COMPARISON OF METHODS FOR UNCERTAINTY ANALYSIS OF HYDROLOGIC MODELS

CHANGJUN CHEN
Zhejiang Design Institute of Water Conservancy and Hydroelectric Power,
Funingxiang66, Hangzhou, Zhejiang, 310002, China

DURGA LAL SHRESTHA, GERALD CORZO PEREZ, DIMITRI SOLOMATINE
UNESCO-IHE, PO Box 3015, 2601 DA
Delft, The Netherlands

The case study of a catchment in Nepal is applied to investigating and illustrating these three methods including Markov Chain Monte Carlo (MCMC), Monte Carlo simulation (MCS) and Latin Hypercube simulation (LHS). The parameter uncertainty calculated by using MCMC method is compared with estimates from LHS, MCS. The comparison between uncertainty in the model output only with parameter estimates and those with total error implies that the model uncertainty is more important than the parameter uncertainty based on the likelihood function used. The results also indicates that the MCS and LHS give very different estimates from the MCMC in this case, although the former two produce almost the same results with each other and are more efficient than the latter one.

INTRODUCTION

Hydrologic models usually include many parameters computed by calibration so that the model output can fit the observed to a great extent. Most people used local search optimization methods (Dawdy and O'Donnell, 1965; Nash and Sutcliffe, 1970; Johnston and Pilgrim, 1976; Pickup, 1977; Gupta and Sorooshian, 1985; Hendrickson et al., 1988), while others applied globally based optimization methods (Brazil and Krajewski, 1987; Wang, 1991). At the same time, many difficulties of finding a unique set of parameter have been reported (Johnston and Pilgrim, 1976; Pickup, 1977; Sorooshian and Gupta, 1983; Gan and Burges, 1990a,b; Duan et al., 1992; Sorooshian et al., 1993; Gan and Bifu, 1996). The poor identifiability of parameter may lead to considerable uncertainty in the model output, whereas the uncertainty analysis methods have begun to appear in the literature only recently. Examples of these approaches for quantification of parameter uncertainty include multi-normal approximation (Kuczera and Mroczkowski, 1998), evaluation of likelihood ratios (Beven and Binley, 1992), parametric bootstrapping and Markov Chain Monte Carlo (MCMC) methods (Tarantola, 1987; Kuczera and Parent, 1998; Engeland and Gottschalk, 2002; Vrugt et al., 2003; Engeland et al., 2005). Among them, MCMC has become very popular method for uncertainty analysis and the Metropolis Hastings algorithm has been used to estimate the parameters. In this paper,
the Bagmati catchment in Nepal is applied to estimating the model parameters from MH algorithm. The other two technique MCS and LHS were also used to give a comparative results and the total uncertainty was calculated using A Meta-Gaussian Model.

THE STUDY AREA AND HYDROLOGIC MODEL

The study area
The study area, Bagmati river basin is a medium sized river basin of Nepal with a catchment area of about 3700 km² up to the Nepal-India border (Figure 1). The mean daily data of hydrometric station lying on the Bagmati River at Pandheradobhan (Lat. 27° 06' 20" and long. 85° 28' 30") is used. In the upper Bagmati basin, the coverage of forest is minimum and occurs in the mountain tops only. The upper middle Bagmati basin has over 45% of the land area under forest cover and the lower middle Bagmati basin has the largest percentage (80%) of forest cover.

The tank model
Actually, there are many tank models proposed by different authors, here, a simpler model (called SIRT, Simplified IHE Rainfall-runoff tank model) consisting of two tanks with double storage and triple outflow is adopted for the present study. The schematic diagram is shown in figure 2. The physical processes are represented in an artificial way as a collection of vertical series of tanks. The upper storage represents the surface and soil storage whereas the lower storage represents the groundwater and slow runoff component. The lower storage can be interpreted as the ground water storage, which will not tend the model to run empty even during dry phases. The model has 8 unknown parameters that have to be identified through the process of calibration. These parameters can be divided according to the tank: tank A : $k_1$, $k_2$, $k_3$, $d_1$, $d_2$ and $s_1$; Tank B : $k_4$ and $s_2$.

SIMULATION METHODS

MCMC simulation
The Metropolis Hastings (MH) algorithm (Hastings, 1970), is applied to estimating the posterior parameter distributions. The likelihood used by Binley and Beven (1991) is chosen as likelihood function in this case, which is based on the sum of squares of the residuals:

$$L(\theta^{(s)}|Y) = \left(\sigma_e^2\right)^{-N}$$  \hspace{1cm} (1)

where N is a parameter chosen by the user. Here N=1000 were used to avoid too wide uncertainty intervals.

Based on the investigation done in Bagmati catchment in the past, the following parameter ranges of tank model shown in the table 1 and uniform priors between these bounds were applied. To ensure the Markov Chain achieve to convergence to a stationary posterior distribution as fast as possible, the amplitudes of step have to be tuned so that
acceptance rate should be about 45% (Chib and Greenberg, 1995). In this study, acceptance rates for some parameters have been tuned between 40% and 50% while those for others with near uniform distributions keep relative high acceptance rates.

![Study area and Bagmati river basin](image)

**Figure 1. Study area and Bagmati river basin**

![2-tank model](image)

**Figure 2. Schematic representation of a 2-tank model**

The parameters were estimated using MH algorithm and obtain a chain with 80,000 iterations. The first 20,000 iterations in the chain were discarded and the remaining 60,000 iterations were used as samples from the distributions. Visual inspection of the trajectories of the chain indicates that it had converged. 60,000 discharge values were obtained after running the tank model using these 60,000 samples from the parameter distributions. The 95% prediction uncertainty only with the posterior distribution of parameter estimates were estimated from these discharge samples. The 95% prediction uncertainty with total error was calculated using A Meta-Gaussian Model whose...
Monte Carlo simulation

For Monte Carlo methods, the uncertain model input is randomly generated according to predefined probability distributions. The assumption that parameters are normally distributed was made to define the probability distributions with mean from optimal value of calibration. To find the reasonable iterations are needed during the simulation, one of the most widely used forms of fitting criterion, the coefficient of efficiency, is introduced. Figure 3 shows that both standard deviation and mean of the efficiency converge to constant values after 3000 iterations, therefore, 3000 simulations are sufficient for the MCS to analyze model output uncertainty in this case.

The 3000 parameter sets were produced by randomly generated, then the tank model was implemented with these parameter sets to compute 3000 river discharge samples, consequently, the 95% prediction uncertainty only with parameter estimates were estimated from these 3000 discharge samples.

Latin Hypercube simulation

In the Latin Hypercube simulation, the same parameter probability distributions as Monte Carlo simulation are assumed and the coefficient of efficiency is also introduced. Figure 4 shows that both standard deviation and mean of the efficiency converge to constant values after 300 iterations, which means that 300 simulations are sufficient for the LHS.

Like CMS, the 300 parameter sets were created randomly, then the 95% prediction uncertainty only with parameter estimates were estimated from those discharge samples which have been calculated by running tank model using these 300 parameter sets created.

RESULTS AND DISCUSSION

The 95% prediction uncertainty for 2000 days was estimated by MCMC, MCS and LHS seperately. Only 100 days between 900th and 1000th is discussed. The 95% prediction uncertainty bounds with parameter estimates for 100 days simulated by (a) MCMC, (c)
MCS, (d) LHS and the 95% total uncertainty bounds calculated using A Meta-Gaussian Model (b) are shown in figure 5. From the Comparison between figure 5(a) and 5(b), it is easily to see the difference between parameters uncertainty and total uncertainty, which shows that the uncertainty originating from parameters can not account for the all the uncertainties. Figure 5(a) displays that the parameters uncertainty bounds are thin, however, for the total uncertainty, the bounds become much wider (see figure 5 (b)). This suggests that the model uncertainty is more important than the parameter uncertainty based on the likelihood function used here.

![Figure 3. The standard deviation and Mean of efficiency by MCS for various runs](image1)

![Figure 4. The standard deviation and Mean of efficiency by LHS for various runs](image2)

It can also be seen from the figure that only a few of points are within the uncertainty bounds estimated by MCMC because the uncertainty bounds are very narrow in this study, nevertheless, the result could be different when wider uncertainty bounds occur on the basis of other likelihood function. When larger parameter ranges are assumed, there will also be more observed points between the bounds, which happened in the figures 5(c) and 5(d) for MCS and LHS. In all, the parameter uncertainty is largely dependent on the likelihood and assumption of the parameter limits. Figure 5(c) and 5(d) show that
MCS and LHS produce almost the same output results, which implies that the LHS method, drawing samples over the full range of parameters distribution more consistently and evenly than MCS, is more efficient than MCS for predicting uncertainty bounds with parameter estimates considering 300 and 3000 parameter sets for them separately.

Figure 5. The 95% prediction uncertainty bounds with parameter estimates for 100 days simulated by (a) MCMC, (c) MCS, (d) LHS and the 95% total uncertainty bounds calculated using A Meta-Gaussian Model (b).

CONCLUSION

This study investigated the uncertainty of model output resulting from parameters based on three simulation techniques (MCMC, MCS and LHS). A chain with 80,000 iterations were calculated by MCMC to ensure enough points generated after it had converged to stationary posterior distribution, whereas only 300 and 3000 parameter sets were sufficient for MCS and LHS to analyze the uncertainty of model output. A Meta-Gaussian Model was also used to estimate the total uncertainty. Results imply that the model uncertainty is more important than the parameter uncertainty based on the likelihood function used. The uncertainty bounds of MCMC are very different from the others because the assumption of parameter distribution on which MCS and LHS are
based is different from posterior distribution obtained by MCMC in this case. Generally speaking, MCMC can get reasonable results than the other two, but lots of simulations are needed, which means that it will takes longer and even unacceptable time if there is a more complex model required to run, therefore, MCS and LHS could be reasonable choice for complex model if the parameter distributions are well understood.

REFERENCES


