Uncertainty in model-based prediction of copper loads in stormwater runoff

Incertitude des prédictions basées sur les modèles des charges de cuivre dans les ruissellements des eaux pluviales

Erik Lindblom*, Stefan Ahlman** and Peter Steen Mikkelsen*

* Institute of Environment & Resources, Technical University of Denmark, Bygningstorvet, Building 115, DK-2800 Kgs. Lyngby, Denmark. E-mail: eul@er.dtu.dk; psm@ er.dtu.dk

** Water Environment Technology, Chalmers University of Technology, SE-412 96 Göteborg, Sweden. E-mail: stefan.ahlman@chalmers.se

RESUME

Cet article décrit une analyse systématique de l'incertitude liée à l'estimation de la charge polluante totale (cuivre) dans un système séparatif de drainage des eaux pluviales, sur la base d'une combinaison spécifique de données d'entrée, d'un modèle d'accumulation-lessivage de polluants, et de mesures (volumes de ruissellement et charges polluantes). Nous avons utilisé la méthodologie d'estimation de mesure généralisée d'incertitude (GLUE) et nous avons généré des distributions de paramètres postérieurs qui résultent des sorties du modèle englobant un nombre significatif de mesures fortement variables. Compte tenu du modèle appliqué d'accumulation-lavage de pollution et un nombre total de 57 mesures pendant un mois, la charge de cuivre totale prévue peut être comprise dans une plage de ± 50% de la valeur moyenne.

ABSTRACT

In this paper, we conduct a systematic analysis of the uncertainty related with estimating the total load of pollution (copper) from a separate stormwater drainage system, conditioned on a specific combination of input data, a pollutant accumulation-washout model and measurements (runoff volumes and pollutant masses). We use the generalized likelihood uncertainty estimation (GLUE) methodology and generate posterior parameter distributions that result in model outputs encompassing a significant number of the highly variable measurements. Given the applied pollution accumulation-washout model and a total of 57 measurements during one month, the total predicted copper mass can be predicted within a range of ± 50% of the median value.

KEYWORDS

GLUE, Micro-pollutants, Load estimation, Stormwater modelling, Uncertainty.

1 INTRODUCTION

Suppose you have access to a number of catchment-specific stormwater quality measurements and wish to make a statement about the magnitude of pollutant loads from the area. Most often, you would base your statement on a more or less complex model, a function that transfers knowledge about the catchment rainfall-runoff processes into a model response. The selection of an appropriate model type is a question in itself. However, whatever model you select, you are required to accept that the model output does not equal one unique true solution to the problem. To post the statement you will need to say something about the reliability and validity of the model output.

In this work focus lies on determining the uncertainty of a conceptual stormwater quality model given a, in this context relatively detailed measurement campaign. The motivation for determining this is an aim to know to what extent micro-pollutant loads in stormwater systems can be estimated. As a reference compound, the heavy metal copper was selected.

The paper sets out from the ideas of the generalized likelihood uncertainty estimation (GLUE) method of Beven and Binley (1992). GLUE was developed for hydrological models and has been applied in the area of hydrology several times since it was first presented (see e.g. Beven and Freer, 2001). The main message is that calibration of conceptual models is to be done keeping in mind that several parameter sets equally well describe your measured observations. The method makes use of Bayes' Theorem and the modeller is allowed to subjectively assess when the uncertainty is adequately described, generally when simulations with a number of derived parameter sets yield an output that cover a significant number of observations.

2 SITE DESCRIPTION AND FIELD DATA

All data in this work come from an urban catchment called Vasastaden in the city of Göteborg, Sweden. The area is densely populated and consists mainly of older residential and commercial buildings. The catchment has a total impervious area of 4.83 ha and a separate sewer system. For a detailed description of the case study the reader is directed to Ahlman (2006).

Measurements of rainfall, stormwater flow and stormwater quality were undertaken in April-May 2002. An ISCO 6700 automatic water sampler was installed in the vicinity of a manhole to take samples in a Φ400 mm separate storm sewer made out of concrete. A flow meter forced the sampler to take flow-weighted samples. Rain data was collected with a tipping bucket rain gauge (type HoBo/MJK), located approximately 60 meters from the sampler on the boundary of the catchment.

For analysis 13 rain events during a period of 30 days were identified. The rainfall for the events ranged between 0.8 and 11.7 mm with durations from 0.4 to 9.7 hours. The maximum intensities (with a one-minute resolution) ranged between 0.4 and 3.7 mm/h. Within 8 hours after each rain event, the collected stormwater samples were transported to the laboratory, where they were analysed for pH, conductivity, total suspended solids (TSS), chemical oxygen demand (COD) and for heavy metals (copper, zinc, lead and cadmium). In total, the 13 events included analyses of 57 copper concentrations, which were representative for 57 runoff volumes. We combined this information to form an observation vector m_k (k=1,2,…,57) [µg], the observed total masses of copper in runoff during the sampling periods.

3 METHOD

3.1 Model description

We apply the pollutant accumulation and wash-off module of SEWSYS (Ahlman, 2006) to simulate the load of copper in the stormwater. SEWSYS is a conceptual stormwater model developed for simulations of substance flows in urban drainage systems, running in the mathematical software MATLAB/Simulink. In the considered model pollutants are accumulated in dry periods and washed off during rainfall, processes described with classical build-up and wash-off functions (Overton and Meadows, 1976). The mass flow of a considered pollutant in the stormwater runoff is obtained from the following mass balance around the (not observed) amount of pollutant $X(t)$ [µg m^{-2}], accumulated on the surfaces:

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = \Theta_1 - \left(\Theta_2 + \Theta_3 \cdot p(t)\right) \cdot X(t), \qquad X(0) = \frac{\Theta_1}{\Theta_2} \tag{1}$$

The dry deposition load Θ_1 [µg·s^{-1}·m^{-2}] is assumed to be constant and represents different sources of pollution, e.g. traffic activities, surface corrosion and atmospheric deposition. The rate coefficient for pollutant dry removal Θ_2 [s^{-1}] describes removal by wind and other means, a process which is assumed to be proportional to the accumulated mass. The wet removal by wash-off is assumed to be proportional to the accumulated mass and the rain intensity $p(t)$ [µm·s^{-1}] with a rate constant Θ_3 [µm^{-1}]. In all simulations in this paper, it is assumed that the system is at steady state at the time for the first event, which yields $X(0)= \Theta_1/\Theta_2$. The simulated observations Y_k [µg] (k=1,2,…,57) represent predictions of the measured observations m_t, which were taken over the sample lengths i_k [s] centred at t_k [s] (see also Figure 1, right):

$$Y_k = \int_{t_k - i_k/2}^{t_k + i_k/2} \left(\Theta_3 \cdot p(t) \cdot X(t)\right) \mathrm{d}t \cdot \mathrm{A} + \mathrm{v}_{\mathrm{Runoff},k} \cdot \mathrm{c}_{\mathrm{Rain}} \tag{2}$$

The effective rain intensity, runoff volume $\mathrm{v}_{\mathrm{Runoff},k}$ [l] and impervious area A [m^2] has been calibrated in a previous study (Ahlman *et al., In prep.*) and are considered to be fixed. The concentration of copper in rain water, $\mathrm{c}_{\mathrm{Rain}}$ [µg·l^{-1}] is assumed to be constant and low (1 µg·l^{-1}) so that it does not affect the results much. Left for analysis is the parameter vector $\Theta = (\Theta_1,\Theta_2,\Theta_3)$.

3.2 Uncertainty analysis

Although the proposed model is complex compared to those often used in the area of micro-pollutants in stormwater, it is a crude simplification of reality. If it is accepted to apply such a simplified model, model parameters yielding solutions that are almost equally good should also be accepted. One single “optimal” solution, e.g. one that minimizes/maximizes a specific objective function is of less interest. This is the basis for the GLUE method, which has inspired this work. The uncertainty analysis presented here is based on the premise that input data uncertainty, model structural uncertainty and measurement uncertainty can be lumped into model parameter uncertainty. This view, that the parameter vector Θ is a random variable, should be interpreted as a way to model the uncertainty involved in all assumptions of the model (Equations 1 and 2). If we let y_k (k=1,2,…,57) and $\theta = [\theta_1, \theta_2, \theta_3]$ be realisations of Y_k and Θ respectively, this uncertainty is, in a Bayesian framework, described through the posterior parameter distribution $f(\Theta|Y_k=y_k,m_k)$, the distribution of the parameters conditioned on data. Following Bayes’ Theorem, this distribution is proportional to the product of the likelihood $L(Y_{tk}|\Theta=\theta,m_k)$, the distribution of simulated observations conditioned on the parameters, and $\pi(\Theta)$, the prior distribution of parameters reflecting our (lack of) knowledge about these before observing the data.

$$f(\Theta|Y_k = y_k, \mathrm{m}_k) \propto L(Y_k|\Theta = \theta, \mathrm{m}_k)\cdot \pi(\Theta) \tag{3}$$

From the posterior distribution $f(\Theta|Y_k{=}y_k,\mathrm{m}_k)$, answers about the statistical properties of Θ then follow directly from probability theory. Of specific interest here is to choose the likelihood function and prior distribution so that the simulated observations (with θ drawn from f) lie in a region A_k, which is large enough to encompass the measured observations with high probability P:

$$P(Y_k(\Theta) \in A_k|Y_k = y_k, \mathrm{m}_k) = \int \mathbf{1}\{Y_k(\Theta) \in A_k\} \cdot f(\Theta|Y_k = y_k, \mathrm{m}_k)\mathrm{d}\Theta \tag{4}$$

where $\mathbf{1}\{Y_k(\Theta) \in A_k\}$ is an indicator function:

$$\mathbf{1}\{Y_k(\Theta) \in A_k\} = \begin{cases} 1 & \text{if } Y_k(\Theta) \in A_k \\ 0 & \text{else} \end{cases} \tag{5}$$

In practice, the e.g. P=95% quantiles are the ranges $[A_1, A_2, \ldots, A_{57}]$, within which 95% of the modelled observations fall. Part of the notation is exemplified to the right in Figure 1 below.

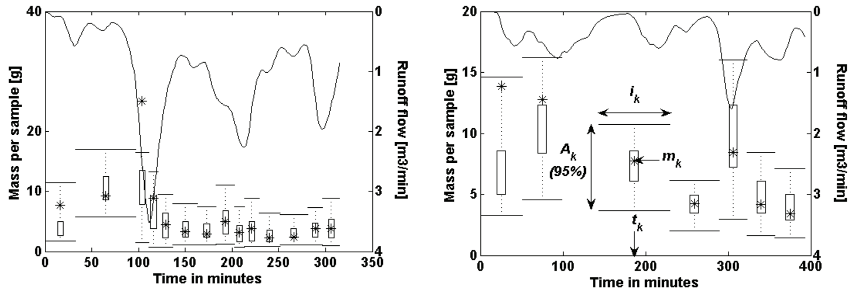

Figure 1: Measured (stars) and simulated (95 and 50% quantiles) observations for 2 out of 13 events. Time normalised to zero at start of event. For further discussion, see the results section.

To calculate probabilities as in Equation 4, two issues need to be formalised: (*i*) how are the prior distribution and the likelihood function defined? and (*ii*) how are draws from the posterior generated? For the first issue, our suggestion follows one of the basic thoughts behind the GLUE methodology, that is to chose a likelihood function which the modeller can manipulate to obtain a reasonable prediction uncertainty (A_k above). For the second issue, two popular techniques (also mentioned in Kuczera and Parent, 1992) to draw from f are implemented: A Markov Chain Monte Carlo (MCMC) method and importance sampling. A dense survey of the theory behind these two methods can be found in e.g. Robert and Casella (2004).

3.2.1 Choosing the prior distribution and likelihood function

In *the prior distribution*, the modeller includes information about the parameters that he/she had before obtaining the observed data. In the case of little prior knowledge, the common choice is to select uniform prior distributions. There are incentives to constrain the priors into a hypercube, one reason being to reduce the parameter space, which the computational algorithms (see 3.2.2 below) need to explore. Based on the discussion in Ahlman *et al.* (*In prep.*) the constraints were chosen as shown in

Table 1. Regarding Θ_2, any model describing build-up as a dynamic process with rate 1.3-13 days are given equal probability 1 whereas faster or slower models are disregarded. Regarding the parameter Θ_3, which determines the rate of storage depletion, we constrain it to be maximum 0.056 μm^{-1}, which corresponds to a storage depletion of approximately 15 minutes for a (hypothetical, rectangular) rain with intensity 0.02 $\mu m\ s^{-1}$. Finally, for the dry deposition load Θ_1 we constrained an upper limit of 1 $\mu g \cdot s^{-1} \cdot m^{-2}$. This was adjusted so that the choice of prior did not affect the possibility of covering the observations.

Parameter description	Notation and unit	Min. value	Max. value
Dry deposition load	Θ_1: [$\mu g \cdot s^{-1} \cdot m^{-2}$]	0	1
Dry removal rate	Θ_2: [s^{-1}]	$1.5 \cdot 10^{-5}$	$1.5 \cdot 10^{-4}$
Wet removal rate constant	Θ_3: [μm^{-1}]	0	$5.56 \cdot 10^{-2}$

Table 1: Overview of parameters used and their prior distribution.

Based on a realisation θ from the prior distribution, *the likelihood function* compares the simulated observation with the corresponding measured one. It can thus be seen as a measure of goodness-of-fit. A common choice is to assume that the residuals between modelled and measured observations are independent and normally distributed. This was done in Kanso *et al.,* (2005) who studied suspended solid concentrations in stormwater runoff with a model similar to the one presented here. In our case this assumption yields confidence limits not large enough to encompass a desirable number of the measured observations. Out of several likelihood functions used in GLUE studies (Beven and Freer, 2001), we instead apply:

$$L(Y_k \mid \Theta = \theta, \mathrm{m}_k) = \exp\left(-\frac{\sum_{k=1}^{57} (y_k - \mathrm{m}_k)^2}{T} \right) \quad (6)$$

where exp denotes the exponential function. Equation 6 was also used by Mailhot *et al.,* (1997) to investigate the uncertainty related to stormwater quality modelling (using artificial data). The parameter T can be seen as a scaling factor whose value depend upon the confidence the modeller has on measurements compared to the model. A small value of T will result in a peaked posterior distribution and tight uncertainty bounds while a larger value will widen the posterior and the uncertainty bounds. By varying T, it is thus possible for the practitioner to manipulate the modelled uncertainty.

3.2.2 Sampling of the posterior distribution

Having defined the prior distribution and likelihood function, the posterior distribution is known up to proportionality (Equation 3). The remaining problem is that to evaluate the probability of Equation 4, independent draws from the posterior distribution $f(\Theta|Y_k=y_k,\mathrm{m}_k)$ are required.

With a *Markov Chain Monte Carlo (MCMC)* method, N draws $\theta^1,\ldots,\theta^N$ are directly generated from f by producing an ergodic Markov chain whose stationary distribution is f. The distribution of Y_k can subsequently be estimated empirically by Monte-Carlo integration of Equation 4. Out of several possible MCMC methods, we implemented the so-called SCEM-UA algorithm proposed in Vrugt *et al.,* 2003. This method has recently also been applied in a GLUE study by Blasone *et al.* (*In prep.*).

Importance sampling, sometimes also called weighted sampling, re-formulates Equation 4 so that draws from f can be replaced by draws from an almost arbitrary importance sampling distribution, in our case the prior uniform distribution π. Although computationally expensive, the algorithm is often used in GLUE applications and is easy to implement with this setting:

Repeat for *n*=1,2, …, N:

1. Draw a parameter set θ^n from the prior hypercube distribution (Table 1).
2. Calculate *X*(0) and simulate the entire 30 day period with the model (Equation 1)
3. Observe $[y_1^n, y_2^n, \ldots, y_{57}^n]$ with Equation 2 , calculate an importance weight (in our case Equation 5) and normalise it:

$$P^n = \frac{L(Y_k \mid \theta^n, m_k)}{\sum_{n=1}^{N} L(Y_k \mid \theta^n, m_k)} \tag{7}$$

The set $\{\theta^n, P^n\}$ know represents *N* weighted random samples from the posterior *f* with P^n interpreted as the probability of drawing θ^n (Kuzcera and Parent, 1998). For each sample *k*, the mass predicted by each parameter set is ranked in order of magnitude and, using the weights associated with each set, a distribution function of the prediction is calculated (Figure 2 below).

4 RESULTS

The posterior distribution was approximated by both importance and MCMC sampling for various values of the scaling parameter *T*. The two methods yielded nearly identical distributions of Y_k. With $T = 3 \cdot 10^8$, we judged that the uncertainty was adequately described. Figure 2 shows the 57 measured and simulated observations.

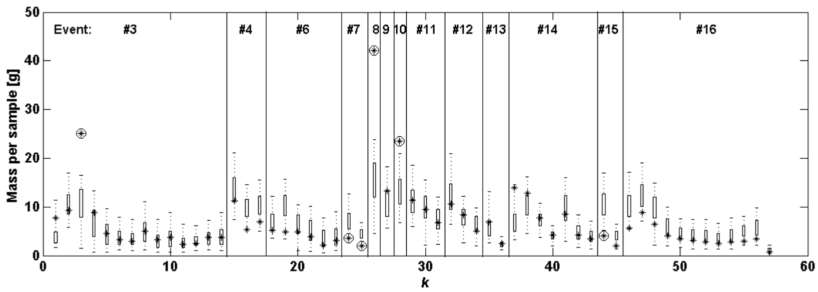

Figure 2: Results from 100 000 importance samples with T=3·10^8. Measured (stars) and simulated observations (95 and 50% quantiles).

There is some subjectivity involved with the judgement. Six out of 57 (10.5%) encircled measurements fall outside the modelled 95% quantiles, which could be argued to be on the upper edge. However, in this work we include few assumptions such as presence of outliers. There are e.g. two measured observations (3 and 10) that the model fails to predict, even if *T* is chosen significantly higher. It should also be noticed that 35 of the measured observations (61%) lie within the 50% prediction limit.

Figure 2 does not show the intra-event dynamics of the model. To study this, one can refer to Figure 1. In general, the model has some difficulties with predicting high initial peaks in the copper concentrations, which can be seen as high copper loads in the beginning of the events. It is also clear from the figure that it is highly uncertain to predict e.g. the copper concentration at a certain time.

In Figure 3, 10 000 draws from the posterior distribution obtained with MCMC sampling are shown. It can be seen, that to simulate the observations with the proposed model, a wide range of parameter sets need to be considered. The banana-shaped correlation between the initial pollutant mass (θ_1/θ_2) and wet removal coefficient in the right plot arises because when the initially accumulated pollutant mass is low, the model output is no longer sensitive to increases of the wet removal rate parameter; all pollutants are washed off any way. To the left it is seen that the posterior distribution has moved quite far away from the prior distribution. To make a similar plot as Figure 3 with the results from importance sampling, only approximately 5% of the parameter sets with highest likelihood weights would be selected. This implies that 95 % of the samples are basically unnecessary. This inefficiency of importance sampling becomes more and more evident as the number of random variables increases, as prior knowledge decreases as well as when the model takes a long time to simulate. In such cases, MCMC sampling is to be preferred.

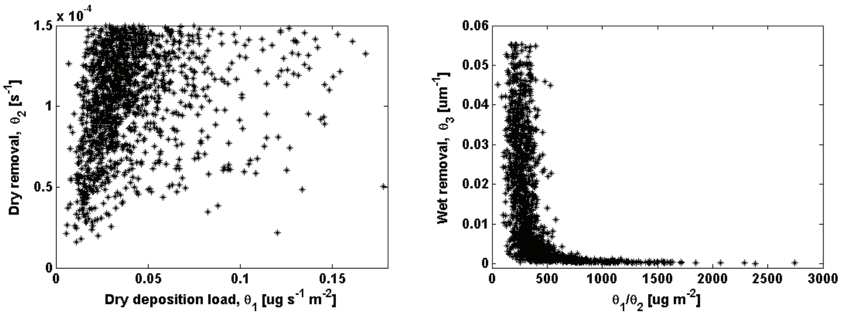

Figure 3: 10 000 draws from the posterior distribution by MCMC sampling.

Besides a simulated observation of the 57 measured masses, each sampled parameter set is associated also with cumulative (the sum of) masses throughout the 30 day period. In Figure 4 (left) the total masses of each event has been divided by the event runoff volumes to give the event mean concentrations (EMCs). The events not encompassed by simulations are each based on few (1-2) measured observations. To the right the cumulative masses, from the first sample 1 to the final sample 57 are shown. The jump at k=26 is caused by the abnormally high observed mass in this sample. The final cumulative mass gives the answer to our original question, which is how accurately we can determine the total mass of copper. Given the proposed model the total sampled copper mass is 206-576 g with 95% probability and 327-459 g with 50% probability. The median of the predicted total mass is 385 g.

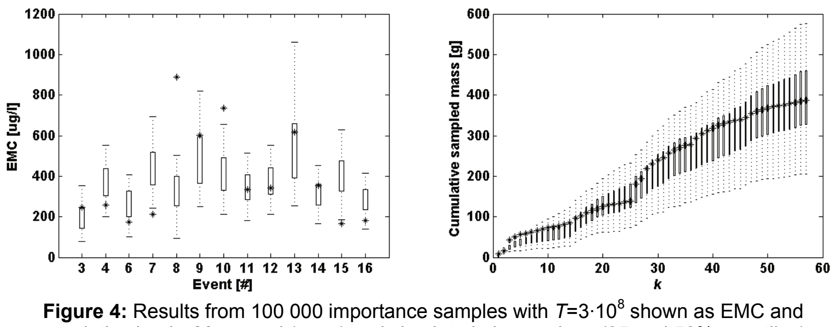

Figure 4: Results from 100 000 importance samples with T=3·10^8 shown as EMC and cumulative loads. Measured (stars) and simulated observations (95 and 50% quantiles).

5 CONCLUSIONS

With the proposed model and input data from the 30 day period, the applied uncertainty assessment methodology states that the total sampled copper mass can be predicted within a range of ± 50% of the median value (385 g). This final uncertainty includes both model structure uncertainty, input data uncertainty as well as measurement uncertainty, but no attempts to distinguish between the various sources of uncertainty have been made. The message is that given the model, the considered catchment and rain data, and the 57 measured runoff volumes and masses, this relatively large uncertainty should be acknowledged in connection with posting statements about the actual copper loads. If the model is to be used for prediction in other areas or for other rainfall conditions, the uncertainty ranges will be different. While viewing the rather high final uncertainty one should bear in mind that few assumptions such as presence of outliers have been made.

The proposed uncertainty analysis method is a great tool for assessment of uncertainty related with predicting micro-pollutants in stormwater. It is fairly simple to implement and flexible to update if e.g. more data become available or if other case studies are to be considered. The method provides and keeps track of the correlation of parameters, information which is vital for the design of future experiments, as well as for motivating new and better model structures. It may furthermore be used to justify a simplified model structure when limited data is available.

LIST OF REFERENCES

Ahlman S., Lindblom E. and Mikkelsen P.S. (*In prep.*). Uncertainty analysis of the substance flow model SEWSYS using sampled field data for calibration.

Ahlman, S. (2006). *Modelling of substance flows in urban drainage systems.* PhD Thesis, Dept. of Civil and Environmental Engineering, Division of Water Environment Technology, Chalmers University of Technology, Göteborg.

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

Beven, K. and Freer, F. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *J. Hydrol.,* 249, 11-29.

Blasone, R. S., Vrugt, J.A., Madsen, H., Rosbjerg, D., Zyvoloski G.A. and Robinson B.A. (*In prep.*). Generalized Likelihood Uncertainty Estimation (GLUE) Using Adaptive Markov Chain Monte Carlo Sampling.

Kanso, A., Chebbo, G. and Tassin, B. (2005). Stormwater quality modelling in combined sewers: calibration and uncertainty analysis. *Wat. Sci. & Tech.,* 52(3), 63-71.

Kuczera, G. and Parent, E. (1998). Monte Carlo assessment of uncertainty in conceptual catchment models: the Metropolis algorithm. *J. Hydrol..* 211, 69-85.

Mailhot, A., Gaume. E. and Villeneuve, J.-P. (1997). Uncertainty analysis of calibrated parameter values of an urban storm water quality model using Metropolis Monte Carlo algorithm. *Wat. Sci. & Tech.,* 36(5), 141-148.

Overton, D. E. and Meadows, M. E. (1976). *Stormwater Modeling*. Academic Press, New York.

Robert, P. R. and Casella, G. (2004). *Monte Carlo Statistical Methods* (2nd Ed). Springer, New York.

Vrugt, J.A., Gupta, H.V, Bouten, W. and Sorooshian, S. (2003). A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resour. Res.,* 39(8), 1201.