Extended Bayesian uncertainty analysis for distributed rainfall-runoff modelling: application to a small lower mountain range catchment in central Germany

Daniela Balin (1), Haytham Shbaita (1), Hyosang Lee (1), Michael Rode (1)

(1) Helmholtz Centre for Environmental Research UFZ, Germany, (daniela.balin@ufz.de)

Abstract

Bayesian approach is largely applied nowadays for estimating parameters and uncertainty of simple, lumped, conceptual hydrological models. Fewer studies applied instead these methods to estimate uncertainty of complex, distributed physically based hydrological models due to limitations of input data and time computing. This work presents the application of a Gibbs within Metropolis MCMC algorithm to estimate parameters and uncertainty of the WASIM-ETH model in a small catchment (100km^2) of Weisse Elster basin in Germany. Simulation has been done for the 200m grid size as current results showed that this is a good compromise between time computing and model efficiency. The use of a statistical likelihood function led to well identified posterior distributions of the parameters and further enabled computation of both parametric and predictive uncertainty of the simulated discharges. The model performed well in both calibration and validation periods, with reliable performance proven by small uncertainty bounds. More than 85% of the observed discharge values find themselves within the predictive uncertainty bounds while only 7% of these find themselves within the parametric uncertainty bounds. This is one of the limitations of the present method and could be an effect of not considering the errors in the input data in the calibration approach. As rainfall is one of the major inputs for hydrological modeling, rainfall measurement errors have been further considered for the Bayesian calibration approach. To the difference of the classical statistical approach of dealing with input uncertainty that considers rainfall data as unknown and then calibrates it together with the model parameters, in this study the rainfall uncertainty has been considered prior to the hydrological model calibration. For each of the five rainfall gages available on the study catchment, the systematic errors have been computed through a well-established procedure in Germany and random errors have been estimated through a Bayesian classical measurement error theory. 10 random rainfall scenarios have been sampled from the rainfall posterior distributions and for each scenario, the MCMC algorithm has been applied and the WASIM-ETH model posterior parameter distributions have been updated. These allowed updating the parametric and predictive posterior distributions of the simulated discharges. In comparison with the calibration case without considering input errors, the results showed improved coverage of the observed discharge values as a result of wider parametric uncertainty bounds (14%). Nevertheless, this improvement is still marginal and indicates that other sources of errors might be more important then point measurement errors such as spatial uncertainty and model structural errors.

Current work focuses on extending the Bayesian methodology in order to take into account the spatial input uncertainty into the calibration approach. In one hand this will allow obtaining more robust parameters and in the other more reliable model performances.

1. Introduction

Uncertainty in hydrological modelling has become an unavoidable topic nowadays in the hydrological research. Although generally only the parameter uncertainty was previously given attention, there is an increased awareness on the importance of other sources of uncertainty in the modelling approach. One of the most used methods to cope with uncertainty is GLUE (Beven and Freer 2001) that uses an informal criterion as likelihood measure and a Monte Carlo based sampling algorithm to compute parameter posterior distributions. Recently two Bayesian methods have been proposed to cope with different sources of uncertainty. One is called BATEA (Kavetski et al. 2006) and the other called IBUNE was introduced by Ajami et al. 2007. The first method was up to now applied only for simple conceptual methods and the second one uses a simplified model for input uncertainty and has been applied with semi-distributed models. The limit of these models is that the rainfall is considered as random and thus is calibrated together with the hydrological model parameters. The approaches adjust the observed rainfall in order to improve prediction of the simulated discharges and reduce its uncertainty. We consider that these approaches may be statistically feasible but we find difficult to defend this kind of approaches as for different models that use the same rainfall input adopting these methods we might end with different posterior distributions of the rainfall. In this paper, Section 2 the Bayesian methodology will be introduced, in Section 3 the study catchment, in Section 4 the main results and discussion will be included and in Section 5 the conclusions.

2. Bayesian methodology

The total uncertainty in rainfall-runoff conceptual modeling stems from several sources of error: input and response uncertainty (errors in measurement of input data such as rainfall, of response data such as water level and discharge or error in computing the potential evapo-transpiration), parameter uncertainty and model uncertainty arising from intrinsic inability of a given rainfall-runoff model to reproduce the physical mechanisms involved in runoff generation

(Montanari 2004). In this study, the uncertainty arising from the input forcing (rainfall), from the parameters (parametric uncertainty) and from a lumped term counting for other sources of uncertainty (predictive uncertainty) has been assessed MCMC Bayesian stochastic methods. At the difference of the classical statistical way of dealing with input uncertainty, the present extended Bayesian approach to deal with input uncertainty has two steps: an input measurement error Bayesian model and a hydrological Bayesian model calibration

2.1 Input measurement error Bayesian model

A classical multiplicative measurement error model has been build for the rainfall that includes both systematic errors and random errors. Errors of the classical type arise when a quantity (i.e. rainfall) is measured by some device (pluviometer) and repeated measurements vary around the true value like. We considered that the multiplicative error model is more appropriate for the rainfall as generally the measurement error increases proportionally with the measured rainfall. Since the multiplicative error is additive on the log-scale, all characteristics of the additive error are valid for the multiplicative error on the log-scale.

Systematic errors for the German Hellmann precipitation gauges in particular are well documented (Rode and Wenk 2006). Systematic error sources in measurement are included in Table 1. The wind effect has been studied and it was quantified in function of 4 shelter classes. For the study catchment we attributed the rain gauges to the class shelter C. The wetting and evaporation losses were added to the wind corrected data.

Source of losses	Year
a) Wind effects loss	12
b) Wetting collecting funnel loss	5.6
c) Wetting collecting can loss	0.3
d) Evaporation loss	0.8
Whole loss (%)	18.7

Table 1 Correction factors for Hellmann precipitation measurement of wind shelter class C in Germany, (from Rode and Wenk, 2006)

The rainfall posterior distribution was finally computed within the WINBUGS software and for each raingauge10 random rainfall scenarios have been computed. For each rainfall scenario the classical MCMC algorithm has been applied to compute posterior distributions of the hydrological parameters.

2.2 Bayesian method for parameter estimation and uncertainty

Bayesian methods use a probability model to fit a set of data and they summarize the results (estimated model parameters and predictions for new observations) by probability distributions. Bayesian theory is built on three probability concepts: the prior distribution, the likelihood function, the posterior distribution through the Bayes theorem.

The knowledge of the model parameters before using the observed data is given by the prior probability density function of the parameters. In this study, vague knowledge about the parameters has been represented by bounded uniform distributions.

The likelihood function summarizes all the information about the parameters available from the data. In this work, the likelihood function assumes that the simulation errors follow a normal distribution meaning that between the observed and predicted values the following relation can be stated:

$$Y = Q_{sim} + \varepsilon \text{, where } \varepsilon = N(0, \sigma^2) \tag{1}$$

As Q_{sim} is a function of observed input data (I) and model parameters (θ) one can say that

$$Y = f(I, \theta) + \varepsilon \tag{2}$$

The likelihood function for n observations, independent and identically distributed IID is given by the product of individual probability distributions:

$$L(\theta) = \frac{1}{\left(\sqrt{2\pi}\right)^n \cdot \sigma^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{t=1}^{n} \left(Y_t - Q_{sim\,t}\right)^2 \right) \tag{3}$$

The variance of the residuals σ^2 is considered as a new statistical parameter to be estimated together with the model hydrological parameters θ. The statistical likelihood function allows thus identifying two types of uncertainty: one associated with the hydrological parameters θ (parametric uncertainty) and the second associated with the variance of the residuals σ^2 (predictive uncertainty) that lumps all other sources of uncertainty.

In order to be able to use this likelihood function, the required statistical hypothesis such as constancy of the variance and time independence of the residuals (ε) should be respected. Exploratory analysis of the modelling errors is suitable and when important departures from these hypotheses are observed, it is important to correct the data by using different techniques.

The posterior distribution $p(\theta \mid y)$ represents the update of the prior distribution $p(\theta)$ with the likelihood function $p(y \mid \theta)$. The posterior distribution of the model parameters conditioned on the observed data is computed with Bayes' rule:

$$p(\theta \mid y) = \frac{p(y \mid \theta) \cdot p(\theta)}{p(y)} \tag{4}$$

p(y) can be considered as a probability of the evidence which is a constant and in this case *Eq.1* can be rewritten as:

$$p(\theta \mid y) \propto p(y \mid \theta) \cdot p(\theta);\ posterior \propto likelihood \cdot prior \tag{5}$$

The posterior distributions of the parameters can be computed either analytically or numerically. The methods presented in this paper use a numerical approach to sample from the posterior distributions of the parameters.

2.3 Metropolis-Hastings algorithm

The Metropolis algorithm is an iterative procedure that generates samples by using a Markov chain that converges to a given probability posterior distribution. The Metropolis algorithm includes three steps:

1) Generation of new parameters by jump specification step
The jump specification is needed to build a Markov chain in order to sample new candidates starting from the previous ones.

2) Acceptance /rejection of the new generated parameters step: Acceptance rule

This step is the central point of the Metropolis-Hastings algorithm. First, one has to compute the ratio of the posterior probabilities densities function between the last accepted and the candidate vectors of parameters:

$$r = \frac{p(\theta_{new} \mid Y)}{p(\theta_{old} \mid Y)} \cdot \frac{p(\theta_{old} \mid \theta_{new})}{p(\theta_{new} \mid \theta_{old})} \tag{6}$$

where θ_{new} and θ_{old} are the candidate and respective the last accepted vectors of parameters. The candidate vector of parameters is added or not to the previous Markov chain based on the Metropolis rule:

- if $r > 1$ than set $\theta^{i+1} = \theta_{new}$, the candidate set of parameters is accepted with probability 1.
- if $r < 1$ than u is generated randomly from the uniform distribution [0,1]:
    - if $r > u$ than set $\theta^{i+1} = \theta_{new}$, the candidate set of parameters is accepted,
    - otherwise set $\theta^{i+1} = \theta_{old}$ the candidate set of parameters is rejected and we keep the last vector set of parameters.

Gibbs within Metropolis-Hastings represents a special case of the Metropolis algorithm. This is called Gibbs within Metropolis-Hastings algorithm as we use both a multivariate normal jump specification for generating hydrological parameters, the Gibbs sampler in order to sample the residuals variance parameter and the Metropolis-Hastings rule as acceptance rule.

$$r = \frac{p(\theta_{new}, z_{new} \mid Y) \cdot p(\theta_{old}, z_{old} \mid \theta_{new}, z_{new})}{p(\theta_{old}, z_{old} \mid Y) \cdot p(\theta_{new}, z_{new} \mid \theta_{old}, z_{old})} \qquad \text{Eq.7}$$

where z = 1/σ.

3) Monitoring convergence of the algorithm

In this work, the Geweke test (Geweke 1992) has been applied in order to check for convergence. After removing the burn-in part, a thinning factor might be used in order to reduce the dependence of the remaining converged chains.

3. Study catchment and hydrological model

3.1 Study catchment

Weida catchment catchment is a sub-basin of the Saale catchment, which is one of the largest tributaries of the river Elbe in eastern Germany. The catchment is about 100 km^2 and the altitudes range between 270 m and to 650 m. Due to its location neighboring the low mountain ranges of Thüringer Wald and Harz, the catchment is sheltered from most rain and consequently receives an annual precipitation of 640 mm. The average annual temperature is around 7°C. There are two reservoirs closely located to this catchment: Zeulenroda reservoir, which is located to the outlet of the catchment and the Loessau reservoir, which is connected by pipes (as pointed as doted line in figure 1), used to provide a constant inflow (i.e. $30 m^3/sec$) into the catchment but which it is not that active anymore. 5 rainfall gauging stations (two inside the catchment and three near it) are used, and three climate stations, situated within 3 km of the catchment.

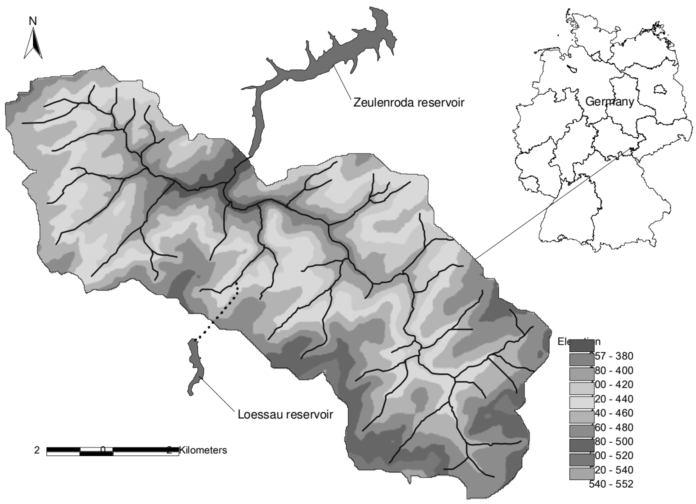

Fig. 1 Weida catchment- geographical localization

3.2 WaSiM-ETH – The Hydrological Model

The deterministic Water balance Simulation Model (WaSiM-ETH; Schulla and Jasper 2001) is chosen for this study as it has been successfully applied in similar mountainous catchments across Europe. The WaSiM-ETH is a process based distributed modelling system which can simulate different hydrological processes at spatial and temporal scales. The WaSiM-ETH uses a modular, object-oriented architecture for the simulation of different hydrological components such as evapotranspiration, snow accumulation, snow melt, infiltration and generation of surface and subsurface flow components. Spatial variability is represented in terms of orthogonal grid cells discretized over the surface of the basin. Spatial data required for the model include digital elevation model, land use and soil type. Precipitation and temperature are basic meteorological data required for the model, which can be supplemented by the climatic variables such as global radiation, relative sunshine duration, wind velocity, relative humidity and vapour pressure.

The WaSiM-ETH consists of an implementation of Green and Ampt/TOPMODEL approach for the simulation of runoff generation from infiltration excess and saturated areas. The Green and Ampt equation calculates infiltration based upon soil moisture conditions and surface runoff occurs when soil infiltration capacity has been exceeded. The TOPMODEL (Beven and Kirby 1979) is a conceptual variable contributing area approach based upon the distribution of saturation deficit. In this WaSiM-ETH implementation of TOPMODEL, soil water balance and runoff generation is simulated separately for each grid cell based on the spatial distribution of soil topographic index. The surface runoff is routed to the streams using a kinematic wave approach (Schulla and Jasper 2001). Five parameters have been calibrated: m the parameter controlling the rate of decline of transmissivity [m], T_{korr} the correction factor for the transmissivity of the soil [-], S_h the maximum content of the interflow storage [m], K_{korr} a scaling factor for the percolation, K_h a storage coefficient of interflow [h]. The model was calibrated on the period 1.Nov. 1999- 30 Oct. 2001.

4. Results and discussions

Figure 2 shows the posterior parameter of distributions without and with input uncertainty. One can see that the noisy rainfall data led to spreader posterior distributions of the hydrological parameters (Figure 2 down) while the statistical parameter, the residuals variance, remained less affected. The median of the posterior distributions has been left almost unchanged after introducing the noisy input data (Table 2).

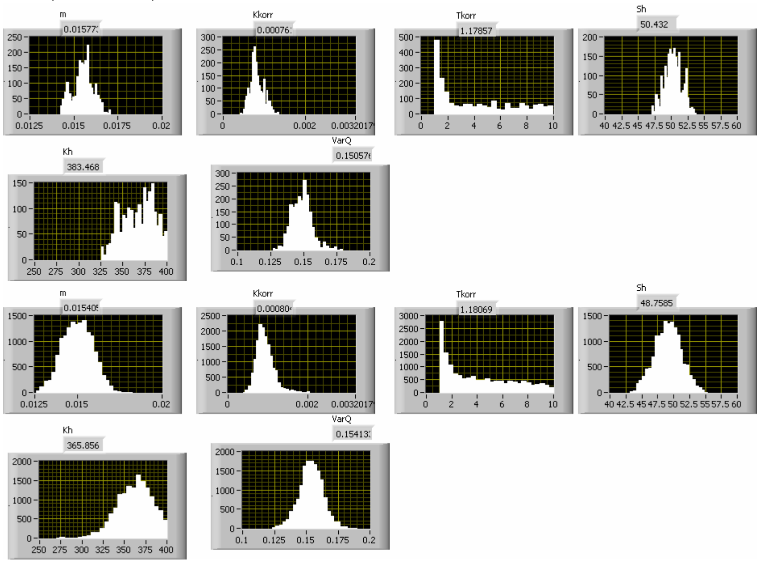

Fig. 2 Modes and posterior distributions of the hydrological parameters (m [*m*], Kkorr[-], Tcorr[-], Sh[m], Kh[h]) and statistical parameter (VarQ) for the cases without input uncertainty (top) and with input uncertainty (down)

Table 2 shows the median and the coefficient of variation of the posterior distributions of the hydrological and statistical parameter.

		m	Tkorr	Kkorr	Sh	Kh	Residuals variance
Without	Mode	0.0157	0.0008	1.17	50.43	383.46	0.15
	CV (%)	3.6	2.1	20.4	2.6	4.8	5.00
With	Mode	0.0154	0.0008	1.18	48.75	365.85	0.15
	CV (%)	5.6	2.6	22.3	4.3	5.5	6.00

Table . 2 Posterior distributions (mode, coefficient of variation-CV) of the hydrological and statistical parameter for the cases with and without input uncertainty

Comparison between the cases without and with uncertainty indicates that introducing input uncertainty led to greater coefficients of variation for all parameters of interest.

The results after propagating the parameter uncertainty in the simulated discharges can be seen in Figure 3.

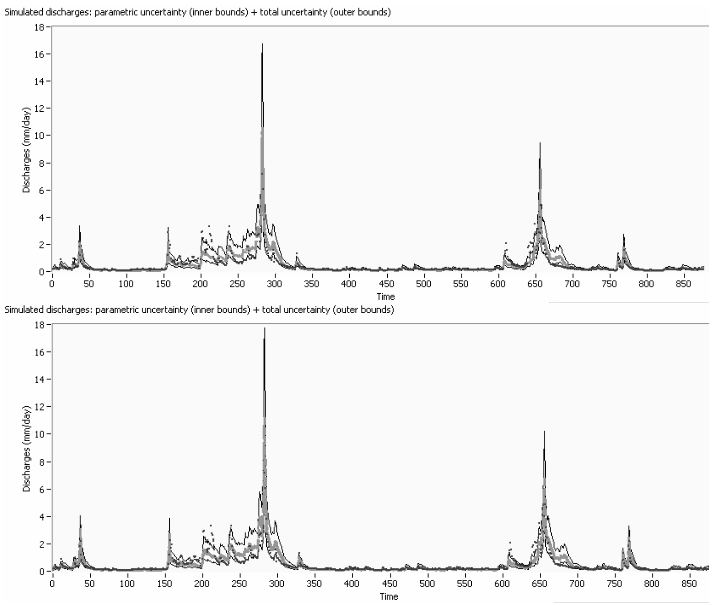

Fig. 3 Parametric (grey inner lines) and predictive (black outer lines) uncertainty bounds for the simulated discharges for the cases without input uncertainty (top) and with input uncertainty (down); dots – observed discharges.

Both plots in Figure 3 present two types of uncertainty bounds: the inner grey lines show the uncertainty bounds generated from the hydrological parameters while the outer black lines indicate the predictive uncertainty bounds after including the residuals variance, a lumped term counting for other sources of error: model structure and response measurement errors. At the difference of the parameters distributions, it is much more difficult to notice any important difference in representing the uncertainty of the simulated discharges between the two cases: without and with input uncertainty. Table3 shows the percentage of observed values covered by the inner parametric uncertainty bounds and the predictive ones.

	Parametric uncertainty	Predictive uncertainty
Without input uncertainty	7	82
With input uncertainty	14	84

Table 3 Percentage (%) of the observed points within the parametric and the predictive uncertainty bounds of the simulated discharges for the cases with and without input uncertainty.

One can see that if no noticeable difference is observed for the predictive uncertainty bounds that cover in both cases (with and without input uncertainty) more than 82% of the observed values, introducing the noisy input has led to an improvement as more than the double of observed points fall within the inner parametric uncertainty bounds (14%) than for the case with rainfall considered measured without error (7%).

Nevertheless, despite the fact that the parametric uncertainty bounds of the simulated discharges are wider after introducing the noisy rainfall, the improvement should be considered only a marginal one. Two possible reasons might explain this behavior. First, it might be possible that a compensation process occurs during the calibration approach with a rather complex physically based model. Second, this behavior might indicate that currently other more important sources of errors than the point measurement errors affect the modeling approach. Currently the second hypothesis is being tested by looking at the effect that the spatial uncertainty of the rainfall might have on the estimated uncertainty bounds of the simulated discharges. This will be done in two steps: first generating different rainfall interpolated fields based on the conditional simulation algorithm and second by applying the MCMC algorithm for each rainfall scenario generated previously and by updating the posterior distributions of the hydrological and statistical parameters.

5. Conclusions

The present paper presents the effects of input uncertainty of WASIM-ETH, a physically based hydrological model on the estimated parameters and the uncertainty of the simulated responses. This is done by first computing the rainfall uncertainty and second by propagating this uncertainty in the hydrological modeling step. While the posterior distributions of the hydrological parameters became wider after including the noisy input data, the effects on the simulated discharges have been much more limited. The first results indicate that although it would be expected that the rainfall input uncertainty might have an important influence on the uncertainty of the simulated discharges, this influence has been rather limited in the present case study. This might be due to other sources of errors that might be much more important for the modeling of the discharges than simply the point measurement error. The results presented here can further help identifying of the most important sources that affect the physically based distributed models and can contribute to obtain more realistic uncertainty bounds which is more suitable for water management schemes and for the decision making process.

6. References

Ajami N. K., Q. Duan, S. Sorooshian (2007), "An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction", Water Resources Research., 43, W01403.

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

Beven, K.J., and Kirkby, M.J. (1979). A physically based variable contributing area model of basin hydrology, Hydrological Sciences Bulletin 24 (1), 43–69.

Kavetski, D., G. Kuczera, S.W. Franks (2006).*"Bayesian analysis of input uncertainty in hydrological models, 1.Theory"* Water Resources Research 42(3), W03407

Kavetski, D., G. Kuczera, S.W. Franks (2006). *"Bayesian analysis of input uncertainty in hydrological models, 2. Application"* Water Resources Research 42(3), W03408

Geweke, J. (1992). "Evaluating the accuracy of sampling - based approaches to calculating posterior moments (with discussion)". Bayesian statistics 4. A. F. M. Smith. Oxford, Clarendon Press: 169-193.

Kuczera, G. and E. Parent (1998). "Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm." Journal of Hydrology 211(1-4): 69-85.

Montanari, A. (2004). *"Uncertainty assesssment in rainfall-runoff modelling: a review"*. Workshop: Statistical and Mathematical Methods for Hydrological analysis, Naples, Italy.

Rode, M. and Wenk, M. (2006). *"Weisse Elster River Basin Data Report"*, HarmoniRiB – EVK1–CT–2002-00109

Schulla, J. and Jasper, K. (2001). Model Description WaSiM-ETH. Internal report. ETH-Zurich.