18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

Impact of runoff measurement error models on the quantification of predictive uncertainty in rainfall-runoff models

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

[1] School of Engineering, University of Newcastle, New South Wales, Australia
[2] Cemagref Lyon, Hydrology-Hydraulics, Lyon, France
Email: mark.thyer@newcastle.edu.au

Abstract: The development of a robust framework for quantifying the parametric and predictive uncertainty of conceptual rainfall runoff (CRR) models remains a key challenge in hydrology. For practical purposes, reliable and robust characterization of predictive uncertainty is important for comparing the impact of management options on key variables of interest (e.g. reservoir yield, meeting low flow criteria for ecological purposes). For research purposes, robust identification of the sources of uncertainty is essential for understanding how to reduce predictive uncertainty, and thereby enhance model predictions. Both these tasks are recognized as a major challenge for hydrological modelling science.

It is generally recognized that CRR modelling is affected by three main sources of uncertainty: (i) input uncertainty, e.g., measurement and sampling errors in the estimates of areal rainfall; (ii) output uncertainty, e.g., rating curve errors affecting runoff estimates; and (iii) structural uncertainty (sometimes referred to as "model uncertainty"), arising from lumped and simplified representation of hydrological processes in CRR models. Various approaches in the literature have aimed to quantify the individual contributions of input, output and structural uncertainties to the total predictive uncertainty. The beneficial impact of quantifying input errors on CRR parameter estimates and the reliability of model predictions has been established and techniques for evaluating model structural errors have begun to appear. However, almost all these studies make assumptions regarding the output (runoff measurement) errors.

This study evaluated whether there is any beneficial impact in utilizing rating curve data to fit a runoff measurement error model. This was undertaken by incorporating this fitted output error (OE) model into the Bayesian total error analysis (BATEA) methodology. BATEA provides a comprehensive framework to hypothesize, infer and evaluate probability models describing input, output and model structural error. BATEA was used to calibrate the GR4J model to the ephemeral Horton catchment. To evaluate the impact of the fitted OE model the calibration results were compared to two other OE models; one representing a commonly assumed OE model and the other representing a conservative "overestimate" of the OE model.

The estimated predictive uncertainty was more consistent with the observed runoff data for the fitted OE model than the assumed OE (which systemically under predicted the observed runoff) and the conservative OE (which overestimated the predictive uncertainty). This result was consistent in model calibration and validation. This illustrates for this case study there was beneficial impact in incorporating a fitted OE model. Comparison of the posterior distributions of parameters showed that the different OE model produced significantly different parameter estimates. This has implications for regionalizing parameters estimations to produce predictions in ungauged basins. Comparison of the estimated input/structural errors also showed substantial differences for different OE. This suggests an interdependency between the error sources, where reliable estimates of input/structural errors will be dependent on reliable estimates of the output error.

Keywords: *Predictive uncertainty, conceptual rainfall-runoff modelling, model calibration, output error, regionalisation*

1. INTRODUCTION

Conceptual rainfall-runoff (CRR) models simulate water balance dynamics at the catchment scale. Given the significance of water in human society, and in aquatic and terrestrial ecosystems, catchment models form an integral part of virtually all environmental models formulated at the catchment scale. Moreover, they underpin most water resources planning models. A key challenge currently faced by hydrological modeling science is the quantification of uncertainty in model predictions.

It is broadly recognized that CRR modelling is affected by three main sources of uncertainty: (i) input uncertainty, e.g., measurement and sampling errors in the estimates of areal rainfall; (ii) output uncertainty, e.g., rating curve errors affecting runoff estimates; and (iii) structural uncertainty (sometimes referred to as "model uncertainty"), arising from lumped and simplified representation of hydrological processes in CRR models.

Recent work has aimed at quantifying the individual contributions of input, output and structural uncertainties to the total predictive uncertainty [*Kuczera et al.*, 2006; *Ajami et al.*, 2007; *Huard and Mailhot*, 2008]. The beneficial impact of quantifying input errors on CRR parameter estimates and the reliability of model predictions has been demonstrated [*Kavetski et al.*, 2006a; b; *Thyer et al.*, 2008]. Several studies have also developed techniques for incorporating the impact of model structural errors [*Kuczera et al.*, 2006; *Smith et al.*, 2008; *Bulygina and Gupta*, 2009]. More recently, *Renard et al.* [2009] have shown the interdependency of input and model structural errors, where prior information on input errors is required to successfully identify the individual contributions of input and model structural errors.

In all these studies an output error model to represent runoff measurement error has been explicitly included in the model calibration but its parameter values have been assumed. Other studies have utilized rating curve data to derive estimates of the runoff measurement error [*Huard and Mailhot*, 2008; *Thyer et al.*, 2008]. *Bulygina and Gupta* [2009] noted that very little information on the actual measurement error of runoff is known and the area is still under development. This study aims to contribute to the research on the impact of the runoff measurement error model, with the following specific aims:

1. Develop runoff measurement error models based on rating curve analysis
2. Compare the derived runoff measurement error models to the commonly used assumed runoff measurement error models and evaluate their impact on (a) Predictive uncertainty estimates (b) CRR parameter estimates (c) Input/structural error uncertainty estimates.

2. BAYESIAN TOTAL ERROR ANALYSIS (BATEA) FRAMEWORK]

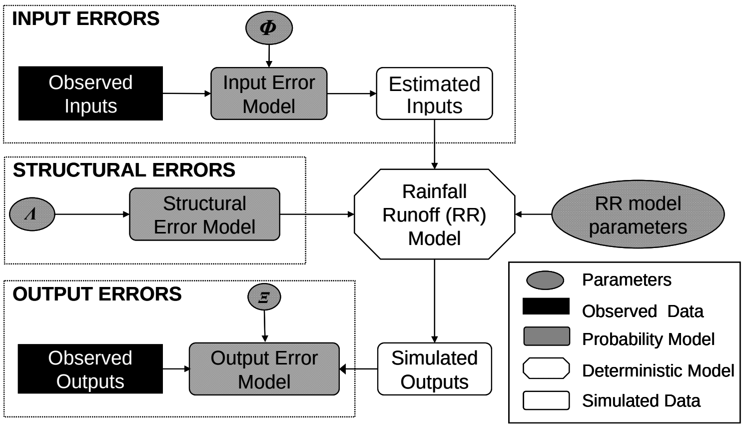
The Bayesian total error analysis (BATEA) methodology provides a comprehensive framework to hypothesize, infer and evaluate probability models describing input, output and model structural error and will be used in this study. The BATEA framework conceptualizes the propagation of error in the CRR model using a hierarchical model. A schematic of this hierarchical model in calibration mode is depicted in Figure 1. Only the components relevant to the study will be explained in this paper, the remainder will be briefly outlined, For a full description the reader of the BATEA framework the reader is referred to *Thyer et al [2008]*. First some notation; Let $\mathbf{R} = \{r_t \; ; t = 1, ..., T\}$ denote the true rainfall inputs of the CRR model and $\tilde{\mathbf{R}} = \{\tilde{r}_t \; ; t = 1, ..., T\}$ be the observed values of these inputs. Similarly, let $\boldsymbol{Q}$ denote the true runoff outputs, $\tilde{\boldsymbol{Q}}$ the observed runoff and $\hat{\boldsymbol{Q}}$ the runoff predicted by the model.

2.1. Input errors

Following, *Kavetski et al.* [2006a] we assume observed rainfall is corrupted by multiplicative errors: $r_t = \varphi_t \tilde{r}_t$, where φ_t is a rainfall multiplier. Following *Kuczera et al.* [2006], the rainfall multipliers are assumed to follow a log-normal distribution $\log \varphi_t \sim N(\mu_r, \sigma_r^2)$, with hyperparameters $\boldsymbol{\Phi} = (\mu_r, \sigma_r^2)$. As the rainfall accuracy is assumed to be unknown the rainfall multipliers and their hyperparameters are included in the inference list. In this study, rainfall multipliers were applied to daily rainfall totals. To reduce the number of rainfall multipliers to be inferred (and therefore the computational burden), a pre-processing heuristic [*Thyer et al.*, 2008] was used to eliminate 'insensitive' daily rainfall multipliers which were associated days with low rainfall which did not produce significant runoff.

2.2. Stochastic Parameters for Structural errors

Stochastic CRR parameters for model structural errors [*Kuczera et al.* 2006] were not inferred in this case study because *Renard et al.* [2009] showed that joint inference of both rainfall multipliers for input error and stochastic parameters for structural error becomes ill-posed when only vague prior information on input error is available. When only input errors are inferred, the rainfall multipliers do not represent input errors exclusively and are also contaminated by model structural errors [*Renard et al.*, 2009]. Hence for this study, the rainfall multipliers will be considered to represent both input/structural errors.



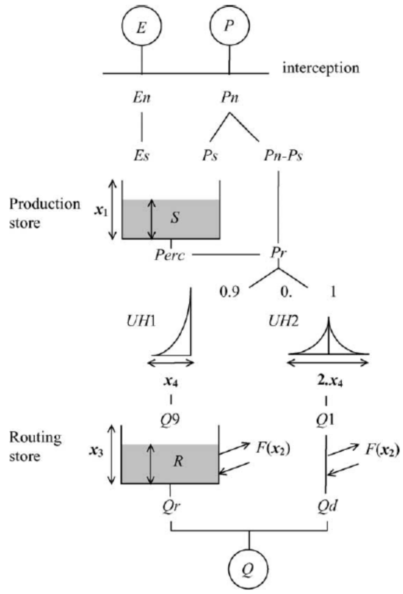


Figure 1. Schematic of BATEA framework

Figure 2. GR4J [*Perrin et al.*, 2003]

2.3. Output Errors: Runoff Measurement Errors

The measurement error in observed runoff is due mainly to rating curve errors, assuming an additive Gaussian error model:

$$\tilde{q}_t = q_t + \gamma_t ; \qquad \gamma_t \sim N(0, \sigma_\gamma^2) \tag{1}$$

Parameter σ_γ is generally estimated using rating curve analysis and set prior to model calibration. In this paper various probability models were trialled and these are outlined in Section 4.2

2.4. Output Errors: Remnant errors

The output measurement error model (1) links the observed runoff with the true runoff. Since the latter is unknown, an additional error model linking the true runoff with the simulated runoff must be specified. Here, we use an additive Gaussian error model with unknown standard deviation σ_ε:

$$q_t = \hat{q}_t + \varepsilon_t ; \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \tag{2}$$

The interpretation of the error term ε_t depends on the error sources explicitly represented in the calibration framework [*Renard et al.*, 2009]. In this study as rainfall multipliers will represent stochastic input/structural errors, ε_t represents "remnant" errors, i.e., errors remaining as results of imperfections of the selected error models. Note that if runoff measurement errors γ_t and remnant errors ε_t are independent, the distribution of observed runoff conditioned on simulated runoff is

$$\tilde{q}_t = q_t + \gamma_t = \hat{q}_t + \varepsilon_t + \gamma_t = \hat{q}_t + \eta_t ; \tag{3}$$

If both runoff measurement error and remanent errors are assumed Gaussian, then $\eta_t \sim N(0,\sigma_\eta^2)$. However, the parameterization of σ_η^2 depends on the probability model chosen for the runoff measurement error and is further detailed in section 4.2.

2.5. Inferring the Posterior distribution

In Bayesian context the goal of model calibration is to infer the posterior distribution of all quantities. The posterior distribution is given by Bayes' theorem as follows (see *Kavetski et al (2006a) and Kuczera et al 2006* for details):

$$p(\boldsymbol{\theta},\boldsymbol{\Phi},\mu_r,\sigma_r,\boldsymbol{\Lambda},\sigma_\varepsilon \mid \tilde{\boldsymbol{Q}},\tilde{\boldsymbol{R}}) \propto p(\tilde{\boldsymbol{Q}} \mid \boldsymbol{\theta},\boldsymbol{\Phi},\boldsymbol{\Lambda},\sigma,\tilde{\boldsymbol{R}})\, p(\boldsymbol{\Phi} \mid \mu_r,\sigma_r)\, p(\boldsymbol{\theta},\mu_r,\sigma_r,\sigma_\varepsilon) \tag{4}$$

It is beyond the scope of this paper to define all the terms in this equation. The reader is referred to *Thyer et al.* [2008] for further details. Development of methods to sample from such high-dimensional posteriors is computationally challenging but not insurmountable. We use a two-stage MCMC strategy detailed by *Kuczera et al.* [2009].

3. CASE STUDY CATCHMENT AND CRR MODEL

To enable comparison to past studies [*Thyer et al.*, 2008] the case study catchment used was the Horton catchment (1920 km^2), located in northern inland New South Wales, Australia. It is an ephemeral catchment, with an annual runoff coefficient of 0.13. The catchment average rainfall was calculated using Thiessen polygons with the available raingauges. The period used for model calibration was 2 years, (03/01/1977 – 30/12/1978), while for model validation a 15 month period was used (10/01/1975-24/03/1976). In both cases a 100 day warm-up was employed prior to the start of the period to reduce the effects of the initial conditions.

The CRR model used for this case study was the lumped GR4J model [*Perrin et al.*, 2003]. This model has a parsimonious form with only four calibrated parameters and has been extensively tested over hundreds of catchments worldwide, with a range of climatic conditions from tropical to temperate and semi-arid catchments. Figure 2 shows a schematic of the GR4J model. The GR4J has four parameters: the capacity of the production store x_1 (mm), the groundwater exchange coefficient x_2 (mm), the capacity of the non-linear routing reservoir x_3 (mm) and the unit hydrograph time base x_4 (days).

4. DEVELOPMENTOF OUTPUT ERROR MODELS

4.1. Estimating Runoff Measurement Error using Rating Curve Analysis

The runoff measurement error was estimated using the runoff gaugings and the rating curves from the Horton catchment. The runoff measurement error model was derived by calculating the actual runoff measurement error, $\tilde{\gamma}_t = q_t - \tilde{q}_t$ for each runoff gauging. The gauged runoff was assumed error free and treated as the true runoff, q_t, while runoff predicted by the rating curve was taken as the observed runoff $\tilde{q}_t$. The following heteroscedastic error model was fitted to the actual runoff measurement errors:

$$\gamma_t \sim N(0,\sigma_\gamma), \qquad \sigma_\gamma = a + b\tilde{q}_t^{\,c} \tag{5}$$

Parameter *a* represents the measurement error at zero observed runoff, parameter *b* represents the linear rate at which the measurement error increases with $\tilde{q}_t$ and the *c* parameter is included to enable non-linear, power form changes in the measurement error as a function of the observed runoff. If *c*=1 Eq. (5) reduces to the standard linear heteroscedastic error model commonly used.

The model was fitted to the measurement error data using WINBUGS [*Spieghalter et al.*, 2003] to estimate the posterior distribution of *a, b* and *c*. Inspection of the posterior of *c* showed that it was well away from the value of *c*=1, providing strong evidence that the power form of the model was justified. The expected value of the posterior distributions from WINBUGS for the power form are provided in Table 1, with models named "PowFit".

4.2. Comparison of Various Output Error Models

The fitted measurement error model will be compared to various other output error (OE) models which are summarized in Table 1. The OE model "LinWRR_NoRE" used by *Thyer et al. [2008]* was also based on rating curve analysis, but ignored low flows, using an arbitrary threshold of 0.5mm (flows <0.5mm comprise over 90% of the runoff time series) and the runoff measurement errors ignored changes in rating curves through time. Comparison of the parameter values shows this OE model has the largest uncertainty and can be considered a "conservative" approach. The model "LinAss_NoRE" assumes a 10% heteroscedastic runoff measurement error, which is a common assumption, with no remnant errors. The model "LinAss_RE" is same as previous, but remnant errors are also estimated. The model "PowFit_RE" uses expected posterior values of parameters from rating curve analysis undertaken in Section 4.1. The model "PowFit_RE_Censor": is same as previous but low flows (runoffs<0.15mm) are excluded from the evaluation of the likelihood. As the catchment is ephemeral, low flows compromise a large proportion of the observed runoff time series. In this situation there is the risk the CRR model is fitting to low flows only. This OE model will test whether censoring the low flows impacts on CRR model parameters.

Table 1. Output error models

OE Model Name	*a*	*b*	*c*	Remnant errors	Parameterization of σ_η^2
LinWRR_NoRE	0.4	0.086	1.0	No	$(0.4+0.086\tilde{q}_t)^2$
LinAss_NoRE	0.0	0.1	1.0	No	$(0.1\tilde{q}_t)^2$
LinAss_RE	0.0	0.1	1.0	Yes	$(0.1\tilde{q}_t+\sigma_\varepsilon)^2$
PowFit_RE	0.1	0.024	0.83	Yes	$\left(0.1+0.024\tilde{q}_t^{\,0.83}\right)^2+\sigma_\varepsilon^2$
PowFit_RE_Censor	0.1	0.024	0.83	Yes	$\left(0.1+0.024\tilde{q}_t^{\,0.83}\right)^2+\sigma_\varepsilon^2,\ for\ \tilde{q}_t>0.15$

5. CALIBRATION RESULTS

5.1. Estimation of Predictive Uncertainty

The predictive QQ plot [*Laio and Tamea*, 2007; *Thyer et al.*, 2008] is used to assess whether the time-varying predictive distribution of runoff is consistent with the observed runoff (refer to *Thyer et al. [2008]* for guidance on interpretation). Basically the closer to the 1:1 line the closer the predictive distribution is to being consistent with the observed data. Figure 3(a) shows that in calibration the LinWRR_NoRE OE model overestimates the predictive uncertainty. The assumed LinASS OE model improves considerably when the remnant errors are included. However, there is still a systematic underprediction of the observed data, and a reasonable proportion of observations (~5%) that are outside the range of the predictions. The fitted (PowFit_RE) OE model performs best, particularly with censoring of low flows ((PowFit_RE_Censor). In validation, (Figure 3(b)) the assumed models (LinAss_RE and LinAss_NoRE) perform relatively poorly, with the best appearing to be the fitted OE model without censoring (PowFit_RE).

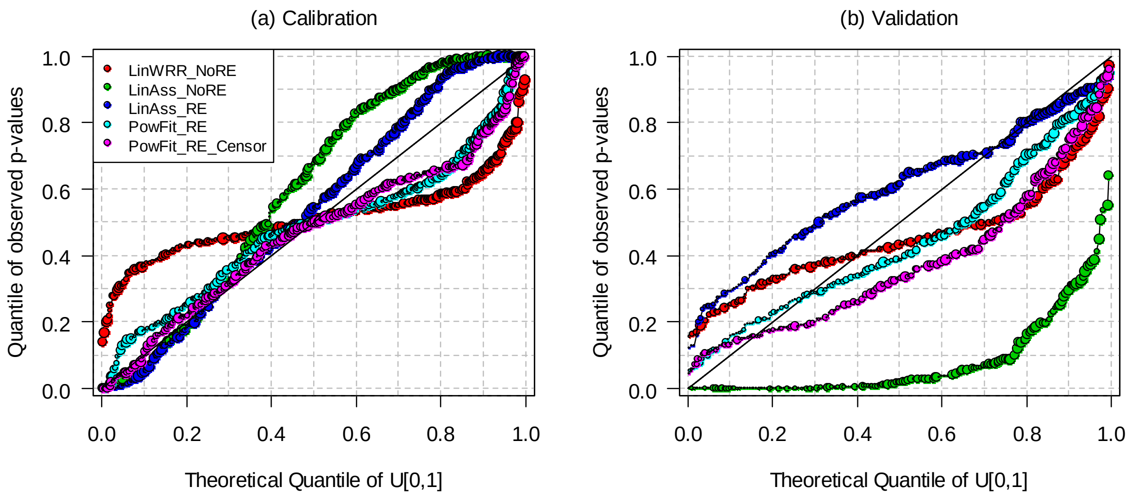


Figure 3. Predictive QQ plots

5.2. Estimation of CRR Parameters

GR4J parameters x_1 and x_3 control the size of the production and routing store respectively. Figure 4 shows that the different OE models produce significantly different posteriors for x_1 (similar trends are evident for x_3). In particular the censoring of low flows significantly reduces the x_1. A reduction in a storage parameter would produce a flashier response, which is associated with quickflows. Thus this result is consistent with a censoring of low flows from the calibration

GR4J parameter x_2 controls the importing/loss of water to the catchment via groundwater transfer (a negative value indicates a groundwater loss). Figure 4 also compares the x_2 parameter to the expected value of the rainfall multipliers, E(rainfall multipliers), which controls the average rainfall input to the model. It shows the complementary nature of these two quantities. The LinASS_NoRE OE model results in a substantial increase in the E(rainfall multiplier), which results in an increase in the groundwater loss parameter. Conversely, the PowFit_RE_Censor results in a decrease in E(rainfall multiplier), and hence a decrease in the groundwater loss parameter to almost zero. These results indicate selection of the output error model has considerable impact on these inferred quantities and further work is needed to understand what is an appropriate error model to provide reliable estimates of these inferred quantities.

5.3. Estimation of Input/Structural and Remnant Errors

Figure 5 compares estimates of the coefficient of variation (CV) of the rainfall multiplier to the remnant error standard deviation, σ_ε. The magnitude of these two quantifies represent the scale of the uncertainty associated with input/structural errors for CV(rainfall multipliers) and remnant errors. Note that the assumed OE models (LineAss_RE and LineAss_NoRE) result in considerably higher CV(rainfall multipliers), compared to the fitted OE models. This is likely to be because there is no minimum error term (a=0) in the assumed OE model. The LineWRR_NoRE produces the smallest CV(rainfall multiplier) most likely because it has the highest minimum error term (a=0.4), which comparing to the fitted value (PowFit) of a=0.1 is a conservative overestimate of the minimum runoff measurement error. If the PowFit runoff measurement error model is closet to the truth then a conservative approach (e.g. LineWRR_NoRE) of overestimating the runoff measurement error would produce an underestimate of the input/structural errors.

Comparing the PowFit_RE_Censor and PowFit_RE OE model the remnant error term is significantly larger when the low flows are censored. This suggests that low flows have different remnant error structures to the high flows.

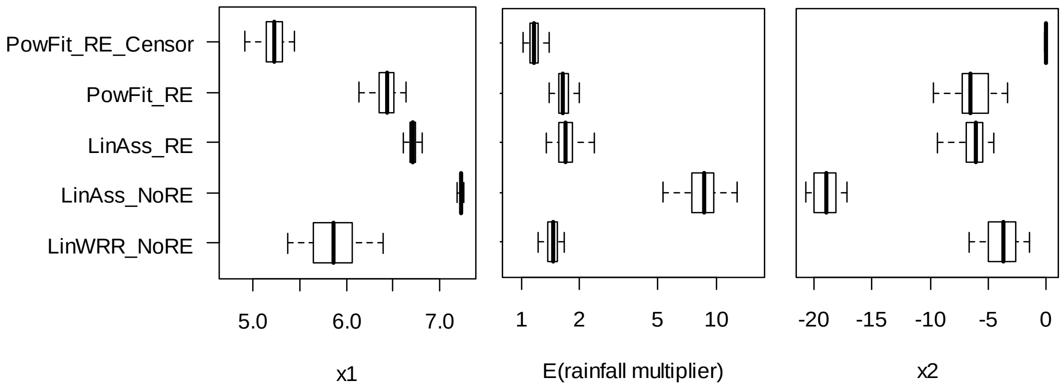


Figure 4. Posteriors of GR4J parameters, x_1 (x-axis scale is in natural log space), E(rainfall multipliers) and x_2 (units, mm/day)

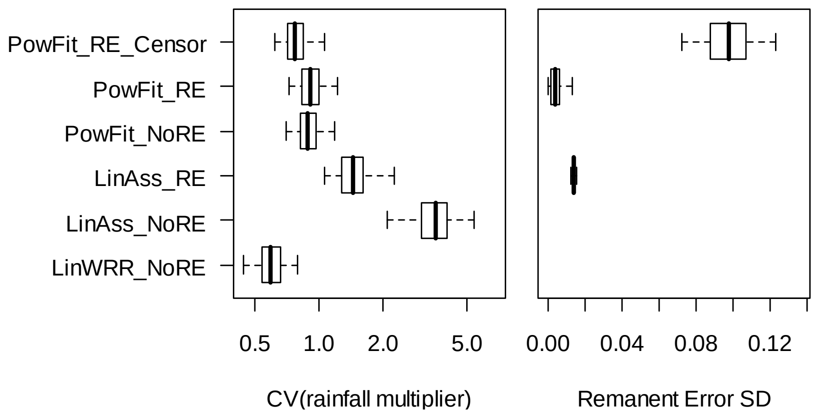


Figure 5. Posteriors of CV(rainfall multipliers) and Remnant Error SD

6. DISCUSSION AND CONCLUSIONS

This study evaluated the impact of different output error (OE) models on the estimates of predictive uncertainty, estimates of CRR parameters and input/structural errors. The OE models compared were a runoff measurement error models based on a common assumption of a heteroscedastic error standard deviation being 10% of runoff; a fitted runoff measurement error model based on rating curve analysis and a runoff measurement error model which resulted in a ‘conservative’ over-estimate of the true runoff measurement error.

Comparison of the predictive uncertainty estimates showed that the fitted runoff measurement error model outperformed both the assumed and conservative runoff measurement error model. This indicates that to obtain reliable estimates of the predictive uncertainty, the runoff measurement error model should be fitted to actual rating curve data, and neither a model based a common assumptions nor a conservative model will suffice.

Comparison of the CRR parameter and input/structural errors estimates revealed that both these quantities are strongly conditioned on the runoff measurement error model. This indicates that poor characterization of the output error model will lead to biased CRR parameter estimates and biased estimates of input/structural uncertainty.

Overall, this study highlights that fitting of the runoff measurement error model to rating curve data is an essential ingredient to improve estimates of the predictive uncertainty and CRR parameters.

REFERENCES

Ajami, N. K., Q. Y. Duan, and S. Sorooshian (2007), An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, *Water Resources Research*, 43(1).

Bulygina, N., and H. Gupta (2009), Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation, *Water Resour Res*, 45, 20, doi:W00b1310.1029/2007wr006749.

Huard, D., and A. Mailhot (2008), Calibration of hydrological model GR2M using Bayesian uncertainty analysis, *Water Resources Research*, 44.

Kavetski, D., G. Kuczera, and S. W. Franks (2006a), Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water Resour Res*, 42(3), doi:10.1029/2005WR004368.

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

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

Kuczera, G., B. Renard, D. Kavetski, and M. Thyer (2009), Markov Chain Monte Carlo Strategies for Bayesian Hierarchical Calibration of Hydrologic Models, *Water Resources Research*, submitted,

Laio, F., and S. Tamea (2007), Verification tools for probabilistic forecasts of continuous hydrological variables, Hydrol. Earth Syst. Sci. J1 - HESS, 11(4), 1267-1277, http://www.hydrol-earth-syst-sci.net/11/1267/2007/hess-11-1267-2007.pdf

Perrin, C., C. Michel, and V. Andreassian (2003), Improvement of a parsimonious model for streamflow simulation, *J Hydrol*, 279, 275-289.

Renard, B., D. Kavetski, G. Kuczera, M. A. Thyer, and S. W. Franks (2009), Understanding predictive uncertainty in CRR modellling: Le challenge of identifying input and structural errors, *Water Resource Research*, submitted.

Smith, P. J., K. J. Beven, and J. A. Tawn (2008), Detection of structural inadequacy in process-based hydrological models: A particle-filtering approach, *Water Resour. Res.*, 44, http://dx.doi.org/10.1029/2006WR005205

Spieghalter, D. J., A. Thomas, and N. Best (2003), WinBugs, Version 1.4, User Manual, MRC and Imperial College of Science, Technology and Medicine.

Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, *Water Resour. Res.*, 45, W00B14, http://dx.doi.org/10.1029/2008WR006825