Bayesian Total Error Analysis For Hydrological Models: Preliminary Evaluation Using Multi-Site Catchment Rainfall Data

Thyer, M.A.[1], B. Renard[1], D. Kavetski[1], G. Kuczera [1] S. Srikanthan [2]

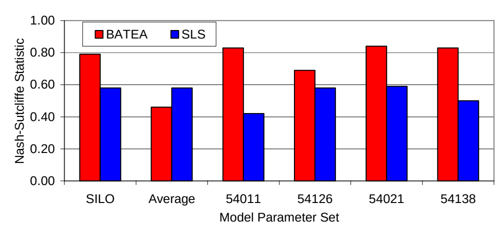
[1] School of Engineering, The University of Newcastle, NSW, Australia
[2]Bureau of Meteorology, Melbourne, Australia
Email: mark.thyer@newcastle.edu.au

Keywords: *Uncertainty, regionalisation, conceptual rainfall-runoff modes, input error, model error, model calibration*

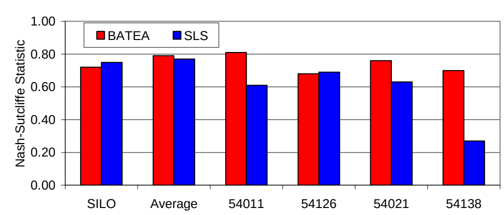
EXTENDED ABSTRACT

The Bayesian Total Error Analysis methodology (BATEA) provides the opportunity to directly address all sources of uncertainty (input, model and response error) in the calibration of conceptual rainfall-runoff (CRR) models. BATEA has the potential to overcome the parameter biases introduced by poor conceptualisations of these sources of errors and enhance regionalisation capabilities of hydrological models. This study is a preliminary evaluation of the robustness of the parameter estimates and the robustness in validation of the BATEA framework using multi-site catchment rainfall data. The aim was to compare how BATEA performed when provided with "degraded" rainfall from a single site compared to average rainfall from the entire catchment. The methodology used was to calibrate the same model to streamflow from the same catchment using six different rainfall time series; catchment average from the SILO gridded rainfall product, four individual gauges and the average of these four gauges. The catchment chosen was the Horton catchment, located west of the Great Dividing Range in Northern New South Wales, Australia. Markov chain Monte Carlo methods were used to compare the parameter estimates and their uncertainty using the BATEA and standard least squares (SLS) approaches for treating the sources of errors. It was found that the BATEA parameter estimates for the different rainfall time series were more consistent with each other, with average deviations from the overall average parameter value of the order of 0.5 to 1.2 times the parameter standard deviation. In comparison the SLS parameters estimates were more sensitive to the differences in the input rainfall data with average deviations from the overall average parameter value varying from 1 to 11 times the parameter standard deviation.

For validation, it was found that using the catchment average rainfall the BATEA and SLS parameters provided similar Nash-Sutcliffe (NS) statistics. However, it was found that catchment average rainfall does not necessarily provide the best streamflow predictions. Figure 1 compares the NS statistic for the 2 year validation period for the twelve parameter sets, which arise from using both BATEA and SLS approaches to calibrate to each of the six rainfall time series. Figure 1(a) shows NS using gauge 054021 as input, while Figure 1(b) shows the NS using the SILO rainfall as input. Gauge 054021 was chosen as it had the lowest BATEA rainfall error and was located in the high rainfall region of the catchment. In general, the BATEA parameter estimates with the 054021 rainfall provide higher NS statistics then SLS. This indicates that BATEA has the potential to utilise rainfall data from more productive areas of the catchment to enhance streamflow predictions. These results have several caveats, as they are based on a single catchment, and the validation period was relatively short, with the potential for biases. Nonetheless, these initial results are promising for the potential of BATEA to improve regionalisation. Future research will investigate the generality of these results with further case studies.



(a) Gauge 054021 rainfall



(b) SILO rainfall

Figure 1. Validation statistics for 12 parameter sets.

1. INTRODUCTION

Rigorous quantification of the impact of the various sources of uncertainty (input error, model structural error, response error) on the calibration of conceptual rainfall-runoff (CRR) models remains a challenging task in hydrological modelling. This has a number of implications not the least of which is that regionalisation of CRR model parameters continues to be confounded by biases in the calibrated parameters and unreliable assessment of parameter and predictive uncertainty.

In a recent review of CRR model calibration, Vrugt *et al*. (2005) note the lack of a robust framework that accounts for all sources of error Recently, several promising approaches have emerged in order to account for various sources of errors (Vrugt *et al*., 2005; Vrugt and Robinson, 2007). Among these is the Bayesian total error analysis (BATEA) methodology, proposed as a general framework to deal with the structural error of the model conceptualisation and measurement uncertainty in forcing/response data (Kavetski *et al*., 2006a,b; Kuczera *et al*., 2006). The core idea for BATEA implementation is to pose the CRR model and its error models as a Bayesian hierarchical model with latent variables describing errors in the data and the CRR model.

This paper examines the performance of BATEA when calibrating the same CRR model to the same streamflow data using different rainfall time series from multi-site rainfall data within a single catchment. These different rainfall time series will include two estimates of the catchment average rainfall. The aim is evaluate how BATEA performs when provided with "de-graded" rainfall data from only a single site compared to rainfall information from the entire catchment. The key issues to be investigated are the robustness of parameter estimates and the robustness in validation. This robustness will be assessed relative to the standard least squares (SLS) approach for parameter estimation. The discussion will focus on the implications of the results for regionalisation of CRR models.

2. OVERVIEW OF BATEA

Figure 2 provides a schematic of the hierarchical BATEA framework. This conceptualization provides a framework to deal with the various sources of input, model and response error. For simulation, the hydrologic time series is partitioned into n epochs $\{(t_i, t_{i+1}-1), i=1,..n\}$ where t_i is the time step index corresponding to the beginning of the i^{th} epoch. At the beginning of each epoch the stochastic parameters for the input error and model error components are sampled from their hyperdistributions, $p(\boldsymbol{\phi}\,|\boldsymbol{\alpha})$, and $p(\boldsymbol{\theta}|\boldsymbol{\beta})$ respectively. The parameters for the hierarchical BATEA are therefore the deterministic parameters, $\boldsymbol{\omega}$, the hyperparameters $\boldsymbol{\beta}$ for the stochastic model error component, the hyperparameters $\boldsymbol{\alpha}$ for the stochastic input error component, and the parameters for the response error component, $\boldsymbol{\gamma}$. For a complete description refer to Kuczera *et al*. (2006).

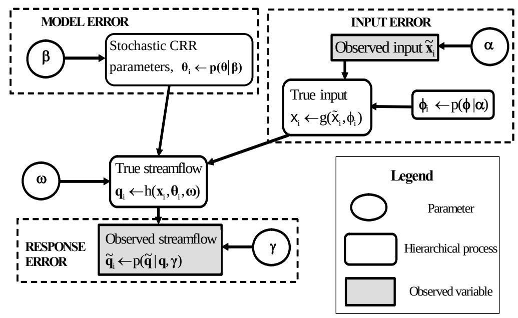


Figure 2. Hierarchical BATEA framework

The primary objective of the BATEA inference problem is to identify the parameters α, β, ω and γ given the observed streamflow time series data $\tilde{\mathbf{Q}} = \{\tilde{\mathbf{q}}_i, i = 1,..,n\}$, the observed forcing time series $\tilde{\mathbf{X}} = \{\tilde{\mathbf{x}}_i, i = 1,..,n\}$ and any prior information. In the Bayesian framework this inference problem is described by the posterior pdf

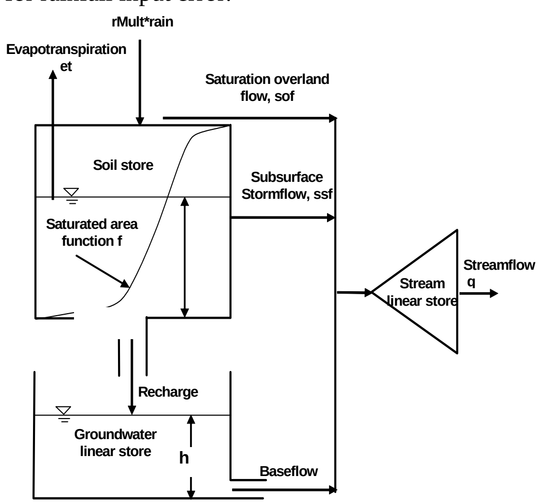
$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\omega}, \boldsymbol{\gamma} \,|\, \tilde{\mathbf{Q}}, \tilde{\mathbf{X}}) = \\ = \int p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\omega}, \boldsymbol{\gamma}, \boldsymbol{\theta}_{1:n}, \boldsymbol{\varphi}_{1:n} \,|\, \tilde{\mathbf{Q}}, \tilde{\mathbf{X}})\, d\boldsymbol{\theta}_{1:n}\, d\boldsymbol{\varphi}_{1:n} \qquad (1)$$

where $p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\omega}, \boldsymbol{\gamma}, \boldsymbol{\theta}_{1:n}, \boldsymbol{\varphi}_{1:n} \,|\, \tilde{\mathbf{Q}}, \tilde{\mathbf{X}})$ is the full posterior pdf, $\boldsymbol{\theta}_{1:n} = \{\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n\}$ contains the set of latent variables which represent the epoch-dependent CRR parameter realisations, and $\boldsymbol{\varphi}_{1:n} = \{\boldsymbol{\varphi}_1, ..., \boldsymbol{\varphi}_n\}$ is the set of latent variables which represent the epoch-dependent input errors for all storms.

Recent advances in the BATEA framework adopted in this paper, include the ability to evaluate the full posterior of the parameters for the BATEA using innovative MCMC methods Kuczera *et al*. (2007). In addition, Renard *et al*. (2007) assessed the sensitivity of the error models used in BATEA. The key results were that parameter estimates are more sensitive to the chosen temporal structure than the chosen hyperdistribution; Secondly, using input errors defined on a daily basis leads to more robust estimates than a storm epoch-based definition. In this paper the input errors, modelled as rainfall depth multipliers, are defined on a daily basis rather than a storm-epoch definition.

3. CONCEPTUAL RAINFALL-RUNOFF MODEL – LOGSPM

Figure 3 illustrates a typical CRR model, a member of the saturated path modelling (SPM) family (Kavetski *et al*., 2003), hereafter referred to as logSPM. The logSPM has 7 parameters (Table 1) and three stores operating at a daily time step. The seventh parameter, rMult, needs further comment. Kavetski *et al*. (2006a,b) use rainfall depth multipliers as an explicit (albeit approximate) representation of input uncertainty, which has the assumption that rainfall errors are multiplicative (i.e., $rain_{true}=rain_{obs}*rMult$). The same approach is used in this study to compensate for rainfall input error.



$$\frac{ds}{dt}=rMult*rain-quick(f)-ssf(f)-rge(f)-et(s) \text{.. surface store water balance}$$

$$f=\frac{1}{1+(sF+99)e^{-ks}}-\frac{1}{sF+100} \text{......saturation} - \text{soil depth function}$$

$$quick(f)=f*rMult*rain\text{..................quickflow}$$

$$ssf(f)=f*ssfMax\text{.........................subsurface stormflow}$$

$$rge(f)=f*rgeMax\text{........................groundwater recharge}$$

$$et(s)=pet*(1-e^{-s})\text{.........................actual evapotranpiration}$$

$$\frac{dh}{dt}=rge-bf(h)\text{............................ groundwater store water balance}$$

$$bf(h)=kBf*h\text{........baseflow}$$

$$\frac{dS}{dt}=quick+ssf+bf-q(S)\text{......... streamflom store water balance}$$

$$q(S)=kStream*S\text{.........................stream runoff}$$

Figure 3. LogSPM model conceptualisation.

4. CASE STUDY CATCHMENT

The catchment chosen for this case study was the Horton catchment, located in northern inland NSW, Australia (Figure 4). This catchment is one of the catchments compiled by Peel *et al*. (2000). Catchment area is 1920 km^2, with average annual rainfall of 819 mm and runoff coefficient of 0.13.

Table 1. Summary of logSPM parameters.

Parameter	Description
sK	Exponent for saturated area fraction
sF	Shift parameter controlling saturated area fraction
ssfMax	Subsurface stormflow at full saturation
rgeMax	Groundwater recharge rate at full saturation
kBF	Groundwater discharge constant
kStream	Stream discharge constant
rMult	Observed storm depth rainfall multiplier

To achieve the study's objectives model calibration was undertaken with six different rainfall time series. The SILO rainfall is the catchment average rainfall derived from gridded (5 x 5km) daily rainfall product produced by QLD Dept. of Natural Resources (Peel *et al.* 2000 provides details). The other four rainfall time series were individual gauges located within the catchment, which had a common 4 year period from 1976-1980. Table 2 lists the rainfall time series and includes an "average" times series based on the arithmetic average of the four individual gauges. The statistics show there is a strong rainfall gradient in the catchment, with higher rainfall in higher elevation south-western areas of the catchment.

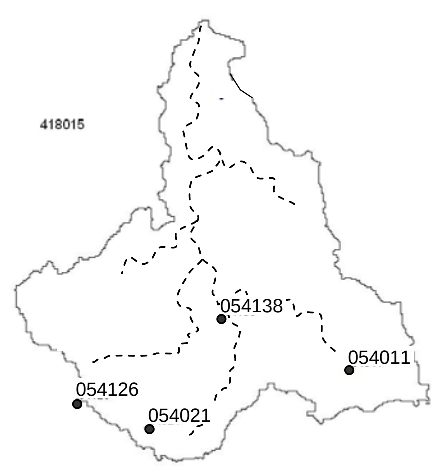


Figure 4. Horton catchment showing raingauge and approximate stream locations.

Table 2. Rainfall Data.

Rainfall Time Series	Elevation(m)	Avg. Daily Rainfall (mm)
SILO	-	1.51
Average	-	2.51
054011	567	1.83
054126	1465	3.40
054021	869	2.66
054138	392	2.15

5. MODEL CALIBRATION

For model calibration, MCMC methods (Kuczera *et al*., 2007) were used to evaluate posterior of the model parameters for both the BATEA and SLS approachs. The four year period was split into two two-year periods. As the second two year period had a larger number of streamflow events, this was chosen for calibration, while the first two-year period was chosen for validation. Using a 100 day warm-up the calibration period was 18/4/1978-7/4/1980, while validation was 25/9/1976-17/4/1978.

During initial BATEA calibration runs achieving convergence of the MCMC algorithm was found to be problematic when both input error and model error were inferred. This is possibly because the logSPM model is overparameterised with respect to the data (further discussed in Section 7). Hence, the parameterisation of model error as stochastic model parameters was not implemented and only deterministic model parameters were inferred. Two of the logSPM parameters (sF and kBF) were fixed at their SLS values. The uncertainty in the remaining parameters, sK, ssfMax, rgeMax, and kStream was inferred. For the input error model, the hyper-mean and hyper standard deviation (hyper-sd) of the stochastic rMult parameters was also inferred.

For the SLS calibration, the uncertainty of four (sK, ssfMax, rgeMax, and kStream) of the six model parameters were inferred, the other two were fixed at their SLS values. The uncertainty in the rMult parameter was also inferred, the difference from BATEA being that it was modelled as a deterministic rainfall multiplier – constant for the entire calibration period.

6. RESULTS & DISCUSSION

6.1. Parameter Estimates

For the parameter estimates, the general trend is that, for the four deterministic parameters (sK, ssfMax, rgeMax and kStream) the BATEA posteriors are reasonably consistent, identified by significant overlap in the probability limits for the different rainfall time series. In contrast the posteriors for the SLS parameters vary considerably for the different rainfall time series. Examples of this are shown in Figure 6. Table 3 quantifies this by providing the number of standard deviations the average parameter value for one rainfall time series is from the overall average parameter value based on all the rainfall times series. For SLS, the average deviation is between 1 and 11 standard deviations. While, for BATEA it is considerably lower, being between 0.5 to 1.4 standard deviations.

Table 3. Number of standard deviation the average parameter value for a certain rainfall time series is from the overall average parameter value.

Calibration Method	Rainfall Time Series	Parameter				
		sK	ssfMax	rgeMax	kStream	rMult hyper-mean
BATEA	SILO	0.09	1.67	1.09	0.96	3.93
	Avg	0.47	0.83	1.52	0.41	0.09
	54011	1.39	1.38	0.43	0.50	0.10
	54126	2.65	0.08	1.53	0.28	3.10
	54021	0.84	0.06	0.16	0.17	0.77
	54138	0.14	0.26	0.72	0.45	0.51
	Average Deviation	**0.93**	**0.71**	**0.91**	**0.46**	**1.42**
SLS	SILO	1.36	2.17	1.50	2.01	1.59
	Avg	3.06	2.39	2.70	1.67	29.37
	54011	3.37	1.98	0.06	0.38	5.41
	54126	8.78	0.12	3.80	2.08	15.82
	54021	1.45	5.41	7.27	0.18	13.05
	54138	0.67	3.07	15.38	0.07	5.77
	Average Deviation	**3.12**	**2.52**	**5.12**	**1.06**	**11.83**

Previous studies have shown using synthetic and real data that BATEA produces parameter estimates that are different from the SLS (Kavetski et al, 2006a). This is confirmed by the results of this study. The additional insight is that when calibrating the same model to the same streamflow data series, using different, but plausible, rainfall inputs, SLS parameters vary considerably, however the BATEA parameters remain relatively robust. This increases confidence that the BATEA parameter estimates are more robust then the SLS parameter estimates. The uncertainty in the BATEA parameters also greatly increases compared to the SLS parameters. This suggests that the model is over parameterised and model simplification may be required

6.2. Model Evaluation

Model evaluation is undertaken by comparing the simulated and observed streamflows for the calibration and validation periods using the Nash-Sutcliff (NS) efficiency. The simulated streamflow was based on a single parameter set - the modal parameter estimates. The results for the BATEA and SLS calibrations for each of the rainfall time series is given in Table 4.

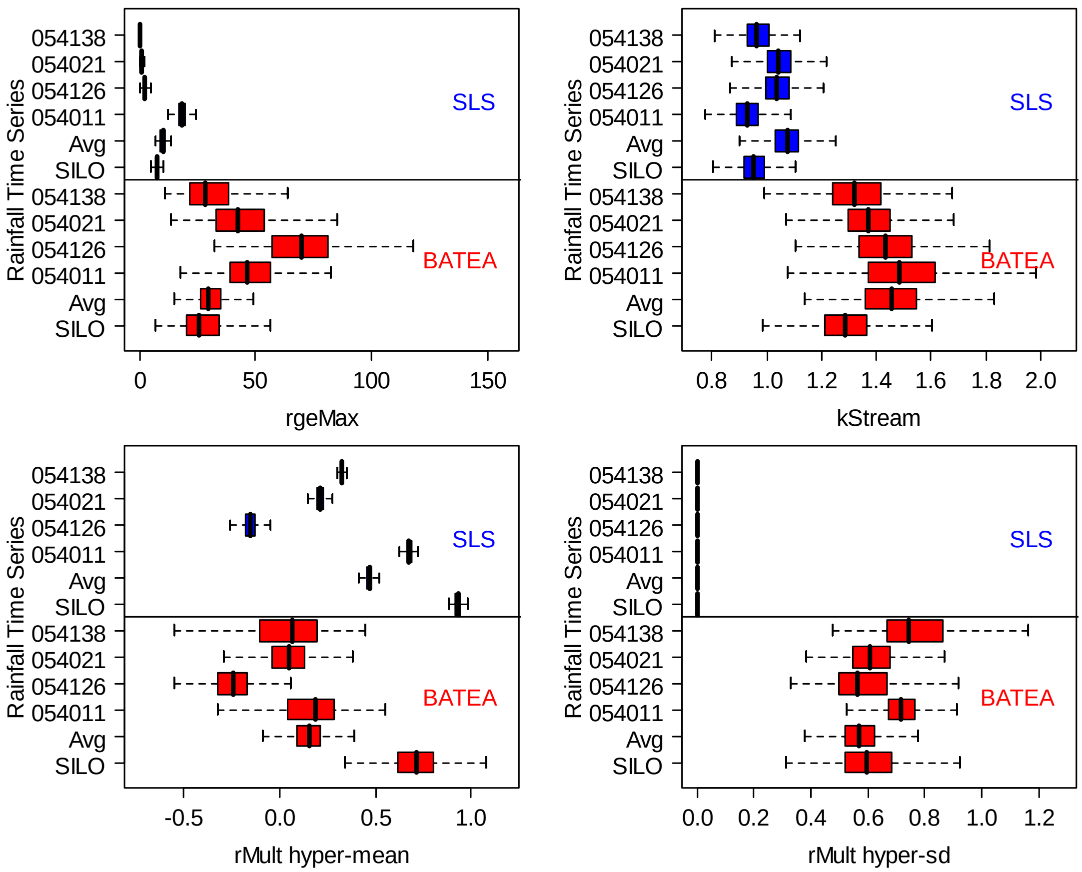


Figure 5. Comparison of parameters posteriors for SLS and BATEA for different rainfall time series.

For calibration the NS was higher for BATEA than the SLS. This is a typical result, the additional degrees of freedom provided by the rainfall multipliers enabling BATEA simulations to almost perfectly match the observed data. In validation, the NS drops significantly. For BATEA this is because the rainfall multipliers are unknown and therefore the modal estimate of the hypermean for rMult is used throughout the simulation.

Table 4. Nash-Sutcliffe (NS) statistics (bold indicates higher of either SLS or BATEA).

Rainfall Time Series	Calibration		Validation	
	BATEA	SLS	BATEA	SLS
SILO	**0.98**	0.91	0.72	**0.75**
Average	**0.96**	0.90	0.71	**0.78**
054011	**0.97**	0.88	0.23	0.23
054126	**0.92**	0.80	**0.42**	0.33
054021	**0.96**	0.87	**0.84**	0.59
054138	**0.97**	0.84	0.57	**0.79**

For validation, the results vary considerably depending on which rainfall time series is used.

For both SILO and average time series the results were similar between BATEA and SLS, with SLS slightly outperforming BATEA. Of the remaining time series, the highest NS overall was for BATEA calibration to rainfall time series 054021, while the areal rainfall estimates were 2nd and 3rd best for SLS and BATEA, respectively. For BATEA, the highest NS was significantly better than both areal rainfall time series. For SLS the best result was for rainfall time series 054138, although the improvement compared to the areal average was minor.

Interesting the catchment average rainfall (SILO or average) did not provide the highest NS in validation. This was provided by gauge 054021, calibrated to the BATEA parameter estimates. This suggests that the catchment average rainfall may not be the "best" rainfall for the prediction of streamflow. This was investigated by comparing the NS in validation using the two rainfall time series, SILO and 054021, and applying them to the 12 different model parameter sets (6 BATEA and 6 SLS) derived from calibrating to the six different rainfall time series. Figure 1 shows that the general trend is that the NS in validation is similar or better for the BATEA parameters compared to the SLS. This illustrates that BATEA parameter estimates are less sensitive to errors produced by using different rainfall time series. Furthermore, using 054021 rainfall time series provides better NS than the SILO rainfall time series when using the BATEA parameter estimates. This result suggests that the 054021 time series is a more appropriate rainfall time series for predicting runoff than the catchment average rainfall. This conclusion has a strong physical basis. Consideration of Figure 4 and Table 2 indicates that the 054021 gauge is located in the high rainfall area of the catchment, with 1.5 to 2.3 times the SILO catchment average rainfall. This area is likely the most productive in terms of generating runoff. The BATEA results concur with this conclusion - the 054021 rainfall had the lowest average rainfall error indicated by the rMult hyper mean, being close to 0 (in log space). These results indicate that rainfall from the most productive runoff generating areas produces better predictions than the catchment average rainfall. Furthermore, BATEA can utilise the most appropriate rainfall time series to provide enhanced runoff predictions as the parameters are less sensitive. The caveat is that for this case study the validation period is relatively short and consists of a low number of major streamflow events. This may introduce biases and further work is needed to determine the generality of these results by testing on more catchments with longer validation periods. BATEA was also implemented without any characterisation of model error. It is possible that the input errors compensate for this lack of model error. Hence further research will investigate the impact of incorporating model error on the results.

6.3. Model Diagnostics

Inspection of the model diagnostics was undertaken to determine if the hierarchical model assumptions were adequate. For the BATEA calibration, it is assumed that the latent variables, rMult, are lognormally distributed. Figure 7 shows the normal probability plot of the latent variables when calibrating to the average rainfall time series. The Kolmogorov-Smirnov (K-S) test indicated there was no significant evidence to reject the null hypothesis that the latent variables were lognormally distributed at the 5% level. The autocorrelation and partial autocorrelation plots showed no statistically significant autocorrelation (5% level). Similar results were found for the other five rainfall time series.

Figure 8 shows the posterior diagnostics for the streamflow residuals. The normal probability plot shows there is significant departure from normality with the distribution exhibiting fat tails - the KS test was statistically significant at the 1% level.

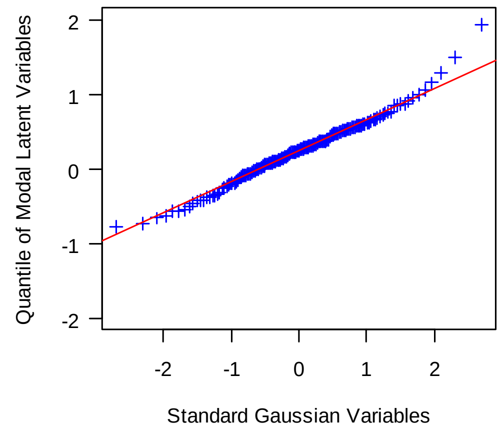


Figure 7. Normal probability plot for rainfall multiplier latent variables.

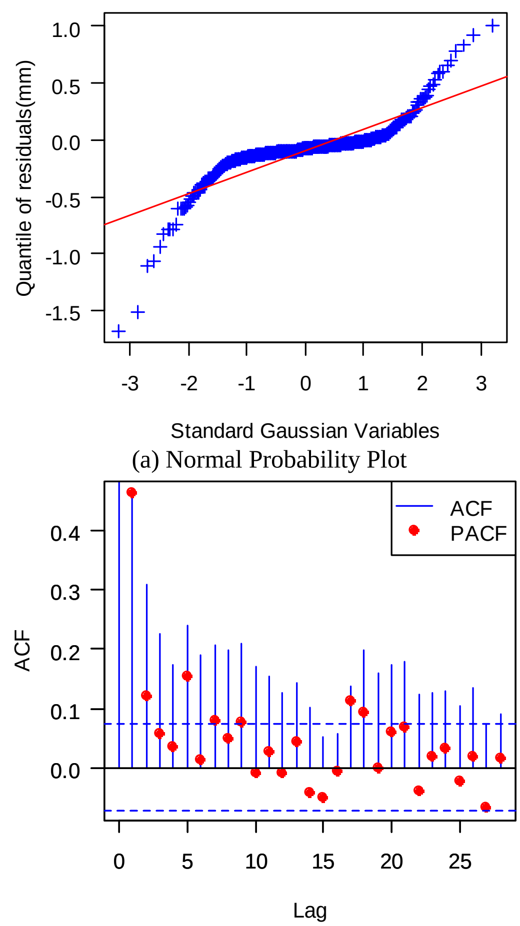


Figure 8. Diagnostics for streamflow residuals

Figure 8(b) shows there is statistically significant autocorrelation at the high lags. However, the majority of this autocorrelation is due to the high autocorrelation in the first few lags, as the partial autocorre1ation has few statistically significant correlations at lags greater than 5. Nonetheless, further research is needed to refine the response error model with heavy tailed distributions and methods to account for the autocorrelation.

7. CONCLUSION

This preliminary evaluation of BATEA using multi-site catchment rainfall data produced the following results: Firstly, calibrating the same model to the same streamflow using different rainfall time series the BATEA parameter estimates were considerably more consistent with each other than the SLS parameter estimates. This illustrates BATEA is able to provide more robust parameter estimates that are less sensitive to input error. This is important for regionalisation where catchment parameters are typically transferred from one catchment to another. Secondly, selection of the rainfall time series was found to be important for providing reliable streamflow prediction during model validation. Although, BATEA provides more consistent parameter estimates for different rainfall inputs during calibration, it is unrealistic to expect it to provide good streamflow predictions in validation if the rainfall input is poor. For this catchment rainfall time series from the high rainfall areas, and therefore most productive in terms of streamflow generation, outperformed the catchment average rainfall time series in validation. Given the strong rainfall gradient that exists in the catchment this result has a strong physical basis. However, this result was only true when the BATEA parameter estimates were utilised, the SLS parameters did not show the same result. This has important implications for regionalisation. It shows BATEA provides more robust parameter estimates that are able to utilise the most appropriate rainfall time series to enhance runoff predictions. The challenge will be identifying the most appropriate rainfall time series in an ungauged catchment.

There are several caveats which limit the generality of these conclusions. Firstly, model error was not incorporated. Secondly, the logSPM model is likely overparameterised. Thirdly, the validation period contained a low number of streamflow events and may be subject to biases. Fourthly, the model diagnostics illustrate further refinement of response error model is required. Although, further work is required to address these issues, the future looks promising for the potential of BATEA to improve the regionalisation of CRR models.

8. REFERENCES

Kavetski, D., Kuczera, G. and Franks, S.W., 2003. Semidistributed hydrological modeling: A "saturation path" perspective on TOPMODEL and VIC - art. no. 1246. Water Resources Research, 39(9), 1246-1246.

Kavetski, D., Kuczera, G. and Franks, S.W., 2006a. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resources Research*, 42(3)

Kavetski, D., Kuczera, G. and Franks, S.W., 2006b. Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. *Water Resources Research*, 42(3)

Kuczera, G., Kavetski, D., Franks, S. and Thyer, M., 2006. Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. *Journal of Hydrology*, 331(1-2), 161-177

Kuczera, G., Kavetski, D., Renard, B. and Thyer, M., 2007. Bayesian Total Error Analysis For Hydrologic Models: Markov Chain Monte Carlo Methods To Evaluate The Posterior Distribution. *Proceedings of the MODSIM 2007*

Peel, M., Chew, F., Weston, A. and McMahon, T., 2000. Extension of Unimpaired Monthly Streamflow Data and Regionalisation of Parameter Values to Estimate Streamflow in Ungauged Catchments, *National Land and Water Resources Audit*

Renard, B., Thyer, M., Kuczera, G. and Kavetski, D., 2007. Bayesian total error analysis for hydrologic models: Sensitivity to error models. *Proceedings of the MODSIM 2007*

Vrugt, J.A., Diks, C.G.H., Gupta, H.V., Bouten, W. and Verstraten, J.M., 2005. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resources Research*, 41(1)

Vrugt, J.A. and Robinson, B.A., 2007. Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. *Water Resources Research*, 43(1)