A simulation based approach for representation of rainfall uncertainty in conceptual rainfall runoff models

Shahadat Chowdhury and Ashish Sharma,
School of Civil and Environmental Engineering,
The University of New South Wales, Sydney NSW 2052 Australia

Abstract:

One of the common contributors to the uncertainty in any rainfall runoff model is the error distribution within the rainfall inputs. The uncertain rainfall introduces systematic bias in the estimated parameters. We present here the application of a method, known as simulation extrapolation (SIMEX), to ascertain the extent of parameter bias. SIMEX requires a knowledge of the standard error associated with the rainfall at any given time step. With this knowledge, it generates multiple sets of rainfall with artificially inflated error variance, parameters associated with whether this leads to any trend in the resulting parameters. This trend is then extrapolated back to assess the most suitable parameter value when the input is error free. The applicability of the method is investigated using a synthetic example where rainfall uncertainty is multiplicative and temporally invariant. This paper ascertained the bias trend in three key storage parameters of the Sacramento Rainfall Runoff Model representing surface and subsurface flow mechanisms respectively. This initial investigation confirmed the stability of SIMEX for use in hydrological model specification studies; which hints the possibility of embedding this simple method to improve runoff estimation.

KEYWORDS Parameter uncertainty; SIMEX; Sacramento, Rainfall-Runoff Model, Rainfall Error.

INTRODUCTION

Rainfall is the prime input variable of any rainfall runoff model. The availability and quality of rainfall data varies in time and across multiple locations of most catchments. Rainfall runoff models require prior transformation of these point rainfall time series to a sub-catchment scale. The transformation of point to aerial scale introduces input error to the rainfall runoff model. What is the effect of this input error on the resulting flows? Given known transformation error profiles, can we reduce the total uncertainty caused by this error on the simulated flows? This paper discusses a method to minimise the impact of rainfall error on flow simulations from any rainfall-runoff model.

Errors in input data introduce systematic bias in model parameterisation during calibration (Carroll et al., 1995; Chowdhury and Sharma, 2007). The bias in parameter estimation of rainfall runoff model due to uncertain rainfall has been studied by Kavetski et al. (2002). Later studies (Kavetski et al., 2006a; Kavetski et al., 2006b) used an orthogonal regression scheme termed as 'total least square' to mitigate the effect input error. An alternative functional estimator termed Simulation Extrapolation, or SIMEX in short, was introduced for use in hydro climatological studies by Chowdhury and Sharma (2007), that can mitigate parameter bias with fewer underlying assumptions. The Chowdhury and Sharma (2007) study was limited to hydrological applications in linear settings whereas this research letter investigates practical non linear extensions of the approach that include rainfall runoff modelling. More details on the rationale used are presented next.

METHODOLOGY

SIMEX, as the name implies, simulates parameter sensitivity to artificially inflated input error which in turn allows the extrapolation of parameters to an error free state. The method was originally developed in mid nineties (Cook and Stefanski, 1994; Stefanski and Cook, 1995) and more recently applied in hydrological area by Chowdhury and Sharma (2007). We briefly repeat the method here for the benefit of the readership, detail of which can be referred to publications mentioned above.

Consider an ideal case of a modelling response variable y in presence of the error free input variable x:

\[ y = G(\beta, x) + \varepsilon. \] (1)

Here, \( \beta \) is the true parameter of the regression \( y \sim x, G(\cdot) \) is an underlying model, and \( \varepsilon \) is the associated total error term. In practice we seldom know the true input variable (x) and need to set up our model using uncertain input variable, say \( \omega \). The input variable \( \omega \) results the naive parameter estimate \( \hat{\beta} \) of the regression \( y \sim \omega \) instead.

The SIMEX methodology requires a known error distribution for \( \omega \) with respect to x. Say, \( \omega \) includes an additive Gaussian error with variance \( \sigma^2 \) and mean zero. Synthetic input data \( \omega^* \) is generated adding artificially inflated error \( \lambda \sigma^2 \) to \( \omega \) where the inflation factor is \( \lambda > 0 \). Say the resulting biased parameter estimate of \( y \sim \omega^* \) is \( \beta^* \). The simulation is repeated for a number of inflation factors for example \( \lambda = 0.5, 1.0, 1.5... \). Hence a regression relationship of \( \beta^* \sim \lambda \) can be derived. The regression is extrapolated to \( \lambda = -1 \) to compute \( \beta_{\text{bias}} \). For a number of generalised linear models (Carroll et al. 1995), analytical solutions dictate that \( \beta_{\text{bias}} \approx \beta \).

RAINFALL RUNOFF MODEL PARAMETERS

The Rainfall runoff relationship can be highly non linear and difficult to model. While an analytical proof of SIMEX in simple linear settings may be viable, a more practical option for implementing the method in a nonlinear rainfall runoff modelling study is through a detailed numerical simulation. This section evaluates the utility of SIMEX in a practical hydrological setting using the Sacramento conceptual rainfall-runoff model.
The Sacramento Model

The Sacramento Model is a water balance model widely used in Australia (Boughton, 2005). Sacramento Model is also alternatively known as National Weather Service River Forecasting System (NWS-RFS) in USA (Burnash, 1975; Burnash et al., 1973). A general specification of the model is:

\[ Q_t = \text{Sac}(I_t, E_t; \theta) + \varepsilon_t \]  

(2)

where \( \text{Sac}(.) \) is the Sacramento model which uses rainfall and evaporation data \((I_t, E_t)\) to generate flow \(Q_t\) at time \(t\) with residual error \(\varepsilon_t\). The model parameter vector \(\theta\) consists of 16 parameters. The parameters define the following five major characteristics of the conceptual modelling system:

a) soil moisture storages,

b) rate of outflow,

c) percolation from upper to lower storages,

d) direct runoff from impervious areas, and

e) evapo-transpiration and deep seepage loss.

The Sacramento model has five soil moisture storages. The model essentially simulates water movements between storages, loss and routing as represented in Figure 1.

We use daily rainfall and evaporation at Golspie, NSW in Australia from 1980 to 1992 as a notional true estimate of catchment rainfall and evaporation, \((I_t, E_t; t = 1,2... (12 \times 365))\), the time series being shown in Figure 2. The average rainfall per wet day in the catchment equals 8 mm/day with a standard deviation of 9 mm. The fraction of wet days in the recorded period is 26%.

We generate flow based on a set of given values for all 16 model parameters. The resulting flow sequence, illustrated in Figure 2, is the synthetic true flow series \((Q_t; t = 1,2... (12 \times 365))\) used in the results presented below. This study monitored only three storage parameters with notional true values of \((60, 150, 38)\) mm. We appreciate that limiting the parameter number reduces the complexity of the Sacramento model. Nevertheless this significantly expedites computation without compromising the objective of demonstrating the application of our proposed approach in a rainfall-runoff modelling setting.

The next step in this synthetic study is to generate a corrupted rainfall data set. The scope for errors in rain records during drier times is low (a dry day reading is error free). Nevertheless a storm may completely miss the rain gauge. Multiplicative error in rainfall (Kavetski et al., 2006b) or a combination of multiplicative and a small additive error (Carpenter and Georgakakos, 2004) have been assumed as appropriate in prior studies. We assumed the rain record error to be multiplicative in this study. Accordingly we artificially corrupt the rainfall series \(I_t\) by multiplying it by a lognormal series \(U_t = \log N(1, \log 0.1^2)\) as in Equation (3).

\[ W_t = I_t U_t \]  

(3)

Three parameters are allowed to vary keeping the remaining 13 parameters constant. They are UZTW (upper zone tension water storage capacity), UZFW (upper zone free water storage capacity) and LZTW (lower zone tension water storage capacity). The relative roles of these parameters are outlined in the Sacramento Model conceptual representation in Figure 1. We now need an objective function prior to calibrating a rainfall runoff model.

The design of objective function depends on the purpose behind the formulation of hydrologic models. Robust hydrologic models require multiple objective functions (Khadam and Kaluarachchi, 2004; Madsen, 2000). Our objective function to calibrate the Sacramento Model is based on flow hydrograph and flow duration curve. Equal weight has been assigned to both criteria as follows.

\[ \text{Obj}(\theta) = \text{Argmin} \left\{ \sum (Q_{\text{est},i} - Q_i)^2 + \sum (Q_{\text{est},i} - Q_i)^2 \right\} \]  

(4)
A SIMULATION BASED APPROACH FOR REPRESENTATION OF RAINFALL UNCERTAINTY

Figure 3. The relationship of true daily rainfall to naive rainfall due to logN(1, log 0.1) error.

Figure 4. The empirical probability density function showing the affect of increase in the variance inflation factor (\(\lambda\)) on parameter estimates.

Figure 5. The biased parameter values are plotted against the multiples of error variance (\(\lambda\)). The point (\(\theta\)) \(\lambda = 0\) is the naive estimate and (\(\theta\)\(\lambda = -1\)) is the SIMEX corrected estimate (s).

Table I. Improvement in parameter values due to SIMEX.

| Parameters (mm) | True | Naive | SIMEX |
|----------------|------|-------|-------|
| UZTW           | 60   | 63    | 59.2  |
| LZTW           | 150  | 151.5 | 149.3 |
| UZFW           | 38   | 38.7  | 38.4  |

where, \(Q_t^\text{est}\) = observed and estimated flow at \(t^{th}\) percentile, the number of percentiles being set equal to the total number of time steps (\(t\)) in the sample. \(\text{Argmin}\) minimises the objective function.

A suitable optimisation scheme is required to efficiently solve Equation (4) and thus estimate \((\theta; p = 1, 2, 3)\). The choice of suitable optimisation schemes to solve non-linear hydrological models is intricate (Duan et al., 1992; Kuczera, 1997; Sorooshian et al., 1993; Yapo et al., 1998). We chose the L-BFGS-B optimisation algorithm (Byrd et al., 1995; Zhu et al., 1997). The L-BFGS-B is a limited-memory quasi-Newton algorithm for solving large non-linear optimization problems subject to simple bounds on the variables. It is intended for problems in which information on the Hessian matrix is difficult to obtain, or for large (or dense) problems.

The naive estimate \((\theta^{\text{naive}})\) is attained by solving the objective function Equation (4) where the rainfall is \(W_t\) in the Sacramento model as shown:

\[
Q_t^\text{est} = \text{Sac}(W_t, E_t; \theta^{\text{naive}} | \theta) \quad p = (1, 2, 3) \text{ and } q = (4, 5, \ldots 16)
\]

The rainfall data with error produces the following naive parameter estimates: \(\theta^{\text{naive}} = \{63, 151.5, 38.7\}\) mm. The naive estimate is a SIMEX step when \(\lambda = 0\). Note the difference between these and the true parameter values (60, 150, 38) mm. Can SIMEX help in modifying these naive estimates so as to offer a better representation of the truth?

Application of SIMEX

We generate replicates of the corrupted rainfall \(W^*_t\), with the addition of increasing levels of error \((\lambda, \sigma)\), thereby resulting in an increasingly biased parameter vector \(\theta^*_t\). In this example, we use variance inflation factors \(\lambda = \{0.25, 0.5, 1.0, 1.5\}\) with 300 random trials each. The expected value of \(\theta^*_t\) is estimated as the sample mean over the 300 trials. The distributions of the 300 trials of each parameter and interdependence among three parameters \((\theta = \theta^*_t, \theta^{*, 2}, \theta^{*, 3})\) are analysed below followed by SIMEX extrapolation.

The probability density functions of the sampled parameters (Figure 4), ascertained using kernel density estimation approaches, show the increase in variance in the parameter sampling distribution with increase in \(\lambda\). The effect of rainfall error is more significant for the upper storage parameter (UZTW), which is expected. The lower storage (LZTW) is less sensitive to input error due to the dampening effect of the prolonged accumulation involved. The three parameters exhibit near independence with respect to each other a feature that allows us to formulate separate relationships between \(\lambda\) and the estimated parameter value for each of the parameters considered.

Figure 5 illustrates the relationship between \((\lambda - \theta^*_t)\) for each \(p\). The presence of a clear trend in the relationship allows the formulation of the following regression relationship:

\[
\theta^*_t \sim G_r(\lambda)
\]

where \(G_r(.)\) denotes generic regression equations that may be linear or non-linear depending on the relationship that exists. In this example, a linear regression relationship is found to be satisfactory. The extrapolation to \(\lambda = -1\) gives us the SIMEX corrected estimate as shown in Figure 5 and listed (along with naive estimates and the true values) in Table I. Note that the SIMEX corrected estimates are closer to the true values compared to the naive estimates.
DISCUSSION AND CONCLUSIONS

Uncertainty in input variable results in systematic bias in parameter estimates. We ascertain that SIMEX helps to mitigate the bias of the Sacramento model parameters. The study demonstrated that parameters with higher influence on direct runoff are more sensitive to rainfall error. This research letter primarily aims to present the stability of SIMEX in a non linear rainfall runoff modelling setting. However the problem definition has been simplified in many respects which need further work. The three model parameters considered were independently related to the variance inflation further work. The three model parameters considered were independently related to the variance inflation further work. The three model parameters considered were independently related to the variance inflation further work.

SIMEX needs prior knowledge of the error distribution and any associated non stationarity in the rainfall time series. One of the future works will be the specification of error in rainfall data, which requires several considerations. The instrumentation error is often non stationary due to change in accuracy and method of measurements over time, which can be obtained from the manufacturers or experimentally quantified (Barry, 1978; Jaech, 1985). The translation of radar rainfall record to ground rainfall (Chumchean and Sharma, 2006) or downscaling of GCM simulations of rainfall and other atmospheric variables to rainfall at the local catchment scale (Mehrotra and Sharma, 2006) introduces uncertainty that may be possible to quantify and specify error distributions for use in a SIMEX or similar framework. The transformation of point rainfalls to aerial rainfall involves interpolation error which can be ascertained using available statistical methods (Kaplan et al., 1997). Alternatively non parametric SIMEX may be useful when multiple replicates (or possibilities) of rainfall scenario are available (Devanarayan and Stefanski, 2002).

Due to less onerous assumptions and computational burden, we envisage application of SIMEX to be a routine practice among practitioners to improve parameter estimation during any rainfall runoff model production.

ACKNOWLEDGEMENTS

We acknowledge the New South Wales Department of Water and Energy, Australia for rainfall and evaporation data. Funding for this research came from the Australian Research Council and the Sydney Catchment Authority. We recognize the contributors of the freely available statistical software R which has been used in this study (R Development Core Team, 2006).

REFERENCES

Barry BA. 1978. Errors in practical measurement in science, engineering, and technology. Wiley: New York; 183 pp.
Boughton W. 2005. Catchment water balance modelling in Australia 1960–2004. Agricultural Water Management 77: 91–116. DOI:10.1016/j.agwat.2005.01.012.
Burnash RJ. 1995. The NWS river forecast system-catchment modelling. Computer Models of Watershed Hydrology, V. P. Singh, Ed. Water Resources Publications: Colorado.
Burnash RJ, Ferral RL, McGuire RA. 1973. A Generalized Streamflow Simulation System-Conceptual Modelling for Digital Computers, Department of Water Resources: California.
Byrd RH, Lu P, Nocedal J, Zhu C. 1995. A limited memory algorithm for bound constrained optimization. Scientific Computing 16: 1190–1208.
Carperone TM, Georgakakos KP. 2004. Impacts of parametric and radar rainfall uncertainty on the ensemble streamflow simulations of a distributed hydrologic model. Journal of Hydrology 298: 202–221. DOI:10.1016/j.jhydrol.2004.03.036.
Carroll RJ, Ruppert D, Stefanski LA. 1995. Measurement Error in Nonlinear Models. Monographs on Statistics and Applied Probability 63, Chapman and Hall; 305 pp.
Chowdhury S, Sharma A. 2007. Mitigating Parameter Bias in Hydrological Modelling due to Uncertainty in Covariates. Journal of Hydrology 340: 197–204. DOI:10.1016/j.jhydrol.2007.04.010.
Chumchean S, Sharma A. 2006. An integrated approach to error correction for real-time radar-rainfall estimation. Journal of Atmospheric and Oceanic Technology 23: 67–79. DOI:10.1175/JTECH1892.
Cook JR, Stefanski LA. 1994. Simulation-Extrapolation Estimation in Parametric Measurement Error Models. Journal of the American Statistical Association 89: 1314–1328.
Devanarayan V, Stefanski LA. 2002. Empirical simulation extrapolation for measurement error models with replicate measurements. Statistics & Probability Letters 59: 219–225. DOI:10.1016/S0167-7152(02)00098-6.
Duan Q, Sorooshian S, Gupta V. 1992. Effective and efficient global optimization for conceptual runoff-models. Water Resources Research 28: 1015–1021.
Jaech JL. 1985. Statistical analysis of measurement errors. Wiley: New York; 293 pp.
Kaplan A, Kushnir Y, Cane MA, Blumenthal MB. 1997. Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures. Journal of Geophysical Research (C, Oceans) 102: 27835–27860.
Kavetski D, Franks S, Kuczera G. 2002. Confronting Input Uncertainty in Environmental Modelling. In Calibration of Watershed Models, Gupta HV, Sorooshian S, Rousseau AN, Turcotte R (eds); pp 49–68.
Kavetski D, Kuczera G, Franks S. 2006. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. Water Resources Research 42: W03407. DOI:10.1029/2005WR004368.
Kavetski D, Kuczera G, Franks S. 2006. Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. Water Resources Research 42: W03408. DOI:10.1029/2005WR004376.
Khadam IM, Kaluarachchi JJ. 2004. Use of soft information to describe the relative uncertainty of calibration data in hydrologic models. Water Resources Research 40: W11505. DOI:10.1029/2003WR002939.
Kuczera G. 1997. Efficient subspace probabilistic parameter optimization for catchment models. Water Resources Research 33: 177–188.
Madsen H. 2000. Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. Journal of Hydrology 235: 276 – 288. DOI:10.1016/S0022-1694(00)00279-1.
Mehrotra R, Sharma A. 2006. A nonparametric stochastic downscaling framework for daily rainfall at multiple locations. Journal of Geophysical Research–Atmospheres 111: D15101. DOI:10.1029/2005JD006837.
R Development Core Team. 2006. R Foundation for Statistical Computing: Vienna. ISBN 3-900051-07-0. URL http://www.R-project.org.
Sorooshian S, Duan Q, Gupta VK, Rousseau AN, Turcotte R. 1993. Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model. Water Resources Research 29: 1185–1194. DOI:10.1029/92WR02617.
Stefanski LA, Cook JR. 1995. Simulation-extrapolation: The measurement error jackknife. Journal of the American Statistical Association 90: 1247–1256.
Yapo PO, Gupta HV, Sorooshian S. 1998. Multi-objective global optimization for hydrologic models. Journal of Hydrology 204: 83–97.
Zhu C, Byrd RH, Lu P, Nocedal J. 1997. Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization. ACM Transactions on Mathematical Software 23: 550–556.