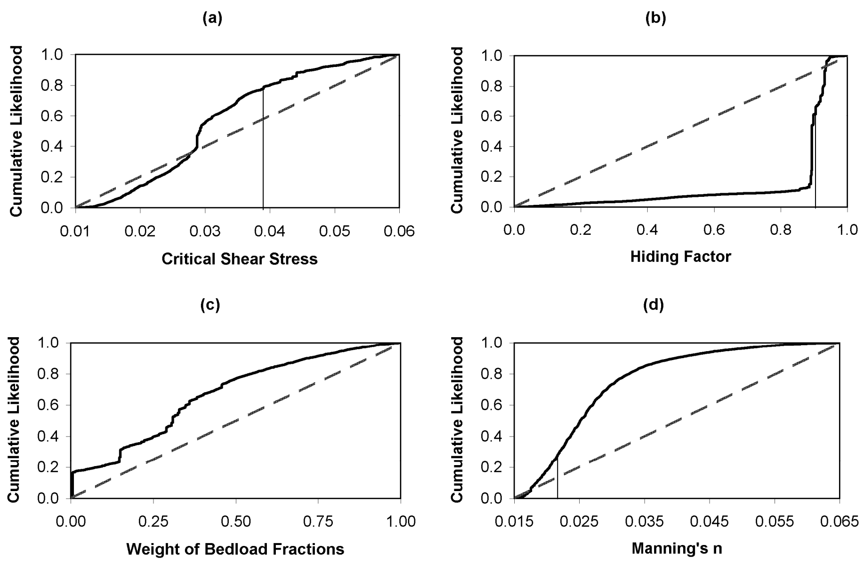

Figure 4. Posterior cumulative likelihood distributions for (a) critical shear stress, (b) hiding factor, (c) weight of bedload fractions, and (d) Manning's n for the Seal et al. (1997) experiment. Dashed lines indicate the assumed initial uniform distribution for each parameter. Vertical lines indicate the parameter value used in the calibrated model. Initial parameter value for weight of bedload fractions is zero.

Error! Reference source not found. shows the posterior cumulative likelihood distributions of critical shear stress, hiding factor, weight of bedload fractions, and Manning's n for the Seal et al. (1997) experiment. Again, these are the parameters found to produce the most variance in the model outputs for this experiment. Steep sections are observed for 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. Critical shear stress and weight of bedload fractions do not exhibit such large steep sections, suggesting that these parameters are more poorly constrained.

5.3 LHS

After the posterior distributions were created, LHS was used to develop samples from them. For both physical experiments, the sample size was 1296 parameter sets (other sample sizes are discussed below). Six values were generated for the four parameters that produced the greatest sensitivity in the outputs, while a single midpoint value was used for the remaining 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.

For the Ashida and Michiue (1971) experiment, six values were generated from each distribution for the posterior distributions of critical shear stress, hiding factor, active layer thickness, and Manning's n. Single values were used for deposition recovery factor, scour recovery factor, bedload adaptation length, and weight of bedload fractions. The first order sensitivity for each of these parameters is less than 1% of the total sensitivity for any of the outputs. Figure 5 plots the histograms for the length-averaged d50, velocity, and bed profile when parameter sets generated from LHS are used to simulate the forecast period (shown as a solid line). For comparison, the figure also shows the histograms of these output variables produced for the calibration period (shown as a dashed line), where the parameters were generated from a uniform distribution via FAST. The vertical axes show relative frequency of occurrence, which was calculated using 30 bins for d50, velocity, and bed profile. The vertical lines represent the observed values for these outputs. The distributions of these outputs have shifted from the calibration to the forecast period due in part to differences in the initial conditions and elapsed simulation time. In particular, the duration of simulation was only 2 hours for the calibration period while it was 8 hours 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 narrow posterior distributions of the parameters used for the forecast period. The narrower distributions of the parameters are likely the reason that the velocity histogram is narrower for the forecast period than for the calibration period. The most common values of d50 range from 3 to 6 mm for the forecast period. The observed value (from the calibrated model) was actually 4.2 mm. The most common values of velocity range from 0.35 to 0.54 m/s for the forecast period, and the observed value was 0.47 m/s. The most common values of the bed profile range from -0.080 to -0.005 m, and the observed value was -0.020 m. Thus, all of the histograms include the actual value for the forecast period. In some cases, the actual values for the forecast period is near the values judged to be most likely from the histograms in the figure.

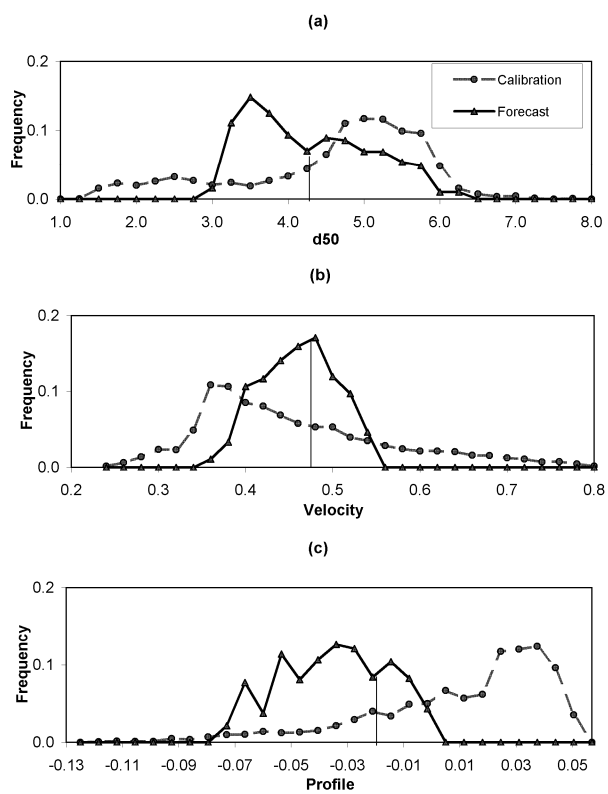

Figure 5. Histograms of (a) d50, (b) velocity, and (c) bed profile for the calibration and forecast periods for the Ashida and Michiue (1971) experiment. Vertical lines indicate the output value from the calibrated model.

For the Seal et al. (1997) experiment, six values were generated from each distribution for the posterior distributions of critical shear stress, hiding factor, weight of bedload fractions, and Manning's n, using the LHS method. Single values were used for the active layer thickness multiplier, deposition recovery factor, scour recovery factor, and bedload adaptation length. Again, as can be seen in Figure 2, the first order sensitivities for these fixed parameters are all less than 1% of the total first order sensitivity for any of the model outputs.

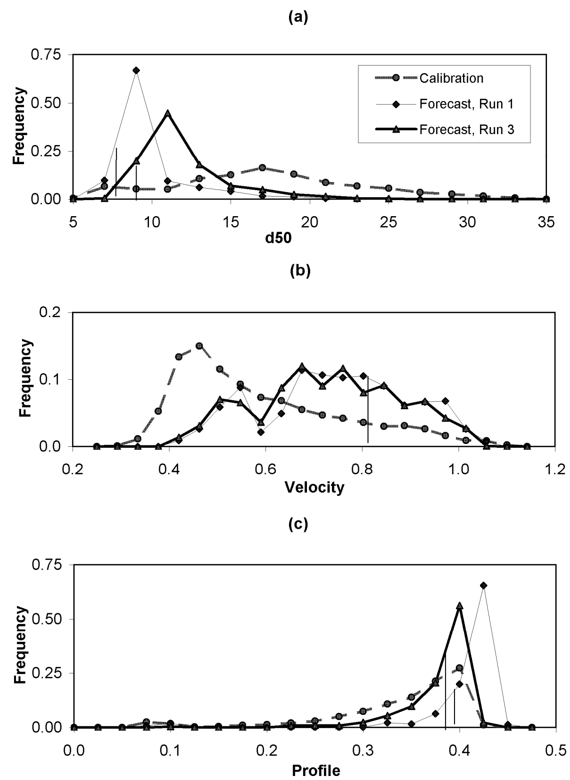

Figure 6. Histograms of (a) d50, (b) velocity, and (c) bed profile for calibration and forecast periods for Seal et al. (1997) experiment. Vertical lines indicate the output value from the calibrated model.

Figure 6 plots the histograms for the length-averaged d50, velocity, and bed profile for the two Seal et al. (1997) cases considered as forecast scenarios (shown as solid lines). These histograms were generated using 23 bins for d50, velocity, and bed profile. The figure also shows the histograms of these output variables that were produced for the calibration period (shown as a dashed line), using parameters generated from a uniform distribution via FAST. The vertical lines represent the observed values for these outputs.

Overall, observed values for the forecast periods typically fall within the range of values included in the histogram, although not always at the most likely value.

6 Analysis

The results described in the previous section are examined based on several decisions made in the application of the methodology. These decisions include: (1) the sample size used in FAST to assess the sensitivities of the outputs to the parameters, (2) the use of the first order sensitivities rather than the total order sensitivities to calculate the weights in the likelihood function, (3) the assumed effective number of independent observations, *m*, (4) the mathematical form of the likelihood function, (5) inclusion of observations of d50, velocity, and bed profile to constrain the parameters, (6) neglecting the possible correlation in the posterior distributions of the parameters, and (7) the sample size used in the LHS method for the forecast period. The impact of each of these decisions is examined below using the Ashida and Michiue (1971) experiment.

The impact of the sample size used in FAST is examined first, as seen in Figure 7. In the previous section, a sample size of 5000 simulations was used (Figure 7a). To test whether this sample size is adequate to quantify the sensitivities, a second sample of equal size was generated and used in FAST to estimate the sensitivities for the Ashida and Michiue (1971) experiment (Figure 7b). As can be seen, the same main parameters were identified in the global sensitivity analysis. Smaller sample sizes of 1160 and 968 simulations (Figure 7c and Figure 7d) were also generated and used to estimate the sensitivities with FAST. Both of these showed good qualitative agreement with the sensitivity analysis of the initial sample. In particular, analysis of both samples identified the same main parameters. A sample size of 520 simulations was also tested. For this simulation size, the sensitivity analysis failed to identify the same parameters previously found to produce the most variance in the outputs. Thus, this sample size is too small to produce reliable estimates of sensitivity, and would greatly impact the assessment of parameter uncertainty, at least for this experiment.

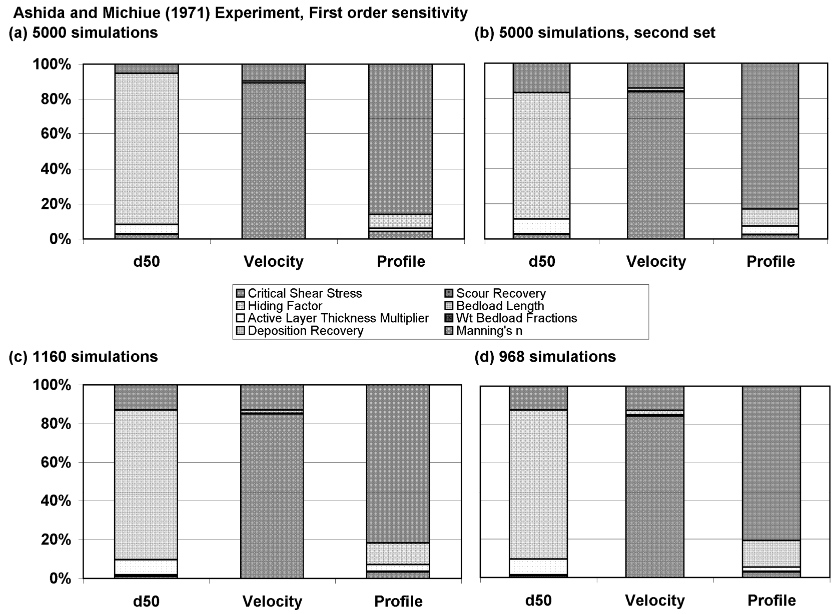

Figure 7. Comparison of first order sensitivity indices for SRH-1D model of Ashida and Michiue (1971) experiment using different sample sizes. Sensitivities are stacked in each column in the same order they are listed in the legend.

Next, the impact of using the first order sensitivities to determine the weights in the likelihood function in Equation (5) is examined. While the total order includes parameter interactions in judging the impact of a given parameter on the outputs, using first order sensitivities potentially allows application of faster methods to estimate the sensitivities (Saltelli et al., 2008; Gatelli et al., 2008). The weights used in this comparison are shown in Table 3. Figure 8 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. As can be seen from Table 3, the weights are similar, so the resulting posterior distributions are very similar in Figure 8. In fact, the cumulative likelihood distributions for critical shear stress are visually indistinguishable. This analysis provides preliminary evidence that the use of first order instead of total order weighting may have relatively little impact on the assessment of parameter uncertainty.

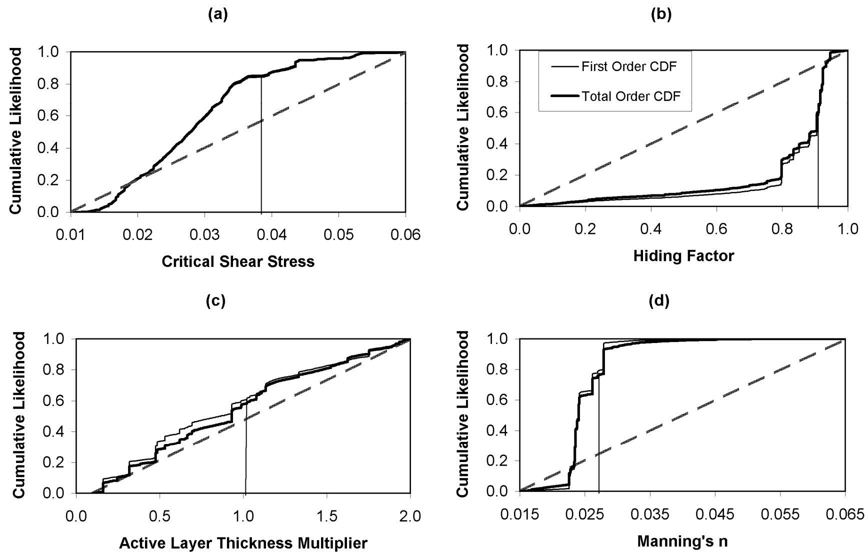

Figure 8. Comparison of posterior cumulative likelihood distributions for (a) critical shear stress, (b) hiding factor, (c) active layer thickness multiplier, and (d) Manning's n when the likelihood function uses the first order or total order sensitivity. Dashed lines indicate the assumed initial uniform distribution for each parameter. Vertical lines indicate the parameter value used in the calibrated model. All results are for the Ashida and Michiue (1971) experiment.

Another key assumption in the methodology above is the effective number of independent observations, *m*. Previously, *m* was assumed to be 1 due to the expected error dependence at different cross-sections in a simulation. Here, 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 Figure 9a and Figure 9b.

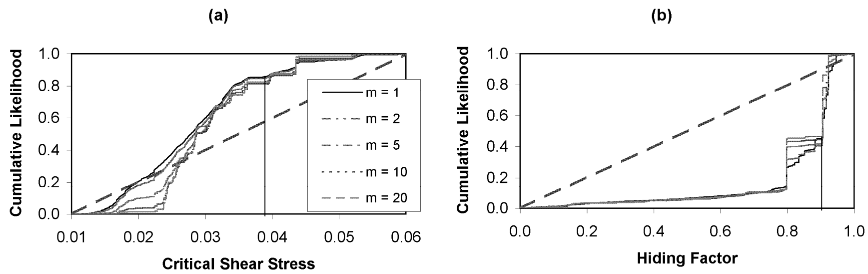

Figure 9. Comparison of the effect of the choice of *m*, the effective number of independent observations, on the posterior cumulative likelihood distributions for (a) critical shear stress and (b) hiding factor. Dashed lines indicate the assumed initial uniform distribution for each parameter. Vertical lines indicate the parameter value used in the calibrated model. All results are for the Ashida and Michiue (1971) experiment.

As can be seen in this figure, an increase in *m* creates a posterior distribution with a more erratic shape, where 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 values of critical shear stress and hiding factor from the calibrated model were 0.0386 and 0.905, respectively, as seen in

Table 2. While the value for critical shear stress is located at a relatively flat portion of the distribution, or one with a lower likelihood, the value of hiding factor is located at a particularly steep portion of the distribution; that is, one with a greater likelihood. The hazard of a large value of *m* is that the method forces an exact match to the observations, essentially neglecting measurement and other possible sources of error.

The form of the likelihood function was also assumed in generating the results in the previous section, based on a conceptual extension of the likelihood function presented by Stedinger et al. (2008) to include multiple criteria. However, an alternative likelihood function could be devised by normalizing each output variable and then including them in the function presented by Stedinger et al. (2008). From this logic, the likelihood function would be:

$$L = \exp\left[\frac{-3m}{2} \cdot \frac{\frac{1}{\sigma_{O_1}^2}\sum_{j=1}^{l}\left(O_{1,j} - M_{1,j}\right)^2 + \frac{1}{\sigma_{O_2}^2}\sum_{j=1}^{l}\left(O_{2,j} - M_{2,j}\right)^2 + \frac{1}{\sigma_{O_3}^2}\sum_{j=1}^{l}\left(O_{3,j} - M_{3,j}\right)^2}{\frac{1}{\sigma_{O_1}^2}\sum_{j=1}^{l}\left(O_{1,j} - M_{1,j}^{MLE}\right)^2 + \frac{1}{\sigma_{O_2}^2}\sum_{j=1}^{l}\left(O_{2,j} - M_{2,j}^{MLE}\right)^2 + \frac{1}{\sigma_{O_3}^2}\sum_{j=1}^{l}\left(O_{3,j} - M_{3,j}^{MLE}\right)^2} \right] \qquad (15)$$

where the indices 1, 2, and 3 are the different system outputs, and $\sigma_{O_1}^2$, $\sigma_{O_2}^2$, and $\sigma_{O_3}^2$ are the variances of the observations for each output. The MLE values are calculated in the manner described earlier. A value of 1 is used for the number of independent observations, m, in this application.

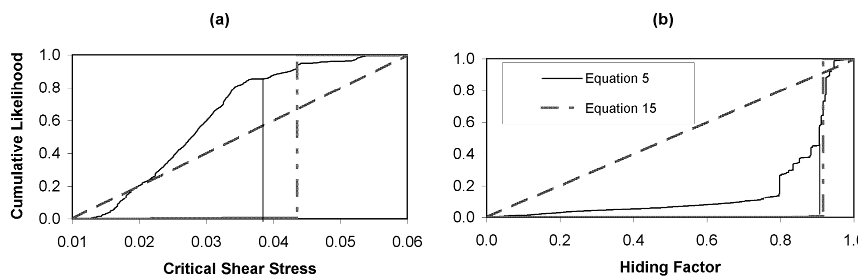

Figure 10. Comparison of the effect of the mathematical form of likelihood function on the posterior cumulative likelihood distributions for (a) critical shear stress and (b) hiding factor, using Equation 5 and Equation 15. Dashed lines indicate the assumed initial uniform distribution for each parameter. Vertical lines indicate the parameter value used in the calibrated model. All results are for the Ashida and Michiue (1971) experiment.

Figure 10 (a, b) shows the posterior cumulative likelihood distributions calculated using both this likelihood function (Equation 15) and the multi-objective likelihood function used earlier (Equation 5). The multiple model likelihood function assigns nearly all of the likelihood to a single parameter value, creating a vertical (or “stair-step”) posterior distribution. The parameter set selected by this function is the MLE parameter set. Part of the reason for this form is that the observations used in this analysis are actually model results, so the MLE is capable of reproducing the results with very little error. Thus, it is judged to have a very high likelihood. However, these results demonstrate that the form of the likelihood function can have major impacts on the results of a GLUE analysis and that the likelihood function contains hidden assumptions about the measurement error and the importance assigned to exactly matching the observations.

The use of calibrated model outputs in place of actual observed data allowed the inclusion of a larger number of observations to constrain the parameters than were actually available in the physical observations. In many cases, smaller numbers of observations or observations of fewer variables are available. To test the impact of the available observations on the results, it is here assumed that no observations for d50 were available. In such a case, the likelihood function will include only two outputs rather than three.

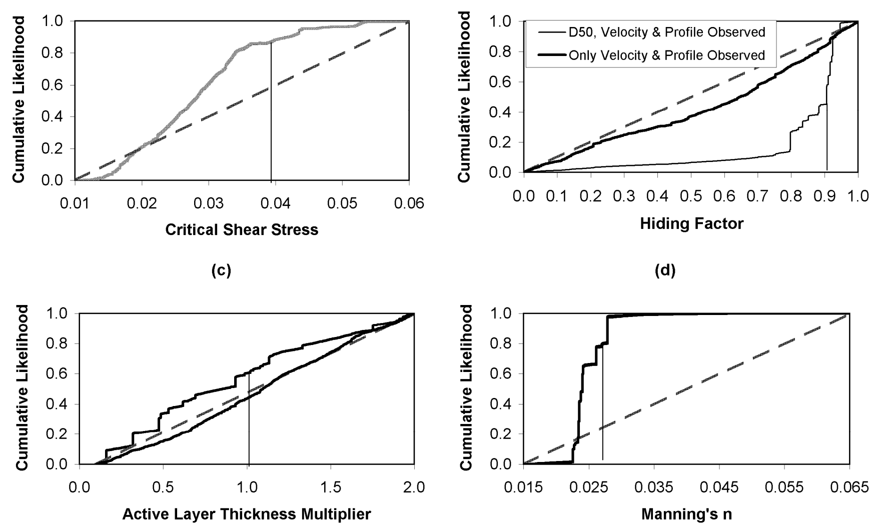

Figure 11. Comparison of posterior cumulative likelihood distributions for (a) critical shear stress, (b) hiding factor, (c) active layer thickness multiplier, and (d) Manning's n when d50 is observed or unobserved. Dashed lines indicate the assumed initial uniform distribution for each parameter. Vertical lines indicate the parameter value used in the calibrated model. All results are for the Ashida and Michiue (1971) experiment.

Figure 11 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

(Figure 11a and Figure 11d) do not change noticeably. In this model, median grain size, d50, is not sensitive to critical shear stress or Manning's n, so its observations have little impact on the likelihood functions for these parameters. The posterior distribution for hiding factor (Figure 11b), however, shows a dramatic change when d50 observations are unavailable, moving closer to a uniform distribution. Because the parameter is assumed to be uniformly distributed in advance of the simulating the calibration period, this implies that the observations from the calibration period are not effective at constraining this parameter. Similarly, the posterior distribution for active layer thickness multiplier (Figure 11c) moves closer to a uniform distribution, implying that the velocity and bed profile observations are of limited effectiveness in restraining this parameter.

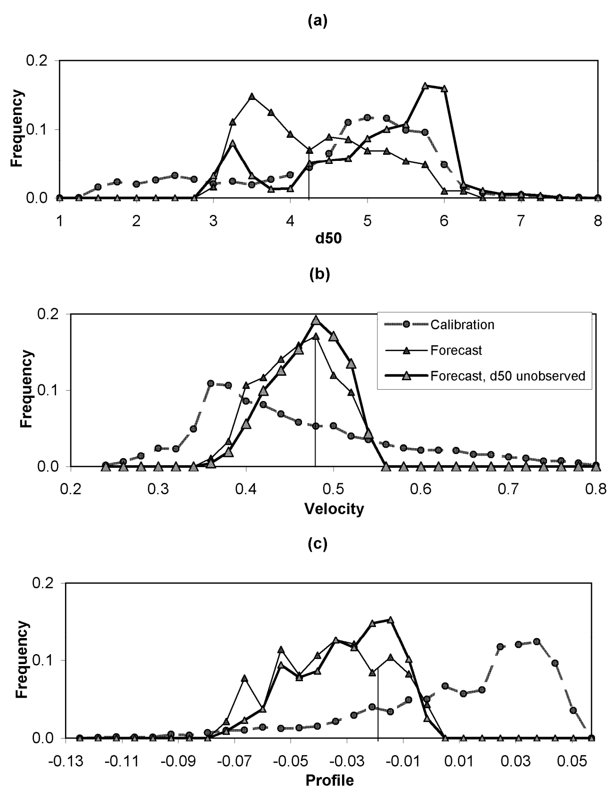

Figure 12. Histograms of (a) d50, (b) velocity, and (c) bed profile for the calibration and forecast periods for the Ashida and Michiue (1971) experiment, with d50 output observed and unobserved. Vertical lines indicate the output value from the calibrated model.

The resulting histograms for d50, velocity, and bed profile for the calibration and forecast periods are shown in Figure 12 (a, b, c). 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 occurs because hiding factor, which is most important in controlling d50 (see Figure 3), was poorly constrained by the calibration observations. Active layer thickness multiplier, also known to impact d50 outputs, is similarly unconstrained. The histograms of velocity and bed profile after the forecast period are similar irrespective of whether d50 was observed or not. 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 identify other, more measurable, outputs that depend on the same parameters.

Another key assumption in the methodology is the choice to neglect correlation between the most likely values of the different parameters. All posterior distributions shown in this paper are marginal distributions, which integrate over all values of the other parameters. These distributions are then used in the LHS method, which implicitly neglects any correlation or dependence in the joint distribution of the parameter values. Figure 13 plots the value of Manning's n against the value of the critical shear stress for the parameter sets in the sample used in the calibration period for the Ashida and Michiue (1971) experiment.

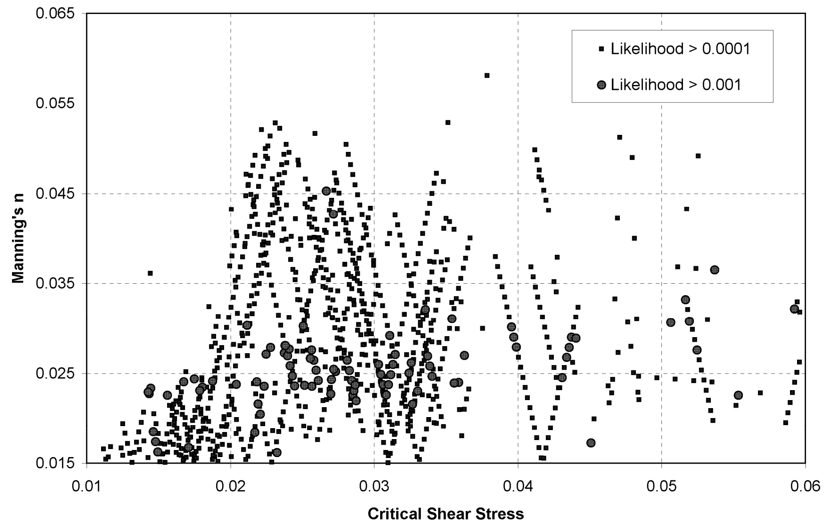

Figure 13. Comparison of likely critical shear stress and Manning's n values, for likelihoods greater than 0.0001, using Equation 5 with equal weights for all outputs. Ashida and Michiue (1971) experiment.

In this plot, only the best performing parameter sets (or most likely parameter sets) are shown. If all points were included, the points would be uniformly distributed. The lines of points visible in the plot occur due to the FAST sampling method described earlier. For this figure, performance was calculated using the likelihood function in Equation (5) with equal weights for d50, velocity, and bed profile. The figure shows that most likely values for critical shear stress and Manning's n values are clustered within a particular region. In addition, the most likely values for Manning's n tend to be larger when the critical shear stress is larger (although much scatter is observed). Based on this observation, some dependence between the distribution of critical shear stress and Manning's n is likely. Similar dependences between critical shear stress and hiding factor, hiding factor and active layer thickness, and hiding factor and Manning's n are suspected. Additional examination would be necessary to determine how such dependences would change the overall assessment of the impacts of parameter uncertainty.

Finally, the impact of the sample size in the LHS method was examined. As described earlier, 6 values were generated for the 4 parameters most important parameters, while the other parameters were fixed at their midpoints. This approach produced a total of 1296 simulations for each experiment. To assess the impact of the sample size, the number of values was varied from 5 to 9 for each parameter, producing sample sizes of 625, 1296, 2401, 4096, and 6561 simulations. Each of these sample sizes produced very similar forecast distributions for the selected outputs. The most likely value varied up to 1.6% for the d50 output (based on the value from the 1296 sample size), up to 0.5% for the velocity output and up to 3.4% for the bed profile output.

7 Conclusions

In this paper, a new method was developed to assess the impacts of parameter uncertainty on the uncertainty of 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 based on 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 model output variables. In the likelihood function, the output variables are weighted using the first order global sensitivities, 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 Ashida and Michiue (1971) experiment, 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 Seal et al. (1997) experiment, median grain size is most dependent on hiding factor and weight of bedload 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. For example, for the erosion case considered here, median grain size is most sensitive to hiding factor and relatively insensitive to the other

parameters. For the depositional case considered here, median grain size is sensitive to critical shear stress, hiding factor, weight of bedload 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. Based on the evaluation of the impacts of parameter uncertainty presented here, weighting the different output variables based on 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 determine the generality of this result and 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 due 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 the structure of measurement and other errors in the analysis. Further research is needed to select and test an appropriate likelihood function. It is recommended that this research begin by applying the methodology to cases where the error and model structures are simple and the likelihood function are known from basic statistics and then begin to transition towards more complex, but interesting, sediment transport cases.

5. In the sampling of the posterior distributions for the parameters, any dependence between the most likely values of different parameters is neglected. Anecdotal evidence in this analysis suggests that dependencies do occur. If present, these dependencies could affect the histograms for the forecast model outputs. Further research should investigate methods to estimate the joint likelihood function.

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., 2007), 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. Additionally, other sediment transport equations such as Meyer-Peter Muller, Laursen, and Ackers-White could be examined to see how the relationships between model parameters and outputs change.

References

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

Ashida, K., and Michiue, M. "An Investigation of River Bed Degradation Downstream of a Dam." *14th International 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." *Hydrology and Earth System Sciences*, 7(3), 304-316.

Beven, K., and Binley, A. (1992). "The Future of Distributed Models: Model Calibration and Uncertainty Prediction." *Hydrological Processes*, 6, 279-298.

Beven, K. J. (2000). *Rainfall-Runoff Modelling: The Primer*, John Wiley & Sons, LTD.

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

Campling, P., Gobin, A., Beven, K., and Freyen, J. (2002). "Rainfall-runoff modelling of a humid tropical catchment: the TOPMODEL approach." *Hydrological Processes*, 16, 231-253.

Chahinian, N., and Moussa, R. (2007). "Comparison of different multi-objective calibration criteria of a conceptual rainfall-runoff model of flood events." *Hydrology and Earth System Sciences Discussions*, 4, 1031-1067.

Chan, K., Saltelli, A., and Tarantola, S. (1997). "Sensitivity Analysis of Model output: Variance-Based Methods make the difference." 1997 Winter Simulation Conference, S. Andradottir, K. J. Healy, D. H. Withers, and B. L. Nelson, eds., 261-268.

Chang, C.-H., and Yang, J.-C. (1993). "Sensitivity and uncertainty analysis of a sediment transport model: a global approach." *Stochastic Hydrology and Hydraulics*, 7, 299-314.

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

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." *The Journal of Chemical Physics*, 59(8), 3873-3878.

Daebel, H., and Gujer, W. (2005). "Uncertainty in predicting riverbed erosion caused by urban stormwater discharge." *Water Science & Technology*, 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 Engineering*, 46, 1-23.

Eidsvik, K. J. (2004). "Some contributions to the uncertainty of sediment transport predictions." *Continental Shelf Research*, 24, 739-754.

Engeland, K., Braud, I., Gottschalk, L., and Leblois, E. (2006). "Multi-objective regional modelling." *Journal of Hydrology*, 327, 337-351.

Engeland, K., and Gottschalk, L. (2002). "Bayesian estimation of parameters in a regional hydrological model." *Hydrology and Earth System Sciences*, 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 Resources Research*, 32, 2161-2173.

Gatelli, D., Kucherenko, S., Ratto, M., and Tarantola, S. (2008). "Calculating first-order sensitivity measures: A benchmark of some recent methodologies." *Reliability Engineering & System Safety*.

Greimann, B. P., Bountry, J., Lai, Y., Mooneyy, D., and Randle, T. "Integrated River Morphology and Vegetation Modeling of the Sacramento River." *Eighth Federal Interagency Sedimentation Conference*, Reno, NV, 972-973.

Greimann, B. P., and Huang, J. "One-Dimensional Modeling of Incision Through Reservoir Deposits." *Eighth Federal Interagency Sedimentation Conference*, Reno, NV, USA, 491-497.

Hall, J. W., Tarantola, S., Bates, P. D., and Horritt, M. S. (2005). "Distributed Sensitivity Analysis of Flood Inundation Model Calibration." *Journal of Hydraulic Engineering*, 117-126.

Harmel, R. D., and King, K. W. (2005). "Uncertainty in Measured Sediment and Nutrient Flux in Runoff from Small Agricultural Watersheds." *Transactions of 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." *Journal of Hydrology*, 362, 89-109.

Holmquist-Johnson, C. L. (2004). "Sediment Model for the Middle Rio Grande - Phase 1: San Acacia Diversion Dam to Elephant Butte Reservoir." U.S. Department of the Interior, Bureau of Reclamation.

Homma, T., and Saltelli, A. (1996). "Importance measures in global sensitivity analysis of nonlinear models." *Reliability Engineering & System Safety*, 52, 1-17.

Huang, J., Greimann, B., and Yang, C. T. (2003). "Numerical Simulation of Sediment Transport in Alluvial River with Floodplain." *International Journal of Sediment Research*, 18(1), 50-59.

Huang, J. V., and Greimann, B. (2007). "User's Manual for SRH-1D V2.1." B. o. R. US Department of Interior, Technical Service Center, Sedimentation and River Hydraulics Group, ed.

Jepsen, R. A. (2006). "Uncertainty in Experimental Techniques for Measuring Sediment Erodibility." *Integrated Environmental Assessment and Management*, 2(1), 39-43.

Legates, D. R., and McCabe Jr., G. J. (1999). "Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation." *Water Resources Research*, 35(1), 233-241.

Mantovan, P., and Todini, E. (2006). "Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology." *Journal of Hydrology*, 330, 368-381.

Mo, X., and Beven, K. (2003). "Multi-objective parameter conditioning of a three-source wheat canopy model." *Agriculture and Forest Meteorology*, 122(1-2), 39-63.

Nash, J. E., and Sutcliffe, J. V. (1970). "River Flow Forecasting through Conceptual Models. Part I - A Discussion of Principles." *Journal of Hydrology*, 10, 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." *Journal of Water Resources Planning and Management*, 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." *Atmospheric Environment*, 38, 2079-2090.

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

Ratto, M., Tarantola, S., and Saltelli, A. (2001). "Sensitivity analysis in model calibration: GSA-GLUE approach." *Computer Physics Communications*, 136, 212-224.

Saltelli, A., and Bolado, R. (1998). "An alternative way to compute Fourier amplitude sensitivity test (FAST)." *Computational Statistics & Data Analysis*, 26, 445-460.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global Sensitivity Analysis: The Primer*, John Wiley & Sons, Ltd.

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." *The Journal of Chemical Physics*, 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." *Journal of Hydraulic Engineering*, 874-884.

Shirmohammadi, A., Chaubey, I., Harmel, R. D., Bosch, D. D., Munoz-Carpena, R., Dharmasri, C., Sexton, A., Arabi, M., Wolfe, M. L., Frankenberger, J., Graff, C., and Sohrabi, T. M. (2006). "Uncertainty in TMDL Models." *Transactions of 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 and Sediment Contamination*, 9(3), 261-289.

Stedinger, J. R., Vogel, R. M., Lee, S. U., and Batchelder, R. (2008). "Appraisal of the generalized likelihood uncertainty estimation (GLUE) method." *Water Resources Research*, 44.

Tarantola, S., Gatelli, D., and Mara, T. A. (2005). "Random balance designs for the estimation of first order global sensitivity indices." *Reliability Engineering & System Safety*, 91, 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." *Journal of Hydraulic Engineering*, 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." *Environmental Modelling & Software*, 20, 19-32.

Wittler, R. J., Klawon, J. E., and Team, F. H. G. (2004). "Upper Gila River Fluvial Geomorphology Study: Final Report." U.S. Department of the Interior, Bureau of Reclamation.

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." *Journal of Hydrologic Engineering*, 440-445.

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

Ziegler, C. K. (2006). "Using Mathematical Models to Assess Sediment Stability." *Integrated Environmental Assessment and Management*, 2(1), 44-50.