Parameter Optimization, Uncertainty Estimation and Sensitivity Analysis in Hydrological Modeling

Rajesh Vijaykumar Kherde and Priyadarshi H. Sawant

Abstract—This paper describes the application of Monte-Carlo simulations for parameter optimization, uncertainty estimation and sensitivity analysis using hydrological model developed by author [7] for Wardha River basin, Maharashtra, India. The Monte Carlo simulations revealed that the average values of parameters for the local optima of the calibration period seem to give good fit to the data and performance measure (NSE) does not differ significantly from the local optima of the respective calibration years. It is interesting to notice that, if the Monte Carlo simulations are carried out all over again, it generates yet another set of random numbers as realizations of model parameters. However, the model objective function (NSE) differs mere by 0.1% by running the new set of realizations and the local optimum parameter values are close to the earlier local optima. It seems that the model structure is in agreement with the “equifinality” or “non-uniqueness” concept as many different parameter sets give good fit to the data. However particular area of the parameter space is observed to be dominant in fitting the available observations.

I. INTRODUCTION

Beven and Binley [1] proposed Generalized Likelihood Uncertainty Estimation (GLUE) for comparing performance of different models. GLUE compares simulated hydrographs based on Nash-Sutcliffe efficiency. The likelihood measures like NSE, SSE, SLE, SAE are determined for each time step value and then normalized. The cumulative distribution function and confidence intervals are then derived using weighs as cumulative probabilities. A. Ghosh Bobba et. al [2] studied quantification of uncertainty in water quality model using two widely used methods viz. functional analysis and Monte Carlo simulations. Krzysztofowicz [8] applied Bayesian Forecasting system to recognize hydrological and precipitation uncertainty and produced probability distribution of rainfall-runoff model outputs. Bansidhar S. Giri et. al [4] studied impact of uncertainty in several model parameters using Monte Carlo simulations based on the assumption that uncertain parameters are uncorrelated and can be modelled by uniform, normal and lognormal probability distributions. L. Ma et. al [9] used a modified Monte Carlo sampling method for model realizations and obtained results of Root Zone Water Quality Model (RZWQM) output response sensitivity to selected model input parameters. A. Rahman et. al [12] presented a Manto Carlo simulation based on joint probability approach for theoretically superior method of design flood estimation. Dmitri Kavetski et. al [6] developed Bayesian total error analysis methodology for hydrological models. Jasper A. Vrugt et. al. [13] presented differential evolution adaptive Metropolis (DREAM) Markov Chain Monte Carlo sampler for estimation of posterior probability density function of hydrologic model parameters. A.J. Kalyanapu et. al. [5] demonstrated that single simulation flood risk approach underestimates flood risk. As the numbers of simulations are increased from 1 to 1000 flood risk increases considerably, thus Monte Carlo flood risk modelling framework has the ability to provide improved accuracy of flood risk. Dušan D. Ostojić et. al. [11] analyzed the accuracy of the point reliability assessment obtained using Monte Carlo simulation method depending on the sample size and number of iterations. James Charalambous et. al [3] applied Monte Carlo simulation technique to a large catchment and obtained more realistic design flood estimates than Design Event Approach. Daniel Marton and Stanislav Paseka [10] studied uncertainty impact of hydrological and operational input data on water management analysis of open reservoir.

II. METHODS AND TECHNIQUES

A. The Monte Carlo Method

Repeated random sampling is the basis of all computational algorithms that are classified as Monte Carlo methods and are used to obtain numerical results. In this technique the distribution of unknown probabilistic entity is obtained by running over the simulations many times. In physical and mathematical problems sometimes it is difficult or even impossible to obtain a closed form expression, or infeasible to apply deterministic algorithms. Under such circumstances Monte Carlo methods have been found useful and are typically being used for optimization, numerical integration and generation of draws from probability distribution.

In hydrology the results of Monte Carlo simulations are used for parameter optimization, uncertainty estimation and sensitivity analysis. Monte Carlo methods vary, but tend to follow a particular pattern:

1. Selecting imprecisely known model input parameters to be sampled.
2. Assigning ranges and probability distributions for each of these parameters.
3. Generating many sample sets (realizations) with random values of model parameters.
4. Running the model for all realizations to estimate uncertainty in model outcomes.

Published on November 29, 2018.
R. V. Kherde is with the Department of Civil Engineering, Dr. D Y Patil Institute of Engineering and Technology, Ambi, Pune, Maharashtra, India.
P. H. Sawant is with Sardar Patel college of Engineering, Andheri(W), Mumbai, Maharashtra, India.

DOI: http://dx.doi.org/10.24018/ejers.2018.3.11.907
B. Uncertainty Analysis Techniques

The sequential process of characterization of model input uncertainties using probability distribution and propagation of the input uncertainties through the system models to the model output is known as uncertainty analysis. In MCS technique the parameter probability distributions are specified and the model is run for all realizations to obtain the probability distributions of model outcomes. The reliability of the MCS uncertainty studies is recognized by many however the results may be greatly affected when (a) parameter uncertainty is not well defined (b) time marching models are computation intensive (c) outcomes of interest are limited in number.

It is usual to present Monte Carlo Simulation results in the form of CDFs (Cumulative density function) and/or histograms. CDF can be extracted at different time intervals for time marching models. Uncertainty with reference to time can also be represented by running 95% and 5% confidence intervals around the outcome of interest. Common output statistic measures used to represent uncertainty are the value at given probability level (say 95%) and the probability of expedience from targeted outcome. MCS offers extensive suppleness for propagation of uncertainty provided that the uncertain inputs have enough data available to define probability distributions. The model can then be run multiple times (say 1000-10,000) to ensure proper output distributions. In actual practice these two conditions seldom meet as (1) Scarcity of data results in making simplifying assumptions regarding input distributions (2) Restriction enforced by computational constraints to reduce number of simulations to mere few tens especially for distributed hydrologic modeling. Also MCS results may produce “overkill” solution in case probabilities of few outcomes of interest are required.

C. Sensitivity Analysis Techniques

Sensitivity analysis involve making small increments in the parameter value, one at a time, from reference value and estimating the corresponding change in the model output. The slope of input-output relationship is thus obtained by dividing change in output by change in input and is defined as sensitivity coefficient. Such analysis reflects the relative sensitivities of input parameters which is valid locally as the functional relationship between output and input of interest are seldom linear. Also the perturbations are applied to one parameter at a time and hence do not make a proper account of synergistic effects between model inputs.

Uncertainty analysis demands investigation into the input-output sensitivity over the full range of parameter variation and combinations. Such global sensitivity analysis techniques examine sensitivity of model results to the uncertainties and assumptions in model inputs. Global sensitivity analysis is useful in identifying key contributors to model output uncertainty, in determining key parameters controlling extreme model outcome and to detect the presence of non-monotonic input-output patterns.

The primary objective of sensitivity analysis is to identify whether the perturbation of parameter significantly affect the model response i.e. the variable of interest. In case it is observed that impact of particular parameter is small, the relevant parameters may be replaced by constants or eliminated altogether. This strategy will not only helps model construction but also model calibration on parameter estimation. In variable source area model the main aim of carrying out sensitivity analysis is to investigate sensitivity of simulated runoff to parameters that generate runoff. The peculiar interest is how these parameters changes with changes in soil, vegetation and climatic conditions.

III. APPLICATION OF MONTE CARLO METHOD IN WARDHA RIVER BASIN

A. Objective Function and Model Performance Indicators

The optimization of the parameters of a model requires the use of an objective function. Objective function is a reference numerical quantity enabling the calibration to be improved. The choice of objective function to be used for a given model is a subjective decision which influences the values of the parameters and the performance of the model.

The objective function for hydrologic simulations used in the present study is given below.

\[ NSE (\eta) = 1 - \frac{\sum_{i=1}^{n}(Q_i - q_i)^2}{\sum_{i=1}^{n}(Q_i - Q)^2} \]

Where,

- \( Q_i = \) Measured stream flow
- \( q_i = \) Simulated stream flow
- \( Q = \) Average measured stream flow
- \( n = \) Total number of time steps in calibration period

The value of NSE can be negative or positive, with a maximum absolute value of 1. A positive value indicate that the simulated values describe the trend of the observed data better than the mean of the observational values, while a negative value indicates that the corresponding model output is dis-similar to the behavior of the studied system. Obviously, a higher NSE value (close to 1) means a better fit of the predicted hydrograph to the observed one. Model simulations with negative NSE values are considered unacceptable. For best fit the other three criterions (performance indicators) should move close to zero. These criteria’s are,

- Sum of squared errors,
  \[ SSE = \sum_{i=1}^{n}(Q_i - q_i)^2 \]
- Sum of squared log error,
  \[ SLE = \sum_{i=1}^{n}[(\log(Q_i) - \log(q_i))]^2 \]
- Sum of absolute error,
  \[ SAE = \sum_{i=1}^{n}|Q_i - q_i| \]

B. Model calibration using Monte Carlo simulation (MCS) results

Model calibration using Monte Carlo simulation involved following steps:
1. Imprecisely known model input parameters (GZWC, SZWC, SZFC & SZPC) to be Sampled were selected.
2. The ranges and probability distribution for each of these parameters were assigned. Table I shows upper and lower bound values used for obtaining Monte Carlo realizations.
3. A total of ten thousand sample sets (realizations) with random values of model parameter were generated.
4. The model was run for all realizations of the calibration period and the Nash-Sutcliffe Efficiency and other performance indices were calculated for each of the calibration year.
5. The parameter set giving maximum efficiency for the calibration years (2002, 2003 & 2004) were extracted from MCS run.
6. The average value of the model parameters giving maximum efficiency for the three years of calibration period is taken as optimum parameter set.

**IV. RESULT AND DISCUSSION**

**A. The Dotty plots**

Figure 1 shows the dotty plots which represents various run of model for Monte Carlo realizations. They are basically scattering diagrams of parameter value verses some objective function value. While each dot represents one run of the model for different randomly chosen parameter value, they essentially represent projection of sample of points on goodness of fit response surface.

**B. Sensitivity Analysis Plots**

Figure 2-4 shows the parameter sensitivity plots depicting the cumulative distributions of parameter values group according to the ranking of each Monte Carlo run for a given likelihood measure or predicted variable. Strong differences between the cumulative distributions for a given likelihood measure or variable indicates sensitivity to that parameter. Distributions plotting close together indicate a lack of sensitivity.

First, for each parameter, the Monte Carlo simulations were ranked and equally divided into ten bins based on the value of the objective function. Thus the first bin contained the best 10% of the simulations, the second bin the next best 10%, and so forth. Next, the values of the objective function in each bin were normalized so they ranged from 0 to 1. Finally, these normalized objective function values were plotted as a cumulative distribution function of the parameter value. Thus, for each panel in Figure 2-4, there are ten curves, each corresponding to a single bin. In general, an insensitive parameter produced a straight one-to-one line whereas a sensitive model parameter showed differences in separation and form between the cumulative frequency distribution curves.

**C. Uncertainty Estimation**

For estimation of uncertainty associated with model parameters GLUE methodology (Beven and Binley 1992) was adopted. Appropriate prior parameter uncertainty distributions were assumed and the model and samples were then taken from these parameter distributions to generate Monte Carlo simulations. The likelihood values were then calculated based upon predefined likelihood measure (i.e., a measure of goodness of fit) to evaluate the degree of correspondence between each simulation and the observed system behavior. In this study Nash-Sutcliff efficiency (NSE) is taken as likelihood measure. The likelihood values were then used to determine whether a model structure-parameter set is “behavioral” or “non-behavioral” according to a subjectively defined threshold of likelihood values; and only behavioral model structure-parameter sets were retained to provide predictions of the system behavior. To assess the uncertainty associated with the predictions, weights of model parameters were calculated by normalizing the corresponding likelihood values so that all the weights sum up to one. The distribution of these weights is then taken as the probabilistic distribution of the predicted variables to reflect the uncertainty impacts of structural and parameter errors on model predictions.

Plots of the histogram and cumulative distribution of a simulated discharge, weighted by likelihood measure are shown in Fig. 5 and Fig. 6 respectively. The 5 and 95% sample quantiles are shown as red dots on the cumulative distribution plot.

**TABLE I: OPTIMIZED PARAMETER SET: RANGE AND ACTUAL VALUES**

| Sr.No | Parameter | Description          | Optimu m Value | Lower | Upper |
|-------|-----------|----------------------|----------------|-------|-------|
| 1     | GZWC      | Ground zone water content | 458            | 408   | 475   |
| 2     | SZWC      | Soil zone water content | 92             | 40    | 150   |
| 3     | SZFC      | Field capacity        | 230            | 150   | 350   |
| 4     | SZPC      | Pore capacity         | 359            | 350   | 450   |

**TABLE II: MODEL CALIBRATION USING MONTE CARLO SIMULATION RUN**

| Year  | Parameter Values | Calibration Period Efficiency | Validation Period Efficiency |
|-------|------------------|------------------------------|------------------------------|
|       | GZWC  | SZWC  | SZPC  | SZPC  | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
| 2002  | 462   | 97    | 228   | 354   | 0.9701 | 0.965 | 0.9637 | 0.9639 | 0.9392 | 0.9412 | 0.9271 |
| 2003  | 468   | 86    | 228   | 358   | 0.9325 | 0.9325 | 0.9356 | 0.954 | 0.9294 | 0.9372 | 0.9237 |
| 2004  | 442   | 92    | 233   | 366   | 0.9445 | 0.9638 | 0.9588 | 0.9593 | 0.9297 | 0.9423 | 0.9242 |
| AVG   | 458   | 92    | 230   | 359   | 0.9673 | 0.936 | 0.9597 | 0.9607 | 0.9326 | 0.9396 | 0.9271 |
Fig. 1(a). Dotty plots from Results of MCS run (GZWC & SZFC): calibration period 2002-2004

Fig. 1(b). Dotty plots from Results of MCS run (SZPC & SZWC): calibration period 2002-2004

DOI: http://dx.doi.org/10.24018/ejers.2018.3.11.907
Fig. 2(a). Parameter sensitivity plots GZWC & SZFC (2002)

Fig. 2(b). Parameter sensitivity plots SZWC & SZPC (2002)

Fig. 3. Parameter sensitivity plots (2003)
Monte Carlo simulations revealed that model structure is in agreement with the “equifinality” or “non-uniqueness” concept (Beven, 1993) and many different parameter sets give good fit to the data. But in contrast to TOPMODEL, it is not particularly impossible to find a unique “Best” parameter set whose performance measure differs significantly from other parameter sets. The Monte Carlo simulation results suggested that optimum parameter set for one period of observation seems to be optimum parameter set for another period, if not at least they are from the same parameter space. The average values of parameters for the local optima of the three years of calibration period seem to give good fit to the data and performance measure (NSE) does not differ significantly from the local optima of the respective calibration year.

It is interesting to notice that, if the Monte Carlo simulations are carried out all over again using the C# program developed in this study, the program generates yet another set of random numbers as realizations of model parameters. However, the model objective function (NSE) differs merely by 0.1% by running the new set of realizations and the local optimum parameter values are close to the earlier local optima. It seems that the model structure is in agreement with the “equifinality” or “nonuniqueness” concept (Beven, 1993) but particular area of the parameter space is observed to be dominant in fitting the available observations. As particular area of parameter space is dominant in fitting the available observations, obtaining the optimum parameter set by taking the average of parameter values for three years of calibration period is justified.

V. CONCLUSION

Fig. 4. Parameter sensitivity plots (2004)

Fig. 5. Histogram of predicted variable (Q_{sim})

Fig. 6. Cumulative distribution of a simulated discharge
The sensitivity analysis showed greater control on performance of the model by the parameters ‘SZWC’, ‘SZFC’ and to some extent by ‘SZPC’, while GZWC does not significantly affect the performance of the model. More information can be made available to catchment manager with respect to prediction error by including uncertainty in model parameters. The model output uncertainty is represented as probability distribution or as specific statistical quantity such as 95th percentile. The 95th percentile resulting from cumulative probability distribution represents the prediction of annual stream flow with 95% probability. By introducing such notions of confidence and probability informs policy developers about the degree of risk associated with particular actions.

REFERENCES
[1] Beven, K., & Binley, A. (1992). The future of distributed models: model calibration and uncertainty prediction. Hydrological processes, 6(3), 279-298.
[2] Bobba, A. G., Singh, V. P., & Bengtsson, L. (1996). Application of first-order and Monte Carlo analysis in watershed water quality models. Water resources management, 10(3), 219-240.
[3] Charalambous, J., Rahman, A., & Carroll, D. (2013). Application of Monte Carlo simulation technique to design flood estimation: a case study for North Johnstone River in Queensland, Australia. Water resources management, 27(11), 4099-4111.
[4] Giri, B. S., Karimi, I. A., & Ray, M. B. (2001). Modeling and Monte Carlo simulation of TCDD transport in a river. Water Research, 35(5), 1263-1279.
[5] Kalyanapu, A. J., Judi, D. R., McPherson, T. N., & Burian, S. J. (2012). Monte Carlo arlo-based flood modelling framework for estimating probability weighted flood risk. Journal of Flood Risk Management, 5(1), 37-48.
[6] Kavetski, D., Kuczera, G., & Franks, S. W. (2006). Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. Water Resources Research, 42(3).
[7] Kherde, R. V., & Sawant, P. H. (2014). A GIS Based Variable Source Area Model for Large-scale Basin Hydrology. Research Journal of Applied Sciences, Engineering and Technology, 7(18), 3797-3802.
[8] Krzyztofowicz, R. (1999). Bayesian theory of probabilistic forecasting via deterministic hydrologic model. Water Resources Research, 35(9), 2739-2750.
[9] Ma, L., Ascough Ii, J. C., Abuja, L. R., Shaffer, M. J., Hanson, J. D., & Royas, K. W. (2000). Root zone water quality model sensitivity analysis using Monte Carlo simulation. Transactions of the ASAE, 43(4), 883.
[10] Marton, D., & Paseka, S. (2017). Uncertainty Impact on Water Management Analysis of Open Water Reservoir. Environments, 4(1),10.
[11] Ostojić, D. D., Pokorni, S. J., Rakonjac, P. I., & Brkić, D. M. (2012). Accuracy of reliability calculated by the Monte Carlo simulation method. Vojnotehnički glasnik, 60(4).
[12] Rahman, A., Weinmann, P. E., Houng, T. M. T., & Laurenson, E. M. (2002). Monte Carlo simulation of flood frequency curves from rainfall. Journal of Hydrology, 256(3-4), 196-210.
[13] Vrugt, J. A., Ter Braak, C. J., Clark, M. P., Hynan, J. M., & Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. Water Resources Research, 44(12).

Dr. Rajesh V. Kherde has obtained undergraduate degree (BE-Civil) from SGB Amravati University, Maharashtra, Post Graduate degree (ME) from Gujrat University, Gujrat, and Ph.D. degree in Civil Engineering from Mumbai University, Maharashtra, India. He was awarded Gold Medal in post graduate studies by Gujrat Institute of Civil Engineers and architects. He is working as Professor & Head of civil engineering department at D.Y.Patil Institute of Engineering & Technology, Pune, Maharashtra, India. His area of specialization is water resources engineering, Geographical Information system (GIS), rainfall runoff and modelling and Application of Monte-Carlo simulations in water resources engineering.

Dr. Priyadarshi. H. Sawant is Principal at sardar Patel college of Engineering, Andheri(W), Mumbai, Maharashtra, India. He has obtained his undergraduate degree (BE-Civil) from Mumbai university, Post Graduate degree (M.Tech) from IIT Mumbai, and Ph.D. degree in Civil Engineering from IISC banglore, Karnataka, India. He has guided number of Ph.D. research scholars and M.Tech students. He has published number of research papers in international journals and conferences. His area of specialization is offshore structures, water resources management and construction management.

DOI: http://dx.doi.org/10.24018/ejers.2018.3.11.907