Bayesian model calibration using structural reliability methods: application to the hydrological *abc* model

W. Betz[1], C.M. Mok[1], I. Papaioannou[1] and D. Straub[1]

[1]Engineering Risk Analysis Group, Technische Universität München, Theresienstr. 90, 80333 München, Germany; email: wolfgang.betz@tum.de

ABSTRACT

Bayesian updating using structural reliability methods (BUS) is applied to calibrate the simple hydrological *abc* model to observations of a hypothetical real world case. The assumed hypothetical real world is chosen such that the *abc* model cannot represent it perfectly. The likelihood function is expressed in terms of the measurement error and the modeling error. The probability distributions of both errors are only approximate, due to a lack of knowledge about the true behavior. The correlation structure of the modeling error is regarded as partially uncertain and inferred in the updating process. It is highlighted that under the presence of modeling errors, the predictive distribution of the model output is not the same as the predictive distribution of the true discharge.

INTRODUCTION

In hydrological model calibration long time series of relevant observed rainfall and discharge are commonly available (Beven 2009). Due to the spatial and/or temporal structure of the observed process, there is a dependency in the data that should not be neglected. Additionally, hydrological models can be rather complex (e.g. non-linear behavior, a single model run might already be computationally demanding) and include several parameters that have to be calibrated. All of these characteristics increase the complexity of the numerical treatment of the calibration process. If Bayesian model calibration is applied to infer the parameters of the model based on available observations, often the problem is tackled by means of Markov Chain Monte Carlo (MCMC) methods (Gilks et al. 1996; Gelman et al. 2003).

In this contribution we present a different approach, termed BUS, that transforms the updating problem to a structural reliability problem and uses techniques from structural reliability to perform the updating (Straub and Papaioannou 2014). To demonstrate that BUS is a rather flexible method, we express the likelihood in terms of the measurement error and the modeling error, and learn not only the parameters of the model but also the correlation in the modeling error.

BAYESIAN MODEL CALIBRATION

Let $\mathcal{M}$ denote a probabilistic model class with uncertain model parameters $\boldsymbol{\theta} \in \mathbb{R}^N$. Model calibration aims at learning the uncertain parameters $\boldsymbol{\theta}$ based on observed behavior $\tilde{\mathbf{d}}$ of the system of interest. However, the parameters $\boldsymbol{\theta}$ cannot be determined explicitly for the following two reasons: First, the true behavior $\mathbf{d}$ of the system of interest is different from the observed behavior $\tilde{\mathbf{d}}$. Second, none of the models in $\mathcal{M}$ can represent reality perfectly, because nature is too complex to be described exactly by any model.

Considering the two points stated above, the engineer responsible for conducting the calibration can ask herself: How well can the observations $\tilde{\mathbf{d}}$ be explained by a certain set of parameters $\boldsymbol{\theta}$? This question can be formalized as the likelihood function $L(\boldsymbol{\theta}|\tilde{\mathbf{d}}) = f(\tilde{\mathbf{d}}|\boldsymbol{\theta})$, i.e.: What is the probability of observing $\tilde{\mathbf{d}}$ given some parameter set $\boldsymbol{\theta}$? Note that, due to our imperfect understanding of the real world, assumptions have to be made when formulating the likelihood. Consequently, the synthesis of model and likelihood function represents the belief of the engineer on how the system that is observed should be described within a Bayesian framework.

The available observations $\tilde{\mathbf{d}}$ represented through their likelihood $L(\boldsymbol{\theta}|\tilde{\mathbf{d}})$ can be used to learn the parameter vector $\boldsymbol{\theta}$ through Bayes' theorem as:

$$f(\boldsymbol{\theta}|\tilde{\mathbf{d}}) = \frac{1}{c_{\mathcal{M}}} L(\boldsymbol{\theta}|\tilde{\mathbf{d}})\, f(\boldsymbol{\theta}) \tag{1}$$

where $f(\boldsymbol{\theta}|\tilde{\mathbf{d}})$ is the posterior distribution of the parameters given the observations, $f(\boldsymbol{\theta})$ is the prior distribution of the parameters, and $c_{\mathcal{M}}$ acts as a scaling constant:

$$c_{\mathcal{M}} = \int_{\boldsymbol{\theta}} L(\boldsymbol{\theta}|\tilde{\mathbf{d}})\, f(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta} \tag{2}$$

The difficulty in Bayesian updating is that usually samples of $\boldsymbol{\theta}$ cannot be drawn directly from the posterior distribution $f(\boldsymbol{\theta}|\tilde{\mathbf{d}})$. Commonly, Markov chain Monte Carlo (MCMC) methods are used to sample from the posterior $f(\boldsymbol{\theta}|\tilde{\mathbf{d}})$ (Gilks et al. 1996; Gelman et al. 2003). The problem with MCMC methods is that the samples used after an initial burn-in phase may not have reached the stationary distribution of the Markov chain (Plummer et al. 2006). An alternative approach that applies structural reliability methods to draw samples from the posterior distribution was proposed by Straub and Papaioannou (2014). The advantage of this technique is that methods from structural reliability can be readily applied to perform Bayesian model calibration.

BUS: Bayesian updating using structural reliability methods

Let P be a standard uniform random variable defined on the interval $[0, 1]$. Combining P with the uncertain parameter vector $\boldsymbol{\theta}$ to be learned gives the augmented outcome space $[\boldsymbol{\theta}, p]$. Note that for the joint probability density function of $\{\boldsymbol{\theta}, p\}$ the following relation holds: $f(\boldsymbol{\theta}, p) = f(\boldsymbol{\theta}) \cdot f(p) = f(\boldsymbol{\theta})$, since $f(p) = 1$; therefore, $f(\boldsymbol{\theta}, p)$ and

$f(\boldsymbol{\theta})$ will be used interchangeably. Furthermore, let the domain Ω be defined as

$$\Omega = \{p \leq cL(\boldsymbol{\theta}|\tilde{\mathbf{d}})\} \tag{3}$$

where c is a positive constant such that $cL(\boldsymbol{\theta}|\tilde{\mathbf{d}}) \leq 1$ for all $\boldsymbol{\theta}$. Straub and Papaioannou (2014) showed that

$$\int_{p\in\Omega} f(\boldsymbol{\theta})\mathrm{d}p = c\, L(\boldsymbol{\theta}|\tilde{\mathbf{d}})\, f(\boldsymbol{\theta}) \tag{4}$$

and, consequently, Eq. (1) can be written as:

$$f(\boldsymbol{\theta}|\tilde{\mathbf{d}}) = \frac{\int_{p\in\Omega} f(\boldsymbol{\theta})\mathrm{d}p}{\int_{[\boldsymbol{\theta},p]\in\Omega} f(\boldsymbol{\theta})\mathrm{d}p\,\mathrm{d}\boldsymbol{\theta}} \tag{5}$$

The denominator in Eq. (5), denoted $p_\mathrm{a} \in (0;1]$, constitutes a structural reliability problem with limit-state function $g(\boldsymbol{\theta}, p) = p - cL(\boldsymbol{\theta}|\tilde{\mathbf{d}})$ (Straub 2011). The scaling constant $c_\mathcal{M}$ defined in Eq. (2) is linked to p_a through $c_\mathcal{M} = p_\mathrm{a}/c$. The limit-state function is defined such that if the outcome $[\boldsymbol{\theta}, p]$ is in Ω, then $g(\boldsymbol{\theta}, p) \leq 0$ will hold; and if the outcome is not in Ω, then $g(\boldsymbol{\theta}, p) > 0$. The important point is that samples generated from the prior distribution $f(\boldsymbol{\theta})$ will be distributed according to the posterior $f(\boldsymbol{\theta}|\tilde{\mathbf{d}})$ if they fall into the domain Ω (Straub and Papaioannou 2014).

The simplest application of this idea is the rejection sampling algorithm (Smith and Gelfand 1992, Straub and Papaioannou 2014) for drawing K samples from the posterior distribution:

1. Set counter $k = 1$.
2. Propose a sample $[\boldsymbol{\theta}^{(k)}, p^{(k)}]$:
   (a) Draw $\boldsymbol{\theta}^{(k)}$ from the prior distribution $f(\boldsymbol{\theta})$.
   (b) Draw $p^{(k)}$ from the standard uniform distribution defined on $[0, 1]$.
3. If $g(\boldsymbol{\theta}^{(k)}, p^{(k)}) \leq 0$:
   (a) Accept the proposed sample $[\boldsymbol{\theta}^{(k)}, p^{(k)}]$.
   (b) Increase the counter $k = k + 1$.
4. Go to step 2 as long as $k < K$.

The above algorithm is equivalent to applying Monte Carlo simulation for solving a structural reliability problem with limit-state function $g(\boldsymbol{\theta}, p)$. The simulation continues until K failure events are observed. On average, the procedure needs to be repeated K/p_a times, where p_a is the probability of accepting a proposed sample; i.e. $p_a = \Pr[g(\boldsymbol{\theta}^{(k)}, p^{(k)}) \leq 0]$. If the posterior distribution does not match the prior distribution well, p_a becomes small and renders the algorithm inefficient. However, as Straub and Papaioannou (2014) pointed out, other structural reliability methods can be used instead of the simple rejection sampling algorithm.

In this paper, subset simulation (SuS) is applied to perform the reliability analysis. SuS was proposed by Au and Beck (2001) and is an adaptive Monte Carlo method that is efficient for estimating small probabilities in high dimensional problems. The domain

Ω is expressed as the intersection of M intermediate nested domains Z_i, where $Z_0 \subset Z_1 \subset Z_2 \subset \ldots \subset Z_M = \Omega$. The domains Z_i are defined as the sets $\{g(\boldsymbol{\theta}, p) \leq b_i\}$, where $b_0 = \infty > b_1 > b_2 > \ldots > b_M = 0$ holds. SuS starts with drawing N_s random samples from the prior distribution $f(\boldsymbol{\theta})$. These samples are realizations from Z_0. The scalar b_{i+1} is picked as the p_t-percentile from the list of values of $g(\boldsymbol{\theta}, p)$ that belong to the N_s realizations of Z_i, where p_t is usually chosen equal to 10% in each step. If the p_t-percentile is negative, b_{i+1} is set to zero and p_t is set to the fraction of negative values in the list. Consequently, $p_t \cdot N_s$ realizations of Z_i are also realizations of Z_{i+1}. In order to obtain N_s realizations of Z_{i+1}, Markov chain Monte Carlo (MCMC) methods are used to generate $(1-p_t)\cdot N_s$ additional samples of Z_{i+1}, where the $p_t \cdot N_s$ original realizations of Z_{i+1} are used as seed values for the Markov chains. The described iterative process is continued until $b_i = b_M = 0$. Note that samples from the domain Z_M follow the posterior distribution. Let K denote the number of samples to be generated from the posterior distribution. If $K > N_s$, the MCMC sampling at the Mth iteration step is continued until K (and not only N_s) realizations from the domain Z_M are available. For a more detailed description of the sampling procedure, the reader is referred to Straub and Papaioannou (2014).

APPLICATION TO THE SIMPLE HYDROLOGICAL *abc* MODEL

We apply Bayesian updating by means of structural reliability methods (BUS) to a hypothetical hydrological example, inspired by the model discussed in Beven et al. (2008). Our aim is to infer the parameters of the simple hydrological *abc* model by means of Bayesian model calibration. We assume that measurements of past rainfall and discharge are available, where the measurements are not exact due to measurement errors. Moreover, the true discharge is generated by a model different from the *abc* model.

The *abc* model is introduced in the next section. Thereafter, we discuss the hypothetical real world case, i.e. the model that was used to generate the rainfall data, the discharge data and the measurement errors. Using hypothetical data has the advantage that the computed model output can be compared to the fictitious truth. Following this section, we set up the Bayesian updating problem; i.e. we define the prior assumptions and the formulation of the likelihood function. Finally, the results obtained by means of the Bayesian model calibration are presented and discussed.

Assumed model to be calibrated

The *abc* model has three parameters (a, b and c) and two equations:

$$Q_t = (1 - a - b) \cdot r_t + c \cdot S_t \tag{6}$$

$$S_{t+1} = (1 - c) \cdot S_t + a \cdot r_t \tag{7}$$

where Q_t is the discharge at time-step t, r_t is the rainfall at t and S_t is an assumed storage of water at t. The parameter $a \in [0, 1]$ describes the proportion of rainfall

that enters the storage, $b \in [0,1]$ is the proportion of rainfall that is lost due to evapotranspiration, and c is the proportion of water seeping from the storage. When selecting a and b, the condition $a + b \leq 1$ must be maintained. Besides the three parameters, the initial storage S_1 at $t = 1$ has to be determined.

The hypothetical real world case

In this section we describe how the hypothetical real world was modeled. This includes generation of true rainfall, the errors attached to the rainfall and discharge measurements, and the hydrological model that evaluates the true discharge given the true rainfall.

The true rainfall was generated by a model similar to the one used in Mantovan and Todini (2006) and described in Beven et al. (2008): At each time-step, the probability that a rainfall event starts is 10%. If a rainfall event is triggered at time-step t, it lasts k_t time-steps, where k_t follows a Poisson distribution with a mean of five. Let $r_{t,i}$ denote the rainfall at time-step t due to a rainfall event that started at time-step $t - i$, where $i = 0, 1, \ldots, k_t - 1$ and k_t is the duration of the rainfall event triggered at time-step $t - i$. The value $r_{t,0}$ follows a Gamma distribution with a mean of 400 and a variance of 800. The $r_{t,i}$ with $i = 1, \ldots, k_t - 1$ follow a Gamma distribution with a conditioned mean of $0.35 \cdot r_{t+i-1,i-1}$ and a conditioned variance of $0.7 \cdot r_{t+i-1,i-1}$. The actual rainfall r_t at time-step t is the superposition of all rainfall events that are active at time-step t.

The true discharge was computed based on the true rainfall by means of the following model:

$$Q_t = (1 - a - b) \cdot r_t + c \cdot S_t \tag{8}$$

$$S_{t+1} = (1 - c - d) \cdot S_t + a \cdot r_t + e \cdot R_t \tag{9}$$

$$R_{t+1} = (1 - e - f) \cdot R_t + d \cdot S_t \tag{10}$$

where S and R are two storages, parameters a, b and c are equivalent to the ones used in the simple *abc* model, parameter d is the proportion of water that goes from storage S to storage R, e is the proportion of water that goes from storage R to storage S, and f is the proportion of water that is leaking from storage R to underground flow. The parameters that represent the real world were fixed as: $a = 0.3$, $b = 0.65$, $c = 0.06$, $d = 0.02$, $e = 0.025$, $f = 0.02$, $S_1 = 200$, $R_1 = 200$. Note that the hypothetical real world model is run for 200 time-steps before the recording of rainfall and discharge measurements begins, to diminish the influence of the initial condition.

The rainfall measurement errors are generated as follows: between the measured rainfall $\tilde{r}_t$ and the true rainfall r_t, the relation $\ln(\tilde{r}_t) = \ln(r_t) + \varepsilon_{\mathrm{r},t}$ holds. The error term $\varepsilon_{\mathrm{r},t}$ is described by a moving-average model as $\varepsilon_{\mathrm{r},t} = \frac{1}{3}\sqrt{\ln(0.1^2 + 1)}\,(\xi_t + \xi_{t-1} + \xi_{t-2}) - \frac{1}{2}\ln(0.1^2 + 1)$, where the ξ_t are uncorrelated standard normal random variables. Note that $\exp(\varepsilon_{\mathrm{r},t})$ is a log-normal random variable with a mean value of 1.0 and a standard deviation of 0.1; the measurement error of the current time-step is correlated with the measurement errors of the two previous time-steps.

For the discharge measurement errors the following relation between measurement $\tilde{q}_t$ and true discharge q_t is assumed: $\tilde{q}_t = q_t \cdot (0.85 + 0.008 q_t - 0.00008 q_t^2) \cdot \varepsilon_{\mathrm{q},t}$, where the $\varepsilon_{\mathrm{q},t}$ are independent Gamma distributed random variables with a mean of 1.0 and a standard deviation of 0.02.

Formulation of the updating problem

The Bayesian updating problem is formulated assuming that the engineer who conducts the analysis does not have perfect knowledge of the physics that drive the real world. The prior distributions of a, b, c and S_0 are assumed as: b is Beta-distributed on $[0;1]$ with $\alpha = 4$ and $\beta = 3$, a is Beta-distributed on $[0;1-b]$ with $\alpha = 8$ and $\beta = 3$, c is Beta-distributed on $[0;1]$ with $\alpha = 1$ and $\beta = 10$, and S_0 is Gamma-distributed with a mean of 300 and a standard deviation of 100.

For calibrating the model, $N_t = 500$ consecutive rainfall and discharge measurements are available. The error in the rainfall measurements is modeled as $r_t = \tilde{r}_t \cdot \tilde{\varepsilon}_{\mathrm{r},t}$, where r_t is the true rainfall, $\tilde{r}_t$ is the measured rainfall, and $\tilde{\varepsilon}_{\mathrm{r},t}$ is a correlated log-normal distributed error term with a mean of 1 and a standard deviation of η. The standard deviation η is assumed uncertain; its prior is a log-normal distribution with a mean of 0.1 and a standard deviation of 0.1. The error term is modeled as $\tilde{\varepsilon}_{\mathrm{r},t} = \exp\left(\sqrt{\ln(\eta^2+1)} \cdot \chi_{\mathrm{r},t} - \frac{1}{2}\ln(\eta^2+1)\right)$, where the $\chi_{\mathrm{r},t}$ are correlated standard normal random variables. The $\chi_{\mathrm{r},t}$ are described by the following autoregressive model of order 1: $\chi_{\mathrm{r},t} = \varphi \cdot \chi_{\mathrm{r},t-1} + \zeta_t$, where ζ_t are uncorrelated normal random variables with zero mean and $\sqrt{1-\varphi^2}$ standard deviation. The value of $\varphi \in [0;1]$ is assumed uncertain with a uniform prior distribution.

Let $\boldsymbol{\theta}_{\mathcal{M}}$ be a vector that contains the 506 uncertain parameters $\{a; b; c; S_0; \eta; \varphi; \zeta_1, \ldots, \zeta_{500}\}$ of the model. Furthermore, let $\mathbf{G}_{\mathcal{M}}\left(\boldsymbol{\theta}_{\mathcal{M}}, \tilde{\mathbf{d}}_{\mathrm{r}}\right)$ be the vector of the discharge computed by the model, where $\tilde{\mathbf{d}}_{\mathrm{r}}$ is a vector that contains the observed rainfall. Assuming that the discharge is always larger than zero, the computed discharge is linked to the observed discharge $\tilde{\mathbf{d}}_{\mathrm{d}}$ through the equation $\tilde{\mathbf{d}}_{\mathrm{d}} = \mathbf{G}_{\mathcal{M}}\left(\boldsymbol{\theta}_{\mathcal{M}}, \tilde{\mathbf{d}}_{\mathrm{r}}\right) \cdot \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ is a multiplicative error that contains both measurement and modeling errors. Using a transformation based on the natural logarithm, the error $\boldsymbol{\varepsilon}$ can be written as the difference $\boldsymbol{\varepsilon}_{\ln} = \tilde{\mathbf{l}}_{\mathrm{d}} - \mathbf{l}_{\mathcal{M}}\left(\boldsymbol{\theta}_{\mathcal{M}}, \tilde{\mathbf{d}}_{\mathrm{r}}\right)$, where $\boldsymbol{\varepsilon}_{\ln}$, $\tilde{\mathbf{l}}_{\mathrm{d}}$, $\mathbf{l}_{\mathcal{M}}$ denote the log-transforms of $\boldsymbol{\varepsilon}$, $\tilde{\mathbf{d}}_{\mathrm{d}}$, $\mathbf{G}_{\mathcal{M}}$, respectively. The dependency on the modeling and measurement error can be stated explicitly by expressing $\boldsymbol{\varepsilon}_{\ln}$ as

$$\boldsymbol{\varepsilon}_{\ln} = \left(\tilde{\mathbf{l}}_{\mathrm{d}} - \mathbf{l}_{\mathrm{d}}\right) + \left(\mathbf{l}_{\mathrm{d}} - \mathbf{l}_{\mathcal{M}}\left(\boldsymbol{\theta}_{\mathcal{M}}, \tilde{\mathbf{d}}_{\mathrm{r}}\right)\right) = \boldsymbol{\varepsilon}_{\ln,\mathrm{m}} + \boldsymbol{\varepsilon}_{\ln,\mathcal{M}} \tag{11}$$

where $\mathbf{l}_{\mathrm{d}}$, $\boldsymbol{\varepsilon}_{\ln,\mathcal{M}}$ and $\boldsymbol{\varepsilon}_{\ln,\mathrm{m}}$ are the log-transforms of the unknown true discharge $\mathbf{d}_{\mathrm{d}}$, the modeling error and the measurement error of the discharge. Modeling the two errors

separately, the likelihood function can be expressed as:

$$L\left(\boldsymbol{\theta}_{\mathcal{M}}\middle|\tilde{\mathbf{l}}_{\mathrm{d}},\tilde{\mathbf{d}}_{\mathrm{r}}\right)=f\left(\tilde{\mathbf{l}}_{\mathrm{d}}\middle|\tilde{\mathbf{d}}_{\mathrm{r}},\boldsymbol{\theta}_{\mathcal{M}}\right)=\int f\left(\tilde{\mathbf{l}}_{\mathrm{d}},\mathbf{l}_{\mathrm{d}}\middle|\tilde{\mathbf{d}}_{\mathrm{r}},\boldsymbol{\theta}_{\mathcal{M}}\right)\mathrm{d}\mathbf{l}_{\mathrm{d}}$$

$$=\int f\left(\tilde{\mathbf{l}}_{\mathrm{d}}\middle|\mathbf{l}_{\mathrm{d}}\right)f\left(\mathbf{l}_{\mathrm{d}}\middle|\tilde{\mathbf{d}}_{\mathrm{r}},\boldsymbol{\theta}_{\mathcal{M}}\right)\mathrm{d}\mathbf{l}_{\mathrm{d}} \tag{12}$$

$$=\int f_{\boldsymbol{\varepsilon}_{\ln,\mathrm{m}}}\left(\tilde{\mathbf{l}}_{\mathrm{d}}-\mathbf{l}_{\mathrm{d}}\right)f_{\boldsymbol{\varepsilon}_{\ln,\mathcal{M}}}\left(\mathbf{l}_{\mathrm{d}}-\mathbf{l}_{\mathcal{M}}\left(\boldsymbol{\theta}_{\mathcal{M}},\tilde{\mathbf{d}}_{\mathrm{r}}\right)\right)\mathrm{d}\mathbf{l}_{\mathrm{d}} \tag{13}$$

where $f_{\boldsymbol{\varepsilon}_{\ln,\mathrm{m}}}$, $f_{\boldsymbol{\varepsilon}_{\ln,\mathcal{M}}}$ are the joint probability density functions of $\boldsymbol{\varepsilon}_{\ln,\mathrm{m}}$, $\boldsymbol{\varepsilon}_{\ln,\mathcal{M}}$. Here we implicitly assume that the modeling error and the measurement error are independent.

The measurement and the modeling error are formulated separately to obtain the predictive distribution of the unknown true discharge $\mathbf{d}_{\mathrm{d}}$ (or rather its log-transform $\mathbf{l}_{\mathrm{d}}$). Alternatively, the likelihood could also be expressed directly through the measurement error $f_{\boldsymbol{\varepsilon}_{\ln,\mathrm{m}}}$ if $\boldsymbol{\varepsilon}_{\ln,\mathcal{M}}$ is included in the vector of model parameters $\boldsymbol{\theta}_{\mathcal{M}}$. As a consequence, samples from the predictive distribution of $\mathbf{l}_{\mathrm{d}}$ would be directly available through posterior samples of $\boldsymbol{\theta}_{\mathcal{M}}$. However, such an approach renders the numerical treatment of the updating problem more demanding compared to formulating the likelihood according to Eq. (13), because the likelihood of the problem has a stronger peak. The disadvantage of expressing the likelihood through Eq. (13) is that the predictive distribution of $\mathbf{l}_{\mathrm{d}}$, denoted by $f_{\mathbf{l}_{\mathrm{d}}}$, is not readily available. The predictive distribution $f_{\mathbf{l}_{\mathrm{d}}}$ can be obtained as

$$f_{\mathbf{l}_{\mathrm{d}}}\left(\mathbf{l}_{\mathrm{d}}\middle|\tilde{\mathbf{l}}_{\mathrm{d}},\tilde{\mathbf{d}}_{\mathrm{r}}\right)=\int f\left(\mathbf{l}_{\mathrm{d}}\middle|\tilde{\mathbf{l}}_{\mathrm{d}},\tilde{\mathbf{d}}_{\mathrm{r}},\boldsymbol{\theta}_{\mathcal{M}}\right)f_{\boldsymbol{\theta}_{\mathcal{M}}}\left(\boldsymbol{\theta}_{\mathcal{M}}\middle|\tilde{\mathbf{l}}_{\mathrm{d}},\tilde{\mathbf{d}}_{\mathrm{r}}\right)\mathrm{d}\boldsymbol{\theta}_{\mathcal{M}} \tag{14}$$

where $f_{\boldsymbol{\theta}_{\mathcal{M}}}\left(\boldsymbol{\theta}_{\mathcal{M}}\middle|\tilde{\mathbf{l}}_{\mathrm{d}},\tilde{\mathbf{d}}_{\mathrm{r}}\right)$ is the posterior distribution of the model parameters, and the distribution of $f\left(\mathbf{l}_{\mathrm{d}}\middle|\tilde{\mathbf{l}}_{\mathrm{d}},\tilde{\mathbf{d}}_{\mathrm{r}},\boldsymbol{\theta}_{\mathcal{M}}\right)$ can be deduced from $f\left(\tilde{\mathbf{l}}_{\mathrm{d}},\mathbf{l}_{\mathrm{d}}\middle|\tilde{\mathbf{d}}_{\mathrm{r}},\boldsymbol{\theta}_{\mathcal{M}}\right)$ up to a proportionality constant for fixed $\tilde{\mathbf{l}}_{\mathrm{d}}$ and $\tilde{\mathbf{d}}_{\mathrm{r}}$. The predictive distribution $f_{\mathbf{l}_{\mathrm{d}}}$ expresses our posterior belief on how the log-transform of the unknown true discharge in the calibration phase $\mathbf{l}_{\mathrm{d}}$ is distributed.

Let the predictive distribution of the log-transform of the discharge in the prediction phase $\mathbf{l}_{\mathrm{d,p}}$ be denoted by $f_{\mathbf{l}_{\mathrm{d,p}}}$. Furthermore, let $\tilde{\mathbf{d}}_{\mathrm{r,p}}$ denote the vector of observed precipitation in the prediction period, and let $\boldsymbol{\theta}_{\mathrm{p}}$ be a vector that contains the parameters $\zeta_{501},\ldots,\zeta_{800}$ describing the uncertainties of the observed rainfall during the prediction step. Using the relations stated in Eq. (11), we can write $f\left(\mathbf{l}_{\mathrm{d,p}}\middle|\mathbf{l}_{\mathrm{d}},\tilde{\mathbf{l}}_{\mathrm{d}},\tilde{\mathbf{d}}_{\mathrm{r}},\tilde{\mathbf{d}}_{\mathrm{r,p}},\boldsymbol{\theta}_{\mathcal{M}},\boldsymbol{\theta}_{\mathrm{p}}\right)=f\left(\boldsymbol{\varepsilon}_{\ln,\mathcal{M},\mathrm{p}}\middle|\boldsymbol{\varepsilon}_{\ln,\mathcal{M}},\tilde{\mathbf{d}}_{\mathrm{r,p}},\boldsymbol{\theta}_{\mathrm{p}}\right)$, where $\boldsymbol{\varepsilon}_{\ln,\mathcal{M},\mathrm{p}}$ is the log-transform of the modeling error in the prediction phase. The probability distribution of $\mathbf{l}_{\mathrm{d,p}}$ can be expressed as:

$$f_{\mathbf{l}_{\mathrm{d,p}}}\left(\mathbf{l}_{\mathrm{d,p}}\middle|\tilde{\mathbf{l}}_{\mathrm{d}},\tilde{\mathbf{d}}_{\mathrm{r}},\tilde{\mathbf{d}}_{\mathrm{r,p}}\right)=\int\int\int f\left(\boldsymbol{\varepsilon}_{\ln,\mathcal{M},\mathrm{p}}\middle|\boldsymbol{\varepsilon}_{\ln,\mathcal{M}},\tilde{\mathbf{d}}_{\mathrm{r,p}},\boldsymbol{\theta}_{\mathrm{p}}\right)$$
$$\cdot f\left(\mathbf{l}_{\mathrm{d}}\middle|\tilde{\mathbf{l}}_{\mathrm{d}},\tilde{\mathbf{d}}_{\mathrm{r}},\boldsymbol{\theta}_{\mathcal{M}}\right)\cdot f_{\boldsymbol{\theta}_{\mathrm{p}}}\left(\boldsymbol{\theta}_{\mathrm{p}}\right)\cdot f_{\boldsymbol{\theta}_{\mathcal{M}}}\left(\boldsymbol{\theta}_{\mathcal{M}}\middle|\tilde{\mathbf{l}}_{\mathrm{d}},\tilde{\mathbf{d}}_{\mathrm{r}}\right)\mathrm{d}\boldsymbol{\theta}_{\mathrm{p}}\,\mathrm{d}\boldsymbol{\theta}_{\mathcal{M}}\,\mathrm{d}\mathbf{l}_{\mathrm{d}} \tag{15}$$

Table 1. Prior mean μ' and prior standard deviation σ', and posterior mean μ'' and posterior standard deviation σ'' of the driving parameters of the model.

Parameter	μ'	μ''	σ'	σ''
a	$3.1 \cdot 10^{-1}$	$2.6 \cdot 10^{-1}$	$1.4 \cdot 10^{-1}$	$2.1 \cdot 10^{-3}$
b	$5.7 \cdot 10^{-1}$	$6.8 \cdot 10^{-1}$	$1.7 \cdot 10^{-1}$	$2.4 \cdot 10^{-3}$
c	$9.1 \cdot 10^{-2}$	$6.8 \cdot 10^{-2}$	$8.3 \cdot 10^{-2}$	$5.7 \cdot 10^{-4}$
S_0	$3.0 \cdot 10^{2}$	$3.4 \cdot 10^{2}$	$1.0 \cdot 10^{2}$	$1.1 \cdot 10^{1}$
η	$1.0 \cdot 10^{-1}$	$6.7 \cdot 10^{-2}$	$1.0 \cdot 10^{-1}$	$1.3 \cdot 10^{-2}$
φ	$5.0 \cdot 10^{-1}$	$8.3 \cdot 10^{-1}$	$2.9 \cdot 10^{-1}$	$1.3 \cdot 10^{-1}$
$l_{\mathrm{c,nr}}$	$1.0 \cdot 10^{1}$	$1.4 \cdot 10^{1}$	$5.0 \cdot 10^{0}$	$7.0 \cdot 10^{-1}$

In this study $f_{\boldsymbol{\varepsilon}_{\ln,\mathrm{m}}}$ and $f_{\boldsymbol{\varepsilon}_{\ln,\mathcal{M}}}$ are chosen to have a multivariate normal distribution with zero mean. The distribution $f_{\boldsymbol{\varepsilon}_{\ln,\mathrm{m}}}$ is modeled with a uniform standard deviation of 0.03, and $f_{\boldsymbol{\varepsilon}_{\ln,\mathcal{M}}}$ is modeled with a uniform standard deviation of 0.08. The correlation structure of $f_{\boldsymbol{\varepsilon}_{\ln,\mathrm{m}}}$ is defined as $\rho_{\boldsymbol{\varepsilon}_{\ln,\mathrm{m}}}(i,j) = \exp\left(-(i-j)^2/2^2\right)$, where i, j are the indices of the corresponding entries in the random vector with distribution $f_{\boldsymbol{\varepsilon}_{\ln,\mathrm{m}}}$. For the correlation structure of $f_{\boldsymbol{\varepsilon}_{\ln,\mathcal{M}}}$ the following model is used: All time-steps are grouped into periods depending on whether it is raining or not. A period of no-rain starts when the observed rainfall is below 40, and the rain-period starts when the observed rainfall is above 40. Entries of a random vector that has distribution $f_{\boldsymbol{\varepsilon}_{\ln,\mathcal{M}}}$ with indices i, j are uncorrelated if i and j belong to different periods. If i and j are in the same period, the following exponential correlation structure is applied: $\rho_{\boldsymbol{\varepsilon}_{\ln,\mathcal{M}}}(i,j) = \exp\left(-|i-j|/l_{\mathrm{c}}\right)$, where $l_{\mathrm{c,r}} = 2$ for rain periods and $l_{\mathrm{c,nr}}$ for no-rain periods is assumed to be uncertain with a log-normal prior distribution that has a mean of 10 and a standard deviation of 5. As a consequence, $l_{\mathrm{c,nr}}$ has to be added to the parameter vector $\boldsymbol{\theta}_{\mathcal{M}}$.

Results and discussion

The length of the calibration phase is 500, the duration of the consecutive prediction phase is 300 time-steps. The true discharge, measured discharge and measured rainfall were obtained by means of the described hypothetical real world model. For the obtained observations, the updating problem was solved with the BUS approach; i.e. it was transformed to a structural reliability problem that has 508 random variables and was solved by means of subset simulation. $1 \cdot 10^4$ random realizations were drawn from the posterior distribution. These realizations were used to estimate the posterior mean and standard deviation of the parameters listed in Table 1. Note that the parameters a and b are strongly correlated: their prior correlation is -0.91, and their posterior correlation is -0.96. The posterior uncertainty of all parameters listed in Table 1 is considerably reduced compared to their prior uncertainties. However, since imperfect knowledge of the real world was assumed, the applied error model is only approximate. As a consequence, the estimate of the posterior parameter uncertainty obtained by means of

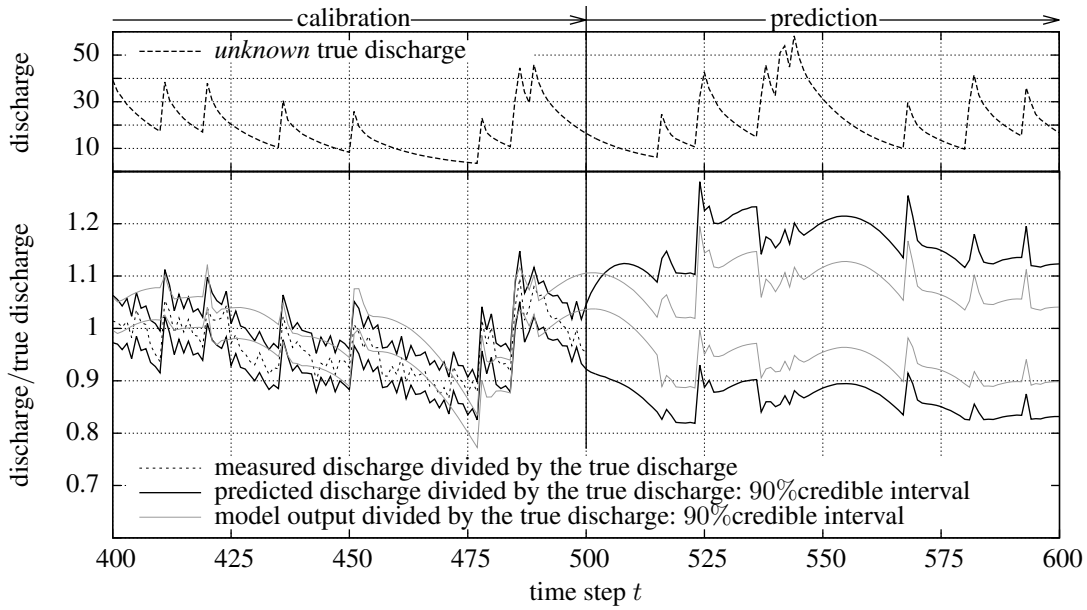

Figure 1. Credible intervals of the predicted discharge and the model output; transition from the calibration to the prediction phase at $t = 500$.

Bayesian model calibration might underestimate the true parameter uncertainty (Beven 2009); i.e. the Bayesian estimate of the model parameters tends to be overly confident.

However, if the focus of the analysis is not on parameter estimation but on prediction, this problem can be compensated by an explicit description of the modeling error. If the modeling error is neglected, the predictive distribution of the discharge is equal to the predictive distribution of the model output. If the modeling error is considered in the analysis, the predictive distribution of the discharge does not match the predictive distribution of the model output. This difference is illustrated in Fig. 1. The time-interval that was chosen for the comparison ranges from the 400th to the 600th time-step, and contains the transition from calibration to prediction occurring at the 500th time-step. In Fig. 1 the 90% credible intervals of the predictive distributions of discharge and the model output are plotted. During the calibration phase, the uncertainty in the predicted discharge is only slightly larger than the uncertainty in the model output. However, the predictive distribution of the discharge is drawn towards the measured discharge. At the end of the calibration phase and the beginning of the prediction phase, the uncertainty in the predictive distribution of the discharge increases considerably. After a transition phase that lasts for approximately 15 time steps, the modeling error acts as an amplifier of the uncertainty bounds of the model output. Therefore, even if one particular parameter set is clearly favored in Bayesian model calibration, i.e. the predicted model output becomes almost deterministic, the uncertainty in the predictive distribution of the discharge will not diminish. However, it should be noted that the underlying true modeling error is not known in hydrological modeling, because we lack a perfect understanding of how the real world behaves.

Therefore, the choice of the modeling error depends on the engineer conducting the analysis, and, consequently, represents his belief about the quality of the model.

CONCLUSION

The BUS approach was applied to infer the parameters of a hydrological model. We showed that the modeler is quite flexible in how to formulate the updating problem. Additionally, we highlighted that the modeling error should not be neglected when the uncertainty in the prediction is assessed.

REFERENCES

Au, S.-K. and Beck, J. L. (2001). "Estimation of small failure probabilities in high dimensions by subset simulation." *Probabilistic Engineering Mechanics*, 16(4), 263–277.

Beven, K. (2009). *Environmental modelling: an uncertain future?* Taylor & Francis.

Beven, K. J., Smith, P. J., and Freer, J. E. (2008). "So just why would a modeller choose to be incoherent?." *Journal of hydrology*, 354(1), 15–32.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian data analysis*. CRC press.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*, Vol. 2. CRC press.

Mantovan, P. and Todini, E. (2006). "Hydrological forecasting uncertainty assessment: Incoherence of the glue methodology." *Journal of Hydrology*, 330(1), 368–381.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). "CODA: Convergence diagnosis and output analysis for MCMC." *R news*, 6(1), 7–11.

Smith, A. F. and Gelfand, A. E. (1992). "Bayesian statistics without tears: a sampling–resampling perspective." *The American Statistician*, 46(2), 84–88.

Straub, D. (2011). "Reliability updating with equality information." *Probabilistic Engineering Mechanics*, 26(2), 254–258.

Straub, D. and Papaioannou, I. (2014). "Bayesian updating with structural reliability methods." *Journal of Engineering Mechanics (ASCE) (under review)*.