Uncertainty analysis of flood inundation modelling using GLUE with surrogate models in stochastic sampling

Yu, J. J.; Qin, Xiaosheng; Larsen, O.

2014

Yu, J. J., Qin, X. S., & Larsen, O. (2014). Uncertainty analysis of flood inundation modelling using GLUE with surrogate models in stochastic sampling. *Hydrological processes*, in press.

<https://hdl.handle.net/10356/79427>

<https://doi.org/10.1002/hyp.10249>

© 2014 John Wiley & Sons, Ltd. This is the author created version of a work that has been peer reviewed and accepted for publication by *Hydrological Processes*. It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [<http://dx.doi.org/10.1002/hyp.10249>].

Downloaded on 26 Apr 2023 18:27:18 SGT

Uncertainty Analysis of Flood Inundation Modeling using GLUE with Surrogate Models in Stochastic Sampling

J.J. Yu^{1,2}, X.S. Qin^{1,3*}, O. Larsen²

1. School of Civil & Environmental Engineering, Nanyang Technological University, 50
Nanyang Avenue, Singapore 639798;

2. DHI-NTU Water & Environment Research Centre and Education Hub, DHI Water &
Environment (S) Pte. Ltd., 1 CleanTech Loop, #03-05, CleanTech One, Singapore 637141;

3. Institute of Catastrophe Risk Management (ICRM), Nanyang Technological University, 50
Nanyang Avenue, Singapore 639798.

***Correspondence:** Dr. X. S. Qin, School of Civil & Environmental Engineering, Nanyang
Technological University, 50 Nanyang Avenue, Singapore 639798; Tel: +65-67905288, Fax:
+65- 67921650; E-mail: xsqin@ntu.edu.sg

Abstract:

A generalized likelihood uncertainty estimation (GLUE) method incorporating moving least squares (MLS) with entropy for stochastic sampling (denoted as GLUE-MLS-E) was proposed for uncertainty analysis of flood inundation modeling. The MLS with entropy (MLS-E) was established according to the pairs of parameters/likelihoods generated from a limited number of direct model executions. It was then applied to approximate the model evaluation to facilitate the target sample acceptance of GLUE during the Monte-Carlo-based stochastic simulation process. The results from a case study showed that the proposed GLUE-MLS-E method had a comparable performance as GLUE in terms of posterior parameter estimation and predicted confidence intervals; however, it could significantly reduce the computational cost. A comparison to other surrogate models, including MLS, quadratic response surface (QRS) and artificial neural networks (ANN), revealed that the MLS-E outperformed others in light of both the predicted confidence interval and the most-likely value of water depths. ANN was shown to be a viable alternative, which performed slightly poorer than MLS-E. The proposed surrogate method in stochastic sampling is of practical significance in computationally expensive problems like flood risk analysis, real time forecasting, and simulation-based engineering design, and has a general applicability in many other numerical simulation fields that requires extensive efforts in uncertainty assessment.

Keywords: GLUE, moving least squares, probability, surrogate, uncertainty, flood inundation

1. INTRODUCTION

Flood inundation models are widely used to simulate historical flood events and predict water depth distributions under various scenarios for flood damage assessment, mitigation planning and engineering design. To quantify effects of uncertainties associated with flood inundation modeling and make predictions more reliable, stochastic modeling techniques are normally applied. The generalized likelihood uncertainty estimation (GLUE) method (Beven and Binley, 1992) is a popular stochastic framework in the field of water resources engineering to evaluate the distribution of predicted variables (e.g. water depth in hydraulics and flow rate in hydrology) and characterize their plausible ranges of fluctuations. GLUE uses a likelihood measure to judge how close the model performance is to the reality with the given parameter set. By adopting a strategy similar to Monte Carlo (MC) simulation and Bayesian updating, GLUE can eventually help identify the parameters' posterior distributions which can then be used for future predictions under uncertainty.

Since the first appearance of GLUE, there are a wide range of applications in water resources area, especially in hydrologic modeling (Christiaens and Feyen, 2002; Blasonea et al., 2008; Jin et al., 2010). It also becomes a popular method for the calibration and uncertainty estimation of flood inundation models. Aronica et al. (1998, 2002) and Bates et al. (2000, 2004) applied GLUE to assess uncertainty in LisFlood model predictions using observed binary pattern information derived from satellite synthetic aperture radars (SAR) images (Horritt and Bates, 2001). Pappenberger et al. (2005, 2006) estimated the

uncertainty of effective roughness parameters, boundary conditions and model structures in HEC-RAS using inundation and downstream level observations. Mason et al. (2009) and Schumann et al. (2009) calibrated the uncertain flood inundation models using water stages. Aronica et al. (2012) derived flood hazard maps incorporating parameter uncertainty of flood inundation models using GLUE. Jung and Merwade (2012) applied GLUE to quantify the uncertainty in generating flood inundation maps, and concluded that the likelihood measure would not impact the estimated flood extent. These studies well demonstrated the capability of GLUE in addressing the effect of uncertainties originated from inputs, parameters and structures of flood inundation models. However, it was also found that the heavy computational burden in the repeated sampling and model evaluation process may seriously restrict the applicability of such a method in large-scale flood modeling studies, where several hours may be required for even a single run.

The surrogate modeling, serving as a statistical emulator, has gained much attention in uncertainty analysis (Reichert et al., 2011; Ratto et al., 2012; Razavi et al., 2012), water resources management (Hemker et al., 2008), and operational control of water systems (Broad et al., 2005; Fu et al., 2009). A wide range of surrogate techniques have been proposed, such as response surface models (RSM) (Seo et al., 2012), artificial neural networks (ANN) (Sreekanth and Datta, 2011), Gaussian process (GP) models (Bilionis and Zabaras), and probabilistic collocation methods (PCM) (Shi et al., 2010). In hydrology/hydraulics community, Shrestha et al. (2009) used an ANN model to approximate a conceptual hydrologic model and concluded that the prediction intervals estimated were reasonably accurate. Hall et al. (2011) constructed a GP emulator of

LisFlood code for model calibration, and argued that the derived probabilistic flood inundation map was more rational than that obtained from GLUE. Song et al. (2008) integrated the RSM-based statistical emulator into the SEC-UA (Duan et al., 1992) method for parameter optimization of a hydrologic model. Nevertheless, there were still limited studies that tried to use surrogate models in the field of flood inundation modeling.

The accuracy of model input-output approximation and the efficiency in implementation are the two major criteria in developing a valid surrogate framework for practical application. Among various alternatives, the global optimum models (like RSM) can be constructed easily but they are generally in short of sufficient accuracy. The ANN model requires a tedious trial-and-error procedure to optimize the network architecture in order to obtain acceptable accuracy. In recent years, the moving least square (MLS) model has shed some light on being an efficient surrogate emulator that is capable of improving the approximation accuracy through construction of local response surfaces, without losing flexibility and efficiency. It has been successfully applied in structure engineering design and optimization (Taflanidis and Cheung, 2012) but has not been tested in the field of flood inundation modeling. In particular, its potential of reducing computational efforts in stochastic sampling process of a GLUE framework has yet to be explored.

Therefore, this study presents a GLUE framework coupled with a MLS-based model in stochastic sampling for flood inundation modeling. Two types of MLS models (with or without entropy) will be examined. A study case in Thames river, UK, simulated using

LisFlood will be selected to demonstrate the applicability of the proposed method. The study results, in terms of target sampling prediction, posterior parameter estimation and uncertain intervals of the predicted water depths, will be compared to those obtained from GLUE. Other alternatives of surrogate models, including quadratic response surface (QRS) and ANN will be compared. The major advantages and limitations of the proposed approach will be discussed afterwards.

2. METHODOLOGY

The proposed GLUE-MLS methodology attempts to construct a surrogate model to approximate the input-likelihood relationship of the flood models through limited number of executions, and then replace the stochastic sampling process for model rejection or acceptance. The original idea of using MLS in stochastic sampling was proposed by Taflanidis and Cheung (2012) and successfully tested in probability modeling of structure mechanics. In this study, the framework of GLUE-MLS is discussed after the introduction of GLUE and formulation of MLS. Other surrogate models like ANN and QRS follow the same uncertainty-analysis framework and will not be introduced in detail.

2.1. Generalized likelihood uncertainty estimation (GLUE)

The GLUE method employs MC simulation to identify target behavior models in order to generate probability distribution functions (PDFs) or cumulative distribution functions (CDFs) of model predictions. To implement GLUE, a large number of parameter sets are

generated from a sampling space with a uniform distribution, or a prior known distribution. Such models with various parameter combinations are then executed and evaluated by a likelihood measure, by comparing the model predictions to observations. The criteria to accept a parameter set could be either the likelihood performance being higher than a cutoff threshold, or a fixed percentage or number of samples from all candidate simulations through ranking. The predicted results (e.g. water depth) with the likelihoods of the retained simulations are finally used to fit the PDF or CDF to generate the confidence intervals of prediction results.

The use of a likelihood function has been criticized as informal and statistically meaningless in a number of literatures (Mantovan and Todini, 2006; Stedinger et al., 2008). However, it is not easy to develop a formal Bayesian-based likelihood measure for statistical inference due to imperfect knowledge and limited observation data. In practical applications, it is usually selected based on the limits of acceptability derived from the observations (Pappenberger et al., 2007; Leedal et al., 2010). In flood inundation modeling, according to the availability of the observed data (e.g. water depth or inundation image), a number of likelihood functions have been reported (Bates and De Roo, 2000; Horritt and Bates, 2001; Aronica et al., 2002). Instead of developing a statistically formal likelihood measure, the focus of this study is to improve the computational efficiency of the GLUE framework, which adopts the MC or Markov chain Monte Carlo (MCMC) strategy for stochastic sampling. Thus, the GLUE methodology could then be extended to large-scale flood inundation problems, where significant computational resources are generally required to run the simulation models.

2.2. Moving least squares response surface with entropy (MLS-E)

The response surface model (RSM) is a general form of surrogate model aiming at approximating the complex process to diminish the cost of repeated functions evaluation (Myers and Montgomery, 2002). The relationship between the model response and input variables is established by fitting the multivariate polynomials using samples obtained from designed numerical experiments over possible parameter sampling space. The QRS is a common choice for approximation, which adopts a second-order polynomials:

$$\hat{f}(x) = \beta_0 + \sum_{i=1}^n \beta_i x_i + \sum_{i=1, i < j}^n \beta_{ij} x_i x_j \quad (1)$$

where x is the input variable; i and j are indexes of variables ($i = 1, 2, \dots, n; j = 1, 2, \dots, n$); β_0 , β_i and β_{ij} are the regression coefficients for intercept, linear terms and quadratic terms, which are usually fitted by the least-squares approach; $\hat{f}$ is the constructed approximation function depending on x .

The drawback of QRS is that it tries to seek a global optimal in fitting equation (1). This may lead to under- or over-estimation in capturing local peaks of a highly nonlinear process. The moving least squares (MLS) response surface (Lancaster and Salkauskas, 1986) aims to construct local response surfaces dynamically for each unknown point, through incorporating the points closest to it within a certain distance. Instead of using constant coefficients (β) on the entire domain estimated from a global response surface, the coefficients are functions of the surrounding influencing points ($\mathbf{x}$). The selection of a second-order approximation in MLS could lead to a direct solution of the coefficients under estimation, which could be represented as (Breitkopf et al., 2005):

$$\hat{f}(\mathbf{x}) = \mathbf{p}(\mathbf{x})^T \mathbf{b}(\mathbf{x}) \quad (2a)$$

$$\mathbf{p}(\mathbf{x}) = [1 \ x_1 \ x_2 \ \dots \ x_n \ x_1^2 \ x_1 x_2 \ \dots \ x_1 x_n \ x_2^2 \ x_2 x_3 \ \dots \ x_n^2] \quad (2b)$$

$$\mathbf{b}(\mathbf{x}) = [b_0(\mathbf{x}) \ b_1(\mathbf{x}) \ \dots \ b_n(\mathbf{x}) b_{11}(\mathbf{x}) \ b_{12}(\mathbf{x}) \ \dots \ b_{1n}(\mathbf{x}) \ \dots \ b_{nn}(\mathbf{x})] \quad (2c)$$

where $\mathbf{p}(\mathbf{x})$ is the basis function; $\mathbf{b}(\mathbf{x})$ is the coefficients under estimation, which is determined by minimizing the weighted least squared errors $J(\mathbf{b})$ over the influencing points $(\mathbf{x})$ between the output of numerical experiments $f(\mathbf{x})$ and the approximation $\hat{f}(\mathbf{x})$ (Breitkopf et al., 2005) :

$$J(\mathbf{b}) = \sum_{i=1}^{ns} w(\|\mathbf{x}_i - \mathbf{x}\|) [\mathbf{p}(\mathbf{x}_i)^T \mathbf{b}(\mathbf{x}) - f(\mathbf{x}_i)]^2 \quad (3)$$

where ns is the total number of the executed simulations with the parameter vectors $(\mathbf{x}_i)$ in the designed experiments; and $w(\cdot)$ is the weighting function depending on the measure of distance. The calculated weight for the designed point $(\mathbf{x}_i)$ increases with the decrease of distance to the interpolation point $(\mathbf{x})$, and vanishing outside the domain of influence. Thus, only limited points $(\mathbf{x}_i, i < ns)$ in the designed experiments are used to construct the local response surface.

The construction of a MLS is sensitive to the choice of the weighting function w . A typical selection of w is the exponential type of function proposed by Taflanidis and Cheung (2012):

$$w(d) = \begin{cases} 0 & d > D \\ \frac{e^{-\left(\frac{d}{cD}\right)^{2k}} - e^{-\left(\frac{1}{c}\right)^{2k}}}{1 - e^{-\left(\frac{1}{c}\right)^{2k}}} & d < D \end{cases} \quad (4a)$$

$$d(\mathbf{x}, \mathbf{x}_i) = \sqrt{\sum_{j=1}^{nc} (x_j - x_{j,i})^2 v_j^2} \quad (4b)$$

where c and k are free parameters, which usually equal to 0.4 and 1 (e.g. applied in this study), respectively, to represent a Gaussian curve; $d(\mathbf{x}, \mathbf{x}_i)$ is the distance measure in the form of weighted quadratic norm, between the interpolation point ($\mathbf{x}$) and each of the influencing point ($\mathbf{x}_i$); v_j represents the relative weight of each parameter x_j ; nc is the number of parameters; D is the search radius of the influencing domain for the point $\mathbf{x}$ under interpolation. The search radius (D) is selected subjectively, on the condition that it should cover sufficient number of influencing points to ensure a successful construction of a local QRS. A good estimation is $nc(nc + 3)/2 + 1$ as proposed by Breikopf et al. (2005). Minimization of the objective function $\mathbf{J}(\mathbf{b})$ yields the solution of Taflanidis and Cheung (2012):

$$\mathbf{b}(\mathbf{x}) = \mathbf{A}(\mathbf{x})^{-1} \mathbf{B}(\mathbf{x}) \mathbf{F} \quad (5a)$$

with

$$\mathbf{A}(\mathbf{x}) = \mathbf{P}^T \mathbf{W}(\mathbf{x}) \mathbf{P}, \quad \mathbf{B}(\mathbf{x}) = \mathbf{P}^T \mathbf{W}(\mathbf{x}), \quad \mathbf{F}(\mathbf{x}) = [f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_{ns})]^T \quad (5b)$$

where:

$$\mathbf{P} = [\mathbf{p}(\mathbf{x}_1), \mathbf{p}(\mathbf{x}_2), \dots, \mathbf{p}(\mathbf{x}_{ns})]^T \quad (5c)$$

$$\mathbf{W}(\mathbf{x}) = \begin{bmatrix} w(d(\mathbf{x}, \mathbf{x}_1)) & 0 & \dots & 0 \\ 0 & w(d(\mathbf{x}, \mathbf{x}_2)) & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & w(d(\mathbf{x}, \mathbf{x}_{ns})) \end{bmatrix}$$

Finally, the MLS approximation in the form of $\hat{f}(\mathbf{x}) = \mathbf{p}(\mathbf{x})^T \mathbf{A}(\mathbf{x})^{-1} \mathbf{B}(\mathbf{x}) \mathbf{F}$ could be obtained for each point through evaluating the basis function and the coefficients at that point.

The relative weight (v_j in equation (4)) of each parameter x_j is crucial for the stochastic sampling process. Taflanidis and Cheung (2012) proposed a quantitative measure to describe the relative importance of each model parameter using relative information entropy, which can be defined as:

$$v_j = \frac{E_j}{\sigma_j} = \frac{\int_{\Theta_j} \pi(x_j) \log \left(\frac{\pi(x_j)}{q(x_j)} \right) dx_j}{\sigma_j} \quad (6a)$$

$$\begin{aligned} E_j &\approx \int_{b_l}^{b_u} \tilde{\pi}(x_j) \log \left(\frac{\tilde{\pi}(x_j)}{\tilde{q}(x_j)} \right) dx_j \\ &\approx \frac{\Delta x_j}{2} \sum_{k=1}^{ne-1} \left[\tilde{\pi}(x_{j,k}) \log \left(\frac{\tilde{\pi}(x_{j,k})}{\tilde{q}(x_{j,k})} \right) + \tilde{\pi}(x_{j,k+1}) \log \left(\frac{\tilde{\pi}(x_{j,k+1})}{\tilde{q}(x_{j,k+1})} \right) \right] \end{aligned} \quad (6b)$$

$$\pi(x_j) \approx \tilde{\pi}(x_j) = \frac{1}{n\sqrt{2\pi}(1.06n^{-1/5}\sigma_{sj})} \sum_{s=1}^n e^{-\frac{(x_j - x_{j,s})^2}{2(1.06n^{-1/5}\sigma_{sj})^2}} \quad (6c)$$

where j indicates the index of parameter; E_j is the relative entropy expressing the difference between the marginal target distributions $\pi(x_j)$ and the marginal proposal distribution $q(x_j)$; σ_j is the variance of the influencing points and Θ_j is the parameter sampling space. Considering the analytical expression for $\pi(x_j)$ and $q(x_j)$ is not available to calculate E_j in equation (6a), an approximation is employed using the integral in the possible sampling region of $[b_l, b_u]$; the integration could be calculated using equation (6b) following trapezoidal rule with ne points where $\Delta x_j = (b_u - b_l) / ne$ and

$x_{j,k} = b_l + k\Delta x_j$. Accordingly, $\tilde{\pi}(x_j)$ is the approximation of $\pi(x_j)$ based on the Gaussian kernel density estimator with totally n samples for x_j ; s denotes the index of samples; and

σ_{sj} is the standard deviation of these samples. The $\tilde{q}(x_j)$ could be evaluated similarly using equation (6c) with samples in the marginal proposal distribution $q(x_j)$. The detailed derivation and background theory can be referred to Taflanidis and Cheung (2012).

The MLS has been widely adopted in computational mechanics as a mesh-free alternative for solving PDEs. It is thus investigated in this study for (1) establishing the procedures of integrating it with GLUE to reach an efficient stochastic sampling; (2) exploring the entropy-based weight selection in MLS for optimum surrogate; and (3) evaluating the capability of MLS by comparing it with other models (i.e. QRS and ANN).

2.3. GLUE with MLS-based surrogate modeling (GLUE-MLS)

In order to estimate the uncertainty associated with flood inundation modeling arising from parameters $\mathbf{x}$ (e.g. roughness coefficients), a stochastic problem is formulated. The uncertain parameters $\mathbf{x}$ are considered to subject to a uniform distribution $p(\mathbf{x})$ initially (i.e. prior distribution). Their beliefs are then updated for posterior PDF summarization through the model calibration process with observations z' . Of particular interest in such a context is to generate samples according to the posterior PDF as Lamb et al. (1998) :

$$\pi(\mathbf{x}|\mathbf{x} \in \Theta) = \frac{L(\mathbf{x}|z')p(\mathbf{x})}{\int_{\Theta} L(\mathbf{x}|z')p(\mathbf{x})d\mathbf{x}} \propto L(\mathbf{x}|z')p(\mathbf{x}) \quad (7)$$

where $L(\mathbf{x}|z')$ is likelihood evaluation relative to model output $z(\mathbf{x})$ and observations; Θ denotes posterior parameter space conditioned on observations; and $\propto$ denotes proportionality. Thus, the denominator is simplified to a normalization constant through scaling the likelihood values. Finally, the uncertain quantity of the interested model

output could be determined through model execution over the drawn samples from $\pi(\mathbf{x})$.

The MC or MCMC simulation is an easy-to-develop algorithm for sample generation from $\pi(\mathbf{x})$; however they are generally inefficient due to the repeated evaluation of the model response through direct model execution for each candidate sample. In this study, a surrogate model based on MLS is established to approximate the likelihood evaluation of the model based on limited execution of the model. It is then incorporated into the framework of MC-based stochastic sampling for model evaluation in order to reduce the computational burden. The workflow of GLUE-MLS includes the following steps (as shown in Figure 1):

Place Figure 1 here

Step 1: Select the number of samples ns and na for surrogate model construction and total candidate samples, respectively ($ns \ll na$); identify the sensitive parameters with the possible sampling space; and define the likelihood function $L(\mathbf{x})$ with the cutoff threshold in GLUE;

Step 2: Generate ns parameter sets from uniform distribution using Latin hypercube sampling (LHS) algorithm (Cheng and Druzdzel, 2000); run the model directly and evaluate the performance with the likelihood function L . Thus, ns pairs of the model input

and likelihood response could be obtained as $\{(\mathbf{x}_1, L(\mathbf{x}_1)), (\mathbf{x}_2, L(\mathbf{x}_2)), \dots (\mathbf{x}_{ns}, L(\mathbf{x}_{ns}))\}$.

Step 3: Construct the surrogate model based on the samples obtained in Step 2.

Optionally, the established surrogate model could be validated by additional model input/likelihoods pairs through direct model runs. It could be an iterative process to include new generated samples for surrogate model construction until the accuracy of the likelihood approximation is acceptable. Such a process will increase the computational burden, but lead to much reliable surrogate models in practical applications. In this study, two types of MLS will be tested as the surrogates, including MLS-E (i.e. MLS with entropy) and MLS (MLS without entropy). They are compared with the alternatives of QRS and ANN.

Step 4: Generate na candidate samples using LHS algorithm; rank the model performance according to the predicted likelihood value $\hat{L}(\mathbf{x})$ from the constructed MLS model; Use GLUE to accept nt samples $\{(\mathbf{x}_1, \hat{\hat{L}}(\mathbf{x}_1)), (\mathbf{x}_2, L(\mathbf{x}_2)), \dots (\mathbf{x}_{nt}, L(\mathbf{x}_{nt}))\}$ out of na candidates for posterior distributions estimation according to the specified threshold;

Step 5: Run the flood inundation model for prediction under uncertainty using the nt samples of parameters obtained in Step 4; the predicted variables of interest could be extracted from model outputs for PDF or CDF fitting. It should be noted that if the surrogate model is constructed for approximation of the model inputs with the model response (e.g. water depth) instead of the likelihoods, the model execution could be avoided since the outputs of the surrogate model could be used for uncertainty estimation

directly.

Determination of n_s for surrogate model construction could be based on the trial-and-error method. An appropriate number (e.g. 150 in this study) of samples is tried first for surrogate model construction, considering affordable computational efforts; the surrogate model should then be verified by results from direct model executions; if not satisfactory, additional samples are added to n_s for surrogate model refinement until an acceptable accuracy is achieved. An alternative way, as proposed by Taflanidis and Cheung (2012), is to start an initial MC simulation to ensure a certain number ($\ll nt$) of acceptable samples; these samples with those rejected ones then form the training points for surrogate model construction.

3. CASE STUDY

3.1. Flood inundation modeling under uncertainty

LisFlood is a raster-based inundation model based on an one-dimensional (1D) kinematic or diffusive wave equation capturing the downstream propagation of a flood wave in channel and a two-dimensional (2D) flood spreading model for overland flow (Bates and De Roo, 2000). The 1D St. Venant equations are solved numerically using finite difference technique with a fully implicit scheme, with the channel parameters of width, bed slope, depth and Manning's n . The 2D surface flow over spilled from the channel is modeled between the adjacent cells according to a function of the free surface height difference between those cells. Although it is not accurate enough to represent the full

hydrodynamics of the surface flow, it has been applied into various real world study cases with acceptable prediction accuracy (Horritt and Bates, 2001; Aronica et al., 2002). More importantly, the code is computationally efficient in application to the selected study case which makes the MC simulation possible for benchmarking.

A study case of a reach in the upper Thames near Buscot, UK is selected to demonstrate the applicability of the proposed approach (Figure 2). It suffered from a 1-in-5 year flood event in December 1992 with a peak discharge of $73 \text{ m}^3 \text{ s}^{-1}$. The test reach is 5 km long, which is bounded upstream by a gauged weir at Buscot. The floodplain inside the study site is 3.8 km in x direction and 2.4 km in y direction, with vertical accuracy of 25 cm and horizontal accuracy of 50 m topographic data available (Horritt and Bates, 2001). Various uncertainty sources associated with flood inundation modeling may influence the prediction accuracy and reliability. They may include input hydrographs, topographical data, selection of the models, representation of hydraulic structures, model parameters, observation data, and approaches employed for data analysis (Jung and Merwade, 2012). Based on the sensitivity analysis results from Hall et al. (2005), three parameters, including channel width (w), roughness for channel (n_c) and floodplain (n_{fp}), are selected in this study to demonstrate the proposed GLUE-MLS framework for uncertainty assessment. The reasonable ranges of the uncertain parameters are set as: (i) 0.01 to 0.05 for the channel roughness, (ii) 0.02 to 0.10 for the floodplain roughness, and (iii) 17 to 23 m for the channel width (Hall et al., 2005).

With the availability of the flood inundation extent derived from a SAR image only, the

likelihood function used for model calibration is defined as (Bates and De Roo, 2000):

$$L(\mathbf{x}) = \frac{\sum_{i=1}^n P_i^{D_1 M_1}}{\sum_{i=1}^n P_i^{D_0 M_1} + \sum_{i=1}^n P_i^{D_1 M_0} + \sum_{i=1}^n P_i^{D_1 M_1}} \quad (8)$$

where i is index of the grids in the model; n is number of the grids; $P_i^{D_1 M_0}$ is a binary variable where its value is 1 when the i th grid is found flooded in the observation data but not flooded in the modeled results and 0 for other conditions; $P_i^{D_0 M_1}$ is a binary variable where its value is 1 when the i th grid is not flooded by observation but flooded in the modeled results, and 0 for other conditions; $P_i^{D_1 M_1}$ represents a binary variable where its value is 1 only when both the observation and modeled results indicate a flooding condition for the i th grid, and 0 for other conditions; $L(\mathbf{x})$ shows the goodness-of-fit of the simulated flood inundation extent, with parameters $\mathbf{x}$ being used for the model. The closer the value of $L(\mathbf{x})$ is to 1, the better the simulation behavior. It should be noted that application of the proposed GLUE-MLS framework is not restricted to specific likelihood functions, where the surrogate model is constructed based on the relationship between the parameters and the evaluated likelihood values. In case the water depths are available at m sites, the mean square error could be used as a likelihood function (Wenner et al., 2005)

$$MSE = 1 - \frac{\sum_{j=1}^m (\hat{h}_j - h_j)^2}{\bar{d}} \quad (9)$$

where $\hat{h}$ and h are the j th water levels from simulation and observation, respectively; and $\bar{d}$ is the mean of the observed water depths. If both water depths and inundation

maps are available, equations (8) and (9) could be combined using the weighted sum approach (Wenner et al., 2005).

Place Figure 2 here

3.2. Result analysis

Initially, 150 samples of parameters (ns) are randomly generated from the possible sampling space using LHS algorithm and then they are simulated from direct evaluation of flood inundation model using equation (8). The surrogate model of MLS-E is then constructed based on the initial samples (as inputs) and the corresponding likelihoods (as outputs). In construction of MLS-E, the search radius D , as shown in equation (4), is selected adaptively to include at least 10 points in order to build local quadratic response surfaces considering the case having 3 model parameters. Then, additional 1,370 samples (nt) are selected from 15,000 candidates (na) whose approximated likelihoods from the constructed surrogate models are larger than 0.74. These 1,370 samples are eventually simulated in LisFlood and the CDF of the predicted water depths could be fitted based on the corresponding likelihood values. As a benchmark for comparing the efficiency of stochastic sampling and accuracy of prediction, a direct GLUE method with 15,000 simulations is applied for the same problem. The surrogate model approach only requires about 1,520 ($ns + nt$) real model runs, and GLUE has to go through all 15,000 runs in this study.

Figure 3 shows a comparison of GLUE and GLUE-MLS-E in inundation prediction

under uncertainty. Figures 3a and 3b illustrate the predicted flood inundation maps at the 5th and 95th percentiles obtained from both methods, respectively. It appears that the flood inundation extent predicted by GLUE-MLS-E is close to that of GLUE at both 5th and 95th percentiles. Generally, the GLUE-MLS-E predicts a slightly smaller inundation area at probability of 5th but a larger area at 95th percentiles. Both predictions could cover over 95% of the observed flood inundation map of the 1992 event in the predicted results. It implies that the GLUE-MLS-E is a viable and efficient alternative of GLUE. Figures 3c to 3e illustrate the water-depth contour maps with 0.2 m interval obtained from GLUE and GLUE-MLS-E at 25th, 50th and 95th percentiles for the selected area (as shown in Figures 3a and 3b). The GLUE-MLS-E method predicts the water depth distributions reasonably well at 50th percentile. At a lower percentile level (i.e. 25%), the GLUE-MLS-E method considerably underestimates the water depth than that from GLUE at the same location and also shows a smaller inundation extent. When the percentiles are above 50% (i.e. 95%), the predicted water depths from GLUE-MLS-E are slightly higher than those from GLUE; the corresponding inundation area is larger. These results are consistent with those inferred from Figures 3a and 3b. These results demonstrate that the error from surrogate approximation is more sensitive in the predictions at both lower and higher cumulative probability levels than those at the mean level. Generally, the spatial distribution of the simulated water depths from GLUE-MLS-E is comparable to that obtained from GLUE. At various levels of percentiles, the GLUE-MLS-E method shows a better fit with GLUE in the locations near the channel than those in the boundary area.

Place Figures 3 here

Figure 4 shows the posterior distributions for n_c , n_{fp} , and w generated from GLUE-MLS-E and GLUE according to the histogram of the acceptable samples. It appears that the GLUE-MLS-E method shows similar posterior distributions for the three uncertain parameters as GLUE does, where the Normal, Weibull and beta distributions best fit parameters of n_c , n_{fp} and w , respectively. GLUE-MLS-E could capture the posterior distribution as good as GLUE for n_{fp} and w , but somewhat poorer for n_c . It is indicated that the fitted mean of n_c from both methods is around 0.026; the mean for n_{fp} is 0.072 from GLUE-MLS-E, which is slightly lower than that (0.073) from GLUE. However, the standard deviation of the fitted posterior distribution of n_c (0.001) is smaller than that of GLUE (0.003), indicating that GLUE-MLS-E is more capable of reducing the degree of uncertainty for sensitive parameters (i.e. n_c) than GLUE. Generally, the surrogate approach employing MLS-E, which has a much higher computational efficiency, shows a comparable performance to GLUE with regard to the prediction of water depth and flood inundation extent, as well as estimation of posterior parameters.

Place Figures 4 here

The proposed GLUE-MLS-E framework is further compared with other three alternatives, including GLUE-MLS, QRS, and ANN in terms of the accuracy of the target sample

prediction and estimation of the posterior parameters. The optimal architecture of ANN employing the back-propagation training algorithm is identified by trial and error. Eventually, a 4-layer neural network is used with 5 and 3 neurons in two hidden layers. The performance of the surrogate model could be evaluated by the following three criteria: coefficient of determination (R^2) (O'Connell et al., 1970), root mean squared error ($RMSE$) (Karunanithi et al., 1994) and mean absolute percentage error ($MAPE$) (Hu et al., 2001). They could be calculated as:

$$R^2 = \left[\frac{\sum_{i=1}^n (f(\theta_i) - \bar{f}) (\hat{f}(\theta_i) - \bar{f})}{\sqrt{\sum_{i=1}^n (f(\theta_i) - \bar{f})^2} \sqrt{\sum_{i=1}^n (\hat{f}(\theta_i) - \bar{f})^2}} \right]^2 \quad (10a)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f(\theta_i) - \hat{f}(\theta_i))^2} \quad (10b)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{f(\theta_i) - \hat{f}(\theta_i)}{f(\theta_i)} \right| \times 100 \quad (10c)$$

where $f(\theta_i)$ and $\hat{f}(\theta_i)$ are the results from direct and surrogate models for the i th point, respectively; $\bar{f}$ and $\bar{\hat{f}}$ are the means of the model results (over n points) obtained from direct model evaluation and surrogate emulator, respectively. R^2 indicates how well the correlation between the model evaluation and the approximation is. The closer the value of R^2 to 1, the better the fit of the approximation. $RMSE$ evaluates the residual between the direct model evaluation and the approximation, and $MAPE$ is a weighted average of the absolute errors. A smaller value of $RMSE$ (or $MAPE$) indicates a better fit. In this study, the constructed surrogate approaches are verified using the same 15,000 samples whose likelihoods are obtained from GLUE. However, the number of

samples for model verification could be selected according to the available computational resources in practical application. From Table 1, the R^2 ranges from 0.787 to 0.874, $RMSE$ ranges from 0.033 to 0.043, and $MAPE$ ranges from 2.569 to 3.506. Generally, the MLS-E slightly outperforms ANN method, and shows much improvement over MLS and QRS.

Place Table 1 here

The capability of deriving the posterior distribution of uncertain parameters could be evaluated by comparing the estimated sample mean and the related statistical results referencing to the most-likely values obtained from GLUE (which runs based on a large number of samples). It was considered to have significant impact on predicting the future events and, thus, on the decision making of engineering design. Table 2 shows the sample mean (μ), the estimated error ($|\beta - \mu|/\beta$), the sample coefficient of variance ($cv_s: \sigma/\mu$), and the relative error of cv ($|cv_g - cv_s|/cv_g$) from various surrogate emulators, along with the sample mean (β) obtained from GLUE which is considered to be the referencing most-likely value of the posterior parameter estimation. It appears that the sample mean estimated from MLS-E and ANN fitted comparatively well for all of the most-likely parameters referencing to GLUE; the estimated error is less than 3.5% for n_c , 4.1% for n_{fp} and 1% for w respectively. The relative error of cv for n_{fp} and w is less than 13% and 3% respectively by MLS-E and ANN; this shows their capabilities of reducing the degree of uncertainty are better than MLS and QRS. In fitting n_c , the cv_s for all of the surrogate

models vary from 9.5% to 14.5%, which are lower than that from GLUE (i.e. 17.8%). It implies that the surrogate models show the characteristics of lower-variance in relation to mean for the posterior samples of sensitive parameter (i.e. n_c) than GLUE. Nevertheless, QRS performs worst since the estimated error of sample mean of n_c increases to 18.85%, although its cv_s appears to be relatively close to that of GLUE. Overall, it is demonstrated that the surrogate approaches based on MLS-E and ANN have comparable performances in estimating the parameter means and reducing the associated uncertainties, and they are both better than MLS and QRS.

Place Table 2 here

The reliability of the predicted water depths is of particular interest in practical applications. Considering the uncertainty parameters are jointly-distributed, the derived posterior samples are directly simulated to obtain the predicted water depth. They are then used to fit the CDF curves of model outputs with corresponding likelihoods by histogram analysis. The 5th and 95th percentiles of the CDF curves are used as the lower and upper boundaries of the confidence intervals. Figure 5 shows a comparison of the predicted intervals for the calibration event from various approaches for the grids intersected with the cross-section line (see Figure 2), which is identified by the distance to the lower-left point. It is indicated that all of the surrogate approaches provides wider predicted intervals in comparison to those obtained from GLUE. Such a prediction discrepancy is caused by the approximated likelihood evaluation during the stochastic

sampling process in compromise of saving the computational effort. Selection of different surrogate models would lead to various rankings for the same parameter set for model rejection/acceptance and thus influence the samples acceptable for posterior distribution. Eventually, the confidence intervals derived from the simulations based on the accepted samples from posterior distributions show different characteristics.

From Figure 5, QRS shows the best fitting of the lower bounds of the confidence intervals of the predicted water depths, and the MLS-E and ANN perform relatively better in fitting the upper bound; this may because QRS aims at achieving global optimum and fits relatively worse in local extremes compared with MLS-based algorithms and ANN; this leads to a better performance of QRS in fitting lower-level values with relatively narrower deviations. The confidence intervals of MLS is wider than MLS-E (i.e. performs relatively worse referencing to GLUE). This shows that introducing the entropy in selection of the weight for MLS could improve the quality of approximation, thus, leads to the better performance. It should be noted that the same posterior samples could be used to simulate the water depths and derive the confidence intervals under various flood events by changing the input flows. Since similar characteristics of performance for various surrogate approaches are found, the related details are not demonstrated.

Place Figure 5 here

In shortage of the observed water depth, the accuracy of the derived confidence intervals is difficult to be verified (as only the coverage probability of the observed values is available). The relative error of the predicted intervals (E_b) and the relative error of the most likely value (E_m) of the water depths are used to evaluate the performance of various surrogate approaches in deriving confidence intervals. They can be calculated by referring to the results obtained from GLUE at target numerical grids, represented as:

$$E_b = \frac{|w_u - \hat{w}_u| + |w_l - \hat{w}_l|}{w_u - w_l} \times 100\% \quad (11a)$$

$$E_m = \frac{|w_m - \hat{w}_m|}{w_u} \times 100\% \quad (11b)$$

where w_l , w_m , and w_u denotes the predicted water depth extracted from the fitted CDF at cumulative probabilities of 5%, 50% and 95%, respectively. The symbol w indicates the value obtained from GLUE and $\hat{w}$ indicates that obtained from surrogate approach. A smaller value of E_b or E_m indicates that the surrogate approach could achieve a closer result as GLUE regarding the predicted uncertain interval and most likely value. Figure 6 shows the levels of E_b and E_m along the cross section of the river (see Figure 2). MLS-E, QRS and ANN indicate a better capability (with E_b being less than 1.0) in estimating the confidence intervals than MLS. However, MLS shows has a better performance in fitting the most likely values (slightly better than MLS-E and ANN with E_m being less than 0.05) for all grids of concern; while, QRS performs poorer. From these results, it implies that MLS-E is a more effective surrogate approach for stochastic sampling in GLUE, in the sense of predictions of both confidence intervals and most-likely values; ANN is also a viable alternative which performs slightly poorer than MLS-E.

Place Figure 6 here

3.3. Discussion

By introducing the surrogate approaches to approximate the model evaluation, the number of the model executions for calibration could decrease to the number of samples selected for posterior distribution estimation (i.e. nt) plus additional number of samples (i.e. ns) required for training the surrogate models. Comparing to GLUE, which requires running the original numerical model for na times (equaling to the total number of candidate samples), a surrogate framework could achieve the following percentage saving of computational effort (C):

$$C = \frac{na - nt - ts}{na} \times 100\% \quad (12)$$

In this study, roughly one-tenth of the computational resources is required (C is approximately 90%) by applying the proposed framework without losing much accuracy. Generally, the purpose of incorporating surrogate models into stochastic sampling is to obtain a trade-off between computational efficiency and reliability of uncertainty assessment. Thus, it could be applied into many computationally expensive problems like flood risk analysis, real-time flood forecasting, and simulation-based engineering design where uncertainty has to be tackled by large number of runs like MC or MCMC methods. Particularly, the proposed framework is also applicable for speeding up the evaluation of parameter uncertainty in physically-based hydrological forecasting models subject to the

adoption of different likelihood functions (like equation 9). It should be noted that the way of computing errors between simulated and observed data in likelihood functions between the two types of models is different due to the fact that the hydrological modeling process normally uses temporal changes of river flow at fixed outlet points; while flood modeling is normally based on spatial information (e.g. water depth and inundation map) at fixed time points. Moreover, introduction of surrogate approaches is not necessary in cases where the data-driven models are employed, as these models are normally efficient enough in implementing MC simulations directly.

The accuracy of the constructed surrogate model is essential in the proposed framework. In this study, the MLS method with entropy (GLUE-MLS-E) has proved to be a comparable surrogate model to ANN, and is superior to both MLS and QRS, in the sense of replacing the time-consuming stochastic sampling process to rank the likelihoods in relation to the candidate samples for simulation rejection/acceptance in a GLUE framework. Besides the first attempt by Taflanidis and Cheung (2012) in applying MLS in building stability assessment during dynamic earthquakes, Taflanidis (2012) also applied it in developing stochastic subset optimization algorithm for engineering design of buildings under uncertainty.

Among various surrogate options, the MLS-based methods could capture the local optimums and thus lead to relatively better performance than QRS, without losing efficiency and flexibility. It is also superior to ANN in the sense of mitigating the over-fitting problem through retaining only the most contributive training samples during

prediction. Although ANN shows a comparable performance in this study, such a feature is still important in saving time for training and trial-and-error based model configuration, particularly in those cases where the number of training samples is large. The MLS-E model outperforms MLS in terms of likelihood approximation through introducing the entropy to quantify the relative importance of each individual parameter. It implies that determination of the relative weights is crucial for surrogate model construction in efficient stochastic sampling. There may be other alternative ways of identifying relative weights, such as Akaike information criterion (Zhao et al., 2008); but their applicability and performance need further exploration. Furthermore, the MLS-based methods employ explicit expressions (i.e. 2nd polynomials) for approximation; this shows a potential of revealing the correlation implied in the system.

4. CONCLUSIONS

A GLUE framework incorporating moving least squares response surface (MLS) with entropy (GLUE-MLS-E) as a surrogate model in stochastic sampling was proposed for uncertainty analysis of flood inundation modeling. The MLS-E was established based on the parameters/likelihoods pairs simulated from limited direct model executions. The GLUE was then applied for uncertainty analysis based on the acceptable samples according to the likelihood rankings from the MLS-E approximation. A study case was selected to demonstrate the application of the proposed GLUE-MLS-E framework. The results showed that it could perform comparably well as GLUE in terms of the posterior parameter estimation and the confidence intervals of the predicted water depth. It was also superior to other options (e.g. MLS and QRS) considering the compromise of the

accuracy and easiness of usage. The ANN could be a viable alternative but performed slightly poorer than GLUE-MLS-E.

This study made the first attempt to incorporate a surrogate emulator into GLUE for improving efficiency of uncertainty assessment. The computational cost associated with stochastic simulation could be notably reduced, which allows GLUE to be more applicable for dealing with large-scale flood risk analysis problems. A number of limitations of the proposed methodology are also expected to be tackled in the future. Firstly, the additional uncertainty introduced by model approximation should be properly considered. Secondly, due to a lack of observation data from multiple flood events, the dependence of effective roughness values on various magnitudes of events through the proposed surrogate-based GLUE framework is to be discussed. Finally, the investigation of impacts of various likelihood measures and cutoff thresholds is also desired to be explored.

ACKNOWLEDGEMENT

This research was supported by Singapore's Ministry of Education (MOM) AcRF Tier 1 Project (M4010973.030) and Tier 2 (M4020182.030) Project. The study also received support from DHI Water & Environment (S) Pte. Ltd. and DHI-NTU Water and Environment Research Centre and Education Hub, Singapore. The authors are deeply grateful to Dr. Paul Bates (University of Bristol) for providing the LISFLOOD-FP model and the relevant test data, and also much appreciate the reviewers for their insightful and valuable reviews.

References

- Aronica GT, Bates PD, Horritt MS. 2002. Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. *Hydrological Processes* **16**: 2001-2016.
- Aronica GT, Candela A, Fabio P, Santoro M. 2012. Estimation of flood inundation probabilities using global hazard indexes based on hydrodynamic variables. *Physics and Chemistry of the Earth* **42-44**: 119-129.
- Aronica GT, Hankin B, Beven KJ. 1998. Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data. *Advances in Water Resources* **22(4)**: 349-365.
- Bates PD, De Roo APJ. 2000. A simple raster-based model for floodplain inundation. *Journal of Hydrology* **236**: 54-77.
- Bates PD, Horritt MS, Aronica GT, Beven KJ. 2004. Bayesian updating of flood inundation likelihoods conditioned on flood extent data. *Hydrological Processes* **18(17)**: 3347-3370.
- Beven KJ, Binley A. 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes* **6**: 279-298.
- Bilionis I, Zabaras N. 2012. Multi-output local Gaussian process regression: Applications to uncertainty quantification. *Journal of Computational Physics* **231(17)**: 5718-5746.
- Blasonea RS, Madsenb H, Rosbjerga D. 2008. Uncertainty assessment of integrated distributed hydrological models using GLUE with Markov chain Monte Carlo sampling. *Journal of Hydrology* **353(1-2)**: 18-32.
- Breitkopf P, Naceur H, Rassineux A, Villon P. 2005. Moving least squares response

surface approximation: Formulation and metal forming applications. *Computers & Structures* **83**:1411-1428.

Broad D, Dandy G, Maier H. 2005. Water Distribution System Optimization Using Metamodels. *Journal of Water Resources Planning and Management* **131**(3): 172-180.

Cheng J, Druzdzal MJ. 2000. Latin hypercube sampling in Bayesian networks. *Proceedings of the Thirteenth International Florida Artificial Intelligence Research Society Conference, AAAI Press*: 287-292.

Christiaens K, Feyen J. 2002. Constraining soil hydraulic parameter and output uncertainty of the distributed hydrological MIKE SHE model using the GLUE framework. *Hydrological Processes* **6**(2): 373-391.

Duan QY, Sorooshian S, Gupta VK. 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resource Research* **28**: 1015-1031.

Fu G, Khu ST, Butler D. 2009. Use of surrogate modelling for multiobjective optimisation of urban wastewater systems. *Water Science and Technology* **60**(4): 1641-1647.

Hall JW, Manning LJ, Hankin RKS. 2011. Bayesian calibration of a flood inundation model using spatial data. *Water Resource Research* **47**(5): DOI:10.1029/2009WR008541.

Hall JW, Tarantola S, Bates PD, Horritt MS. 2005. Distributed Sensitivity Analysis of Flood Inundation Model Calibration. *Journal of Hydraulic Engineering* **131**(2): 117-126.

Hemker T, Fowler KR, Farthing MW, Stryk O. 2008. A mixed-integer simulation-based

-
- optimization approach with surrogate functions in water resources management. *Optimization and Engineering* **9(4)**: 341-360.
- Horritt MS, Bates PD. 2001. Predicting floodplain inundation: raster-based modelling versus the finite element approach. *Hydrological Processes* **15**: 825-842.
- Hu TS, Lam KC, Ng ST. 2001. River flow time series prediction with a range dependent neural network. *Hydrological Sciences Journal* **46 (5)**: 729-745.
- Jin XL, Xu CY, Zhang Q, Singh VP. 2010. Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. *Journal of Hydrology* **383(3-4)**: 147-155.
- Jung YG, Merwade V. 2012. Uncertainty Quantification in Flood Inundation Mapping Using Generalized Likelihood Uncertainty Estimate and Sensitivity Analysis. *Journal of Hydraulic Engineering* **17(4)**: 507-520.
- Karunanithi N, Grenney WJ, Whitley D, Bovee K. 1994. Neural networks for river flow prediction. *Journal of Computational Civil Engineering* **8(2)**: 201–220.
- Lamb R, Beven KJ, Myrabbø S. 1998. Use of spatially-distributed water table observations to constrain uncertainty in a rainfall-runoff model. *Advances in Water Resources* **22**: 305-317.
- Lancaster P, Salkauskas, K. 1986. *Curve and surface fitting: an introduction*. Academic Press.
- Leedal DT, Neal J, Beven KJ, Young P, Bates PD. 2010. Visualization approaches for communicating real-time flood forecasting level and inundation information. *Journal of Flood Risk Management* **3**: 140-150.
- Mantovan P, Todini E. 2006. Hydrological forecasting uncertainty assessment:

-
- Incoherence of the GLUE methodology. *Journal of Hydrology* **330(1-2)**: 368-381.
- Mason DC, Bates PD, Dall' Amico JT. 2009. Estimating uncertainty associated with water stages from a single SAR image. *Advances in Water Resources* **31**: 1038-1047.
- Myers RH, Montgomery DC. 2002. *Response surface methodology, 2nd ed.* New York.
- O'Connell PE, Nash JE, Farrell JP. 1970. River flow forecasting through conceptual models part II — The Brosna catchment at Ferbane. *Journal of Hydrology* **10(4)**: 317-329.
- Pappenberger F, Beven KJ, Horritt M, Blazkova S. 2005. Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations. *Journal of Hydrology* **302**: 46-69.
- Pappenberger F, Frodsham K, Beven KJ, Romanovicz R, Matgen P. 2007. Fuzzy set approach to calibrating distributed flood inundation models using remote sensing observations. *Hydrology and Earth System Sciences* **11(2)**: 739-752.
- Pappenberger F, Matgen P, Beven KJ, Henry J, Pfister L, Fraipont de P. 2006. Influence of uncertain boundary conditions and model structure on flood inundation predictions. *Advances in Water Resources* **29**: 1430-1449.
- Ratto M, Castelletti A, Pagano, A. 2012. Emulation techniques for the reduction and sensitivity analysis of complex environmental models. *Environmental Modelling & Software* **34**: 1-4.
- Razavi S, Tolson BA, Burn DH. 2012. Review of surrogate modeling in water resources. *Water Resources Research* **48**: DOI:10.1029/2011 WR011527.
- Reichert P, White G, Bayarri MJ, Pitman EB. 2011. Mechanism-based emulation of dynamic simulation models: Concept and application in hydrology. *Computational*

Statistics & Data Analysis **55(4)**: 1638-1655.

Schumann G, Pappenberger F, Matgen P. 2009. Calibration of uncertain flood inundation models using remotely sensed water levels. *Journal of Hydrology* **368**: 224-236.

Seo J Duenas-Osorio L, Craig JI, Goodno BJ. 2012. Metamodel-based regional vulnerability estimate of irregular steel moment-frame structures subjected to earthquake events. *Engineering Structures* **45**: 585-597.

Shen ZY, Chen L, Chen T. 2012. Analysis of parameter uncertainty in hydrological and sediment modeling using GLUE method: a case study of SWAT model applied to Three Gorges Reservoir Region, China. *Hydrology and Earth System Sciences* **16(1)**: 121-132.

Shi LS, Zhang, DX, Lin L, Yang JZ. 2010. A multiscale probabilistic collocation method for subsurface flow in heterogeneous media. *Water Resources Research* **46**. DOI:10.1029/2010WR009066.

Shrestha DL, Kayastha N, Solomatine DP. 2009. A novel approach to parameter uncertainty analysis of hydrological models using neural networks. *Hydrology and Earth System Sciences Discussions* **13**: 1235-1248.

Song XM, Zhan CS, Xia J. 2012. Integration of a statistical emulator approach with the SCE-UA method for parameter optimization of a hydrological model. *Chinese Science Bulletin* **57(26)**: 3397-3403.

Sreekanth J, Datta B. 2011. Comparative Evaluation of Genetic Programming and Neural Network as Potential Surrogate Models for Coastal Aquifer Management. *Water Resources Management* **25(13)**: 3201-3218.

Stedinger JR, Vogel RM, Lee SU, Batchelder R. 2008. Appraisal of the generalized

likelihood uncertainty estimation (GLUE) method. *Water Resources Research* **44**(12): DOI:10.1029/2008WR006822, 2008.

Taflanidis AA. 2012. Stochastic subset optimization incorporating moving least squares response surface methodologies for stochastic sampling. *Advances in Engineering Software* **44**: 3 - 14.

Taflanidis AA, Cheung SH. 2012. Stochastic sampling using moving least squares response surface approximations. *Probabilistic Engineering Mechanics* **28**: 216-224.

Werner M, Blazkova S, Petr J. 2005. Spatially distributed observations in constraining inundation modelling uncertainties; *Hydrological Processes* **19**: 3081–3096.

Zhao ZY, Zhang Y, Liao HJ. 2008. Design of ensemble neural network using the Akaike information criterion. *Engineering Applications of Artificial Intelligence* **21**: 1182 - 1188.

Table Caption List:

Table 1. Goodness-of-fit of the likelihood approximation with various surrogate models

Table 2. Comparison of the statistical results for posterior parameter estimation

Table 1. Goodness-of-fit of the likelihood approximation with various surrogate models

<i>Methods</i>	R^2	<i>RMSE</i>	<i>MAPE</i>
MLS-E	0.874	0.033	2.569
MLS	0.824	0.039	3.016
QRS	0.787	0.043	3.506
ANN	0.870	0.033	2.571

Notes: the surrogate models are verified using the 15,000 samples, with likelihoods being obtained from GLUE.

Table 2. Comparison of the statistical results for posterior parameter estimation

<i>Parameter</i>		β	μ	$ \beta - \mu / \beta$ (%)	$cv_s: \sigma / \mu$ (%)	$ cv_g - cv_s / cv_g$ (%)
MLS-E	n_c	0.0268	0.0262	2.05%	9.62%	45.96%
	n_{fp}	0.0674	0.0647	4.04%	46.12%	6.61%
	w	20.0876	20.077	0.05%	54.83%	2.14%
MLS	n_c	0.0268	0.0278	3.6%	13.01%	26.91%
	n_{fp}	0.0674	0.0586	13.58%	61.46%	42.07%
	w	20.0876	20.1745	0.43%	57.82%	3.19%
QRS	n_c	0.0268	0.0318	18.85%	14.50%	18.54%
	n_{fp}	0.0674	0.0586	56.6%	68.78%	59.00%
	w	20.0876	19.9585	0.64%	52.50%	6.30%
ANN	n_c	0.0268	0.0259	3.34%	10.17%	42.87%
	n_{fp}	0.0674	0.0680	1.86%	37.73%	12.78%
	w	20.0876	20.0392	0.24%	56.31%	0.50%

Notes: β is the sample mean of the posterior parameters using GLUE; μ is the sample mean of posterior parameters using surrogate approaches; σ is the standard derivation; cv_g (σ / β) and cv_s (σ / μ) are the sample coefficient of variance using GLUE and surrogate approaches respectively.

List of Figure Captions

Figure 1. Workflow of the GLUE-MLS framework. Note: ns , nt , na are the numbers of samples for construction of surrogate models, estimation of posterior distributions and total candidate samples, respectively.

Figure 2. Map of the study area (adapted from Hall et al. 2011 and Aronica et al. 2012).

Figure 3. GLUE-MLS-E vs. GLUE in inundation prediction under uncertainty. Note: the top two sub-figures include maps of inundation probability predicted from (a) GLUE-MLS-E and (b) GLUE; the bottom three sub-figures include water-depth contour maps generated from GLUE-MLS-E and GLUE at (c) 25th, (d) 50th, (e) 95th percentiles, respectively.

Figure 4. Posterior parameter distributions of (a) n_c , (b) n_{fp} , and (c) w from GLUE-MLS-E and GLUE.

Figure 5. Comparison of the water-depth predictions along the river cross section through different surrogate models. Note: d is the distance to the lower-left grid; the top and bottom lines of the box represent 75th and 25th percentiles of the results, respectively; the bars at the top and bottom represent upper and lower whiskers, respectively.

Figure 6. Relative errors of the predicted (a) confidence interval, and (b) mean along the cross sections through different surrogate models. Note: the results from GLUE are used as benchmarks for calculating relative errors.