Journal of Hydrology 540 (2016) 549–564

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Effect of formal and informal likelihood functions on uncertainty assessment in a single event rainfall-runoff model

Mahrouz Nourali [a], Bijan Ghahraman [b,*], Mohsen Pourreza-Bilondi [c], Kamran Davary [b]

[a] Department of Water Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, International Campus, Mashhad, Iran
[b] Department of Water Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran
[c] Department of Water Engineering, College of Agriculture, University of Birjand, Birjand, Iran

ARTICLE INFO

Article history:
Received 5 June 2016
Accepted 10 June 2016
Available online 17 June 2016
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Jozsef Szilagyi, Associate Editor

Keywords:
Uncertainty
DREAM$_{(ZS)}$algorithm
Formal/ Informal likelihood function
HEC-HMS
First-order autoregressive

ABSTRACT

In the present study, DREAM$_{(ZS)}$, Differential Evolution Adaptive Metropolis combined with both formal and informal likelihood functions, is used to investigate uncertainty of parameters of the HEC-HMS model in Tamar watershed, Golestan province, Iran.

In order to assess the uncertainty of 24 parameters used in HMS, three flood events were used to calibrate and one flood event was used to validate the posterior distributions. Moreover, performance of seven different likelihood functions (L1–L7) was assessed by means of DREAM$_{(ZS)}$approach. Four likelihood functions, L1–L4, Nash–Sutcliffe (NS) efficiency, Normalized absolute error (NAE), Index of agreement (IOA), and Chiew–McMahon efficiency (CM), is considered as informal, whereas remaining (L5–L7) is represented in formal category. L5 focuses on the relationship between the traditional least squares fitting and the Bayesian inference, and L6, is a hetereoscedastic maximum likelihood error (HMLE) estimator. Finally, in likelihood function L7, serial dependence of residual errors is accounted using a first-order autoregressive (AR) model of the residuals.

According to the results, sensitivities of the parameters strongly depend on the likelihood function, and vary for different likelihood functions. Most of the parameters were better defined by formal likelihood functions L5 and L7 and showed a high sensitivity to model performance. Posterior cumulative distributions corresponding to the informal likelihood functions L1, L2, L3, L4 and the formal likelihood function L6 are approximately the same for most of the sub-basins, and these likelihood functions depict almost a similar effect on sensitivity of parameters. 95% total prediction uncertainty bounds bracketed most of the observed data. Considering all the statistical indicators and criteria of uncertainty assessment, including RMSE, KGE, NS, P-factor and R-factor, results showed that DREAM$_{(ZS)}$ algorithm performed better under formal likelihood functions L5 and L7, but likelihood function L5 may result in biased and unreliable estimation of parameters due to violation of the residualerror assumptions. Thus, likelihood function L7 provides posterior distribution of model parameters credibly and therefore can be employed for further applications.

1. Introduction

During the past decades, conceptual rainfall-runoff models have been extensively used for watershed management policies and operational and research purposes. Uncertainty in model predictions are caused by the natural randomness, the measurement errors in input (forcing) and output data, the uncertainty in model parameters and the model structure (Blasone, 2007; Alazzy et al., 2015). Hydrologic models often include parameters that cannot be measured directly, so parameter estimation through calibration process is prone to error because the data, which were employed for calibration, generally contain measurement errors (Vrugt et al., 2003). Therefore, accurate calibration and uncertainty analysis is an important step for these models (Beven, 2006).

In order to estimate predictive uncertainty of the hydrologic models, infer the parameters, and predict model outputs, various methodologies may be adopted, including first-order approximation (Kool and Parker, 1988; Vrugt and Bouten, 2002), state-space filtering (Salamon and Feyen, 2009; DeChant and

* Corresponding author at: Department of Water Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran. Tel.: +98 513 8805709; fax: +98 513 8807145.

E-mail addresses: mahrouznourali@yahoo.com (M. Nourali), bijangh@um.ac.ir (B. Ghahraman), mohsen.pourreza@birjand.ac.ir (M. Pourreza-Bilondi), kamdav@um.ac.ir (K. Davary).

http://dx.doi.org/10.1016/j.jhydrol.2016.06.022

Moradkhani, 2012; Vrugt et al., 2013), multi model averaging (Ajami et al., 2007; Vrugt and Robinson, 2007), and various Bayesian approaches (Kavetski et al., 2006a,b; Kuczera et al., 2006; Reichert and Mieleitner, 2009; Renard et al., 2011; Rings et al., 2012; Vrugt et al., 2008, 2009b).

Among these approaches, Bayesian statistics have been widely used in hydrology for statistical inference of parameters and model output prediction (Kuczera and Parent, 1998; Bates and Campbell, 2001; Vrugt et al., 2003; Marshall et al., 2004; Liu and Gupta, 2007). Under Bayes theorem, posterior distribution combines the data likelihood with the prior distributions of parameters.

In majority of hydrological models, posterior distribution cannot be estimated by analytical approximation and, hence, simulation methods such as Markov chain Monte Carlo (MCMC) sampling can be adopted to implement Bayesian approach successfully. This method can efficiently estimate posterior probability density function (pdf) of the parameters.

MCMC methods are stochastic simulation algorithms that successively meet the solutions in parameter space, where solutions finally converge to posterior probability distributions. For any situation, different approaches of MCMC samplers may be considered by using suitable sampling or proposed distribution, while the convergence to the target posterior distribution is guaranteed (Vrugt et al., 2003; Blasone, 2007).

In hydrologic studies, in order to estimate parameter uncertainty of the hydrologic models, a suitable likelihood function has to be considered which provides reliable parameters of model. Formal or informal likelihood functions in Bayesian approaches have been used to estimate parameter uncertainty (Mantovan and Todini, 2006; Beven et al., 2008; Stedinger et al., 2008; McMillan and Clark, 2009; Vrugt et al., 2009b; Cheng et al., 2014). Formal likelihood functions are derived from an assumed statistical model for the residual errors (Box and Tiao, 1992). For example, the standard least squares (SLS) approach is used to derive the formal likelihood function under the assumptions that error residuals are uncorrelated (independent) and identically distributed by normal or Gaussian distribution with zero mean and constant variance (e.g. Vrugt et al., 2009b).

This approach is criticized, as it is highly depended on the assumptions of the residual error (Beven et al., 2008; Thyer et al., 2009), while in fact in many cases residual errors are correlated (dependent), nonstationary (heteroscedasticity), and non-Gaussian distributed (Kuczera, 1983). Revoking SLS assumptions may result in biased estimations of the parameters and affect either parameter or prediction uncertainties.

Informal likelihood functions are subjective likelihood probabilities and are not derived from a known model for the stochastic error series (Smith et al., 2008). For example Generalized Likelihood Uncertainty Estimation method (GLUE) (Beven and Binley, 1992), presented in the hydrologic literature, is often applied with a statistically informal likelihood function (Vrugt et al., 2009b). An informal approach may be used to estimate the uncertainty interval, where traditional error assumptions are violated. But this approach does not adhere to the formal statistical principles, and an informal likelihood function is not explicitly linked to an underlying error model (Schoups and Vrugt, 2010).

Choosing likelihood function requires a reasonable description of the distribution of the model errors in order to estimate the parameters, uncertainties and the statistical inferences accurately (He et al., 2010). If a likelihood function is arbitrarily applied that does not reasonably represent the distribution of the model errors, the results are unreliable.

When the assumptions of the residual error are violated, formal likelihood function must be applied based on a general error model. The general error model allows for the model bias and the correlation; nonstationarity (heteroscedasticity) and the nonnormality of the model residuals (e.g. Schoups and Vrugt, 2010). In addition, various methods may be used to relax common assumptions about residual errors, e.g. Box-Cox transformations to induce homoscedasticity (constant variance) and a first-order autoregressive (AR-1) scheme of the residuals to remove the temporal autocorrelation (e.g. Sorooshian and Dracup, 1980; Bates and Campbell, 2001).

Residuals of the rainfall–runoff models are often autocorrelated, because of the observed data and model structural uncertainties (Laloy et al., 2010). To account for the correlated errors, one common applied approach is using a first-order autoregressive (AR) scheme of the error residuals and considering the effect of model structural error (Vrugt et al., 2009b).

The hydrological modeling literature has mostly focused on the effect of choosing a likelihood function on the uncertainty analysis in the GLUE method and has showed that selection of likelihood function can directly affect the uncertainty analysis and the sensitivity of parameters (e.g. Freer et al., 1996; Stedinger et al., 2008; Freni et al., 2009; Alazzy et al., 2015).

Recently, a new Markov chain Monte Carlo (MCMC) sampler, namely $DREAM_{(ZS)}$ (DiffeRential Evolution Adaptive Metropolis algorithm), was used under a Bayesian framework as an efficient and robust sampler. Compared to the generalized likelihood uncertainty estimation (GLUE), the main advantage of DREAM (using MCMC simulation) is separating the effects of input (forcing), parameters and model structural uncertainties from total predictive uncertainty (Vrugt et al., 2009b).

$DREAM_{(ZS)}$ is based on the original DREAM algorithm (Vrugt et al., 2009a) that was modified for an efficient estimation of the posterior probability density function of parameters of a complex hydrologic model, high-dimensional posterior exploration problems.

Since results which are influenced by different likelihood functions, are important and considerable, this research demonstrates the importance and impact of choosing likelihood function on the parameter posterior distributions in a single event based rainfall-runoff model (HEC-HMS).

So, the influences of four informal likelihood functions and three formal likelihood functions were evaluated on estimating the parameters of HEC-HMS under $DREAM_{(ZS)}$ framework.

In this paper, study area is briefly described, and then the hydrologic model is presented. Afterwards, details of the procedures used to implement $DREAM_{(ZS)}$ method with different likelihood functions in the HEC-HMS hydrologic model are fully explained. Then the description of criteria is followed which are used to compare the effects of likelihood functions on the results of $DREAM_{(ZS)}$ method. Finally, the results and discussion are presented which are followed by a summary of the most important conclusions of this study.

2. Materials and methods

2.1. Case study and data

The study area is located in Gorganroud river basin, Golestan province, Iran, with an area of 3626.5 km^2 and is divided into three sub-basins, Tamar, Tangrah, and Galikesh, with areas of about 1530, 1724 and 372.5 km^2, respectively. In the study area, flash floods occasionally occur which cause some damages to lives, so flood control management plans are urgent in the basin. Annual rainfall varies between 200 and 850 mm in the basin (IWRI, 2008).

In the present study, Tamar basin was selected due to the availability of more reliable data of this basin. This basin is located between longitudes from 55°30′00″ to 56°04′37″E and latitudes from 37°24′49″ to 37°47′48″N (Fig. 1). The elevation of the basin ranges from 113 m at the basin outlet to about 2160 m at the