How to quantify uncertainty in water allocation models? An exploratory analysis based on hypothetical case studies

J. LERAT[1], K. TOMKINS[1], Q. SHAO[2], L. PEETERS[1], A. YANG[1] & D. RASSAM[1]

1 CSIRO Land and Water, GPO Box 1666, Canberra, Australian Capital Territory 2601, Australia
julien.lerat@csiro.au

2 CSIRO Mathematical and Information Sciences, Private Bag no. 5, Wembley, Western Australia 6913, Australia

Abstract Water allocation models are the principal tools used to build water sharing plans in regulated river systems across Australia. These models associate components describing the physical system (e.g. rainfall–runoff transformation and flow routing) and the management rules (e.g. operation of dams and irrigation extractions). In a context of growing pressure on water resources, the uncertainty associated with these highly parameterised models needs to be quantified in a defensible and transparent way. This paper is an exploratory analysis based on the application of two uncertainty methods to hypothetical river system case studies. First, a simplified model structure is developed by using existing characteristics from six regions of the Murray-Darling basin. Each model contains a schematic representation of the region with: (1) one upstream storage, (2) ungauged and gauged tributaries inflows, and (3) one downstream irrigation extraction point. Second, the model outputs (streamflow) are corrupted and an alternative model is calibrated based on the corrupted data using the standard least-squares method. Finally, the uncertainty in the output is generated using two uncertainty post-processors and compared with the uncorrupted ("true") outputs using deterministic and probabilistic scores. The uncertainty post-processor based on the empirical distribution of the residuals proved simple yet effective, especially when compared to the more advanced bootstrapping method. It performed systematically better when applied to the stored volumes in the reservoir. However, the predictive uncertainty was greatly improved by the bootstrapping method to assess the uncertainty on streamflow. This highlights the important differences that can occur in the uncertainty analysis for a multivariate model.

Key words water allocation models; synthetic experiments; uncertainty assessment; calibration; bootstrap

INTRODUCTION

Water allocation models are used to compare water sharing scenarios in regulated catchments in terms of the impacts for the water users and the environment. These models include a representation of the physical system with modules such as flow routing, rainfall–runoff modelling or groundwater/surface water interactions, as well as management components to take into account infrastructure such as dams, canals or extraction points. Water allocation models can be complex modelling structures with a large number of parameters to be calibrated on limited datasets, especially regarding the management aspects. At the same time, water allocation models are used to aid in the making of long-term decisions with important social and environmental impacts. As a result, the assessment of uncertainty becomes a critical task to inform the decision makers about the strengths and weaknesses of the model.

Since the early work of Beven & Binley (1992) with the Generalized Likelihood Uncertainty Estimation (GLUE) method, the assessment of uncertainty has become a central topic in hydrological sciences. A wide range of methods have been developed with two main objectives: (1) locate and quantify the sources of uncertainty associated with the model components (inputs, parameters, model structure), and (2) quantify the predictive uncertainty. The first objective encompasses the second one because predictive uncertainty can be seen as a by-product of a detailed analysis of all uncertainty sources. For example, the Bayesian total error analysis (BATEA, Kavetski *et al.*, 2006) is a calibration method that jointly estimates the input and parameter uncertainty for a rainfall–runoff model. Once the distributions of both elements are estimated, the predictive uncertainty can be easily generated (Thyer *et al.*, 2009). This type of approach is appealing because it offers simultaneously a diagnostic on the model behaviour and a better understanding of the predictive uncertainty. Methods such as GLUE (Beven & Binley, 1992), Parameter ESTimation (PEST, Doherty, 2010) and the Shuffle Complex Evolution

Metropolis algorithm (SCEM,Vrugt *et al.*, 2003) have been developed with the same objective. However, the application of these methods on complex models remains limited due to long runtimes and lack of interface between model codes and Monte Carlo samplers. Moreover, they require a recalibration of the models, which is not always possible if the model has been developed during a previous study.

If the objective of the uncertainty analysis is limited to a quantification of the predictive uncertainty, a simplified approach can be applied by lumping all the uncertainties in the error term. Practically, the model is run once on the calibration period to obtain a series of deviations from the observations (model residuals). The predictive uncertainty is then generated based on these residuals. This paper uses this method as a simple alternative to more sophisticated methods and the results are reported in later sections.

The estimation of the predictive uncertainty based on the model residuals is a popular technique in the forecasting community where deterministic forecasts are "dressed" to generate probabilistic forecasts (Krzysztofowicz, 1999). For simulation models, Montanari & Brath (2004) have proposed a parametric error model based on Normal Quantile transformed variables. Because water allocation models are highly parameterised, the addition of another layer of parameterisation in the error model does not seem an appropriate choice in a first instance according to the parsimony principle. Alternatively, non-parametric statistical methods, like the nearest-neighbour algorithm proposed by Lall & Sharma (1996), may offer a solution. This paper presents a comparison of two non-parametric methods to generate predictive uncertainty. The methods are applied on simplified water allocation models described in the following section. In the remaining part of this paper, the uncertainty methods are designated as the "uncertainty post-processor".

METHOD

The comparison of the two uncertainty post-processors was conducted in four steps using a synthetic experiment modelled after the work of Renard *et al.* (2010):

(a) Simplified water allocation models were built on six regions within the Murray-Darling Basin to generate "true" (albeit synthetic) time series of streamflow and stored volume (see Fig. 3 for an example) on the period 1986–2005.
(b) The streamflow time series resulting from step (a) were corrupted with a random multiplicative error to introduce gauge uncertainty. The corruption was repeated 10 times.
(c) A model is calibrated against the corrupted streamflows with a minimization of the sum of squared errors on streamflow during the period 1986–1995. The calibration was repeated 10 times on the 10 corrupted data sets.
(d) The uncertainty post-processors were applied on the calibrated model to generate probabilistic outputs which were finally validated against the uncorrupted ("true") time series via two performance scores computed on the validation period (1996–2005).

Simplified water allocation models to generate "true" values

The simplified water allocation models were developed at the daily time step with the Integrated Water Quantity and Quality simulation Model software (IQQM, Simons *et al.*, 1996) for the period 1986–2005. Figure 1(a) presents the location of the regions; their characteristics are given in Table 1. The simplified models were not built to accurately reproduce the behaviour of the real systems but to provide a realistic modelling scenario regarding the model components and the amount of information available to calibrate them. The structure of the models is presented in Fig. 1(b). It includes three sub-catchments, one reservoir and one irrigation extraction point. The inflows from the three sub-catchments were set to three observed streamflow time series located in the region. The streamflow data have been selected from the high quality data set collated by Vaze *et al.* (2010); the corresponding catchments are located on Fig. 1(b). Because these catchments are small compared to the size of the whole region, the flows were rescaled so that the sum of the three mean annual flows equals the surface water availability of the region (see column 6 in

Table 1). A linear lag and route model was used to route the reservoir releases with a lag and attenuation parameters set to 1 and 3 days, respectively.

The capacity of the reservoir was set to the sum of the capacities of all public storages in the region (see column 4 in Table 1). The irrigation extractions were computed with the second crop model of the IQQM software to obtain a mean annual extraction equal to the one mentioned by the Australian Commonwealth Scientific and Research Organization (CSIRO, 2008) for each region (see column 5 in Table 1). This component of the model was assumed to be perfectly known (i.e. no uncertainty) and the extractions were treated as a fixed time series in all the tests presented here.

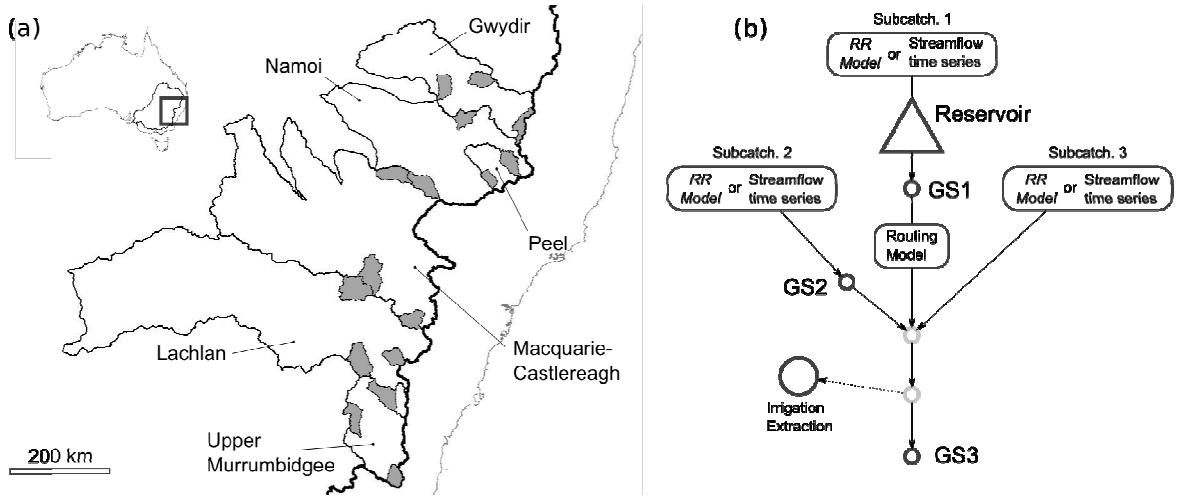

Fig. 1 (a) Location of the six test regions and the catchments used to generate the inflows (shaded area) (b) Schematic diagram of the simplified water allocation model applied to each region.

Table 1 Characteristics of the regions used to develop the simplified water allocation models (CSIRO, 2008).

(1) Model	(2) Region name	(3) Area (km^2)	(4) Public storage ($10^6 m^3$)	(5) Irrigation use ($10^6 m^3$/year)	(6) Surf. water availability ($10^6 m^3$/year)
GWYD	Gwydir	24 947	1346	313	782
NAMO	Namoi	39 781	813	244	965
PEEL	Peel	4 670	65	16	189
LACH	Lachlan	85 532	1252	273	1139
MACQ	Macquarie	73 453	1530	343	1567
UBID	Upper-Murumbidgee	13 100	0*	0	1600

* The UBID model does not take into account the reservoirs located in the Australian Capital Territory.

Four "true" outputs were generated with this model: (1) the stored volume in the reservoir, (2) the streamflow downstream of the reservoir at GS1, (3) the streamflow at GS2, and (4) the streamflow at the downstream end of the system at GS3. These outputs were chosen to match with the measurement points generally available in this type of catchment (e.g. the sub-catchment 3 is considered as ungauged).

Corruption of the "true" streamflow data

The "true" streamflows generated with the previous model were corrupted with an auto-regressive multiplicative error:

$$q^*_j = q_j \,.\, \max(0, \mu_j + 1) \quad (1)$$

$$\mu_j = \alpha \,.\, \mu_{j-1} + (1 - \alpha^2)^{0.5} . \xi_j \quad (2)$$

where q_j is the true streamflow on day j, q^*_j the corrupted streamflow, α the autoregressive coefficient and ξ_j an uncorrelated Gaussian noise having a mean of 0 and a standard deviation of 0.15 (streamflow error of 15%). The coefficient was set to 0.997 so that the lagged-autocorrelation of the error μ becomes non-significant for lags greater than 500 days. The corruption model was applied ten times to generate ten corrupted data sets.

Calibration of the model against corrupted streamflow data

A model with a similar structure was calibrated against the corrupted data on the calibration period 1986–1995. In this model the inflows from the three sub-catchments were generated with the GR4J rainfall–runoff model (Perrin *et al.*, 2003). The water allocation model was calibrated independently on the three areas delimited by the three gauging stations GS1, GS2 and GS3 (see Fig. 1(b)). The Levenberg-Marquardt gradient descent algorithm (Marquardt, 1963) was applied to minimise the sum of squared errors on the streamflows. 14 parameters were calibrated: four parameters for the GR4J mode on each sub-catchment and two parameters for the routing model. This procedure was repeated 10 times for each corrupted data set.

Application of two uncertainty post-processors and validation with performance scores

The calibrated model was run on the validation period (1996–2005) and for each day j of the period, 51 values (or ensembles) were generated for each of the four model outputs using two uncertainty post-processors. The first one, named EMP, computes the ensembles using the distribution of the residuals obtained from the calibration period according to:

$$\hat{V}^{EMP}_{j,k} = \hat{v}_j + \varepsilon'_{N.k/52} \quad (3)$$

with k = 1, ..51 the ensemble rank, $\hat{V}^{EMP}_{j,k}$ the kth ensemble generated with EMP on day j, $\hat{v}_j$ the simulated value from the calibrated model on the same day, N the number of time steps in the calibration period, (ε'_i) the series of N residuals obtained from the calibration period rearranged in ascending order, and *k/52* the Weibull estimate of the sampling frequency. With this uncertainty post-processor, the distribution of the residuals is invariant and equal to the one of the (ε'_i).

A second post-processor, named KNN, was applied based on the KNN bootstrapping algorithm proposed by Lall & Sharma (1996). For each day j of the validation period, M days l_1,.., l_M are selected from the calibration period where $\hat{v}_{l_1}$, ... $\hat{v}_{l_M}$ are the closest simulated value to $\hat{v}_j$. These days are called the "nearest neighbours" of j. The residuals ε_{l_1} , ... ε_{l_M} corresponding to these M days are then sampled 51 times according to the sampling kernel proposed by Lall & Sharma (1996) that gives more weight to the closer neighbours:

$$K(n) = (1/n) \,/\, \sum_{n=1}^{M} 1/n \quad (4)$$

with n the rank of the neighbour and $K(n)$ the probability to sample this neighbour. Several values were tested for M and the value of 50 was retained. Finally, the ensembles are generated according to:

$$\hat{V}^{KNN}_{j,k} = \hat{v}_j + \varepsilon''_k \quad (5)$$

with (ε''_i) a rearrangement of the 51 values sampled from the M nearest neighbours in ascending order. Because the nearest neighbours are different from one time step to another, the distribution of the residuals generated by this post-processor vary in time with variations driven by the simulated value $\hat{v}_j$.

The two post-processors were finally compared using the following relative performance score:

$$R_f = \frac{f(v,\hat{V}^{EMP}) - f(v,\hat{V}^{KNN})}{f(v,\hat{V}^{EMP}) + f(v,\hat{V}^{KNN})} \quad (6)$$

with f a performance metric and v the true value of the variable. If $R_f = 0$, the two post-processors are equivalent. If $R_f > 0$, the KNN brings an improvement over EMP. Two metrics f were used to compute the score:

$$RMSE(v,\hat{V}) = \sqrt{\sum_j \left(v_j - \hat{V}_{j,25}\right)^2} \quad (7)$$

with j the day from the validation period, v_j the true value of the variable v on this day and $\hat{V}_{j,25}$ the median ensemble. *RMSE* is a deterministic score assessing the performance of the median ensemble only. The second metric is the Continuous Rank Probability Score (*CRPS*) following the definition of Hersbach (2000):

$$CRPS(v,\hat{V}) = \sum_j \sum_{k=0}^{52} \int_{\hat{V}_{j,k}}^{\hat{V}_{j,k+1}} \left[H(w - v_j) - k/52\right]^2 dw \quad (8)$$

with H the Heaviside function, $\hat{V}_{j,0} = -\infty$ and $\hat{V}_{j,52} = +\infty$. The *CRPS* is a probabilistic score comparing the distribution of the ensembles and a distribution centred on the true value v_j.

The two relative scores R_{RMSE} and R_{CRPS} are finally computed for the four variables (streamflow on GS1, GS2 and GS3 and stored volume), the 10 corrupted data sets and the six regions.

RESULTS

Figure 2 presents the relative performance scores for the streamflow at the end of the system (GS3) and the stored volume (VOL) on each region. The box-plots represent the distribution of the scores over the ten corruption scenarios.

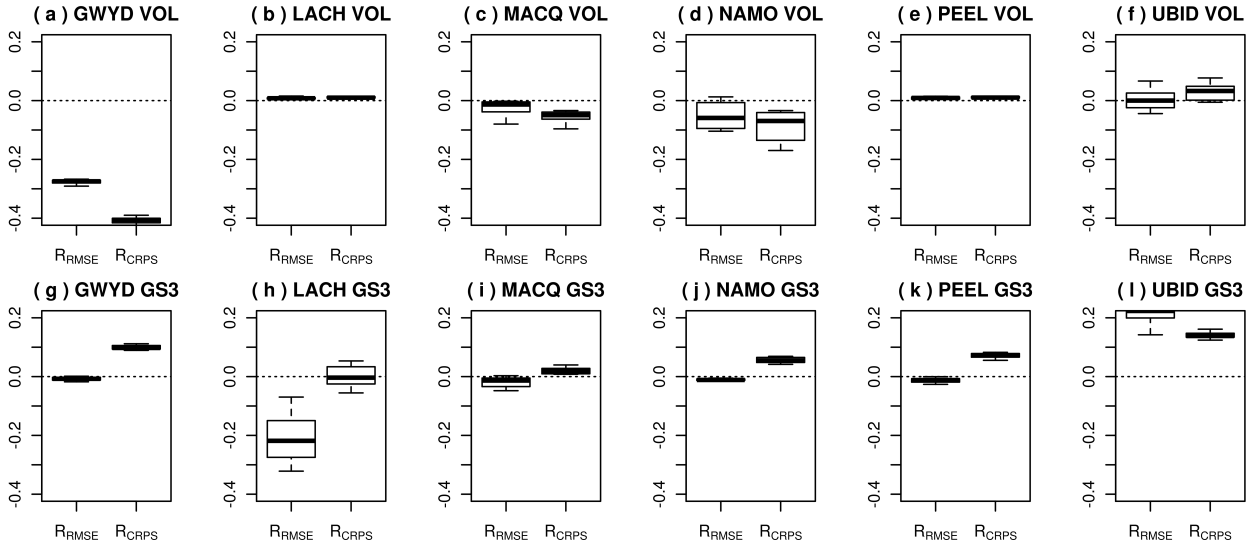

Fig. 2 Relative performance scores R_{RMSE} and R_{CRPS} obtained by comparing the two uncertainty post-processors EMP and KNN in validation mode on the six test regions (GWYD, LACH, MACQ, NAMO, PEEL and UBID) and for two variables (VOL, the stored volume in the reservoir and GS3, the end of system flow).

Figure 2 calls for the following comments:

(a) For the stored volumes (VOL), the two relative scores are mostly negative except in Fig. 2(f). As a result, the KNN did not provide a better predictive uncertainty compared to EMP for this variable.

(b) For the end of system flow (GS3), the R_{RMSE} score remains close to 0, except in Fig. 2(h) where it is systematically negative and Fig. 2(i) where it is systematically positive. Conversely, the R_{CRPS} score is always significantly positive. As a result, the performance of the KNN is not similar if a deterministic or a probabilistic score is used. When a probabilistic score is used, the KNN performs consistently better for this variable.

(c) Some regions (LACH, see Fig. 2(h) and NAMO, see Fig. 2(d)) present a much greater spread in the distribution of the relative scores than others. This suggests that the effects of the streamflow corruption is not identical on all regions and supports the use of a large number of test cases to validate the performance of an uncertainty method.

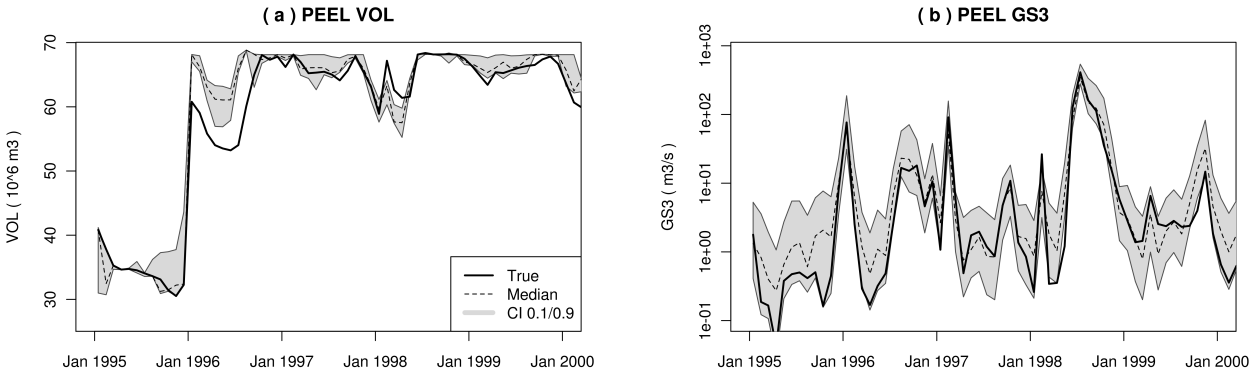

Fig. 3 Time series of the stored volume in the reservoir (a) and the end of system flow (b) for the Peel region. The figure shows the true value (black line), the median simulation (dotted line) and the confidence interval [10–90%] from the uncertainty post-processor (shaded area).

CONCLUSION

This paper compared two methods to generate the predictive uncertainty from deterministic simulations produced by a water allocation model. Both methods do not require intensive computing, which is particularly interesting for a model of this type. The use of the empirical distribution of the residuals (EMP) proved surprisingly powerful compared to a more advanced bootstrapping method (KNN). EMP performed better than KNN when applied to the stored volumes. However, the predictive uncertainty was greatly improved by KNN to assess the uncertainty on streamflow. This highlights the important differences that can occur in the uncertainty analysis for a multivariate model. Further work is needed to merge the strength of both methods and compare them with more advanced uncertainty methods.

REFERENCES

Beven, K. & Binley, A. (1992) The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Processes* **6**(3), 279–298.

CSIRO (2008) Water availability in the Murray-Darling Basin. A report to the Australian Government from the CSIRO Murray-Darling Basin Sustainable Yields Project (regional reports on the Gwydir, Namoi, Macquarie-Castlereagh, Lachlan and Murumbidgee). Technical report, CSIRO, Australia.

Doherty, J. (2010) PEST: Model independent parameter estimation. User Manual 5th edition. Technical report, Watermark Numerical Computing.

Hersbach, H. (2000) Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting* **15**(5), 559–570.

Kavetski, D., Kuczera, G. & Franks, S. W. (2006) Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resour. Res*. 42(3), W03407.

Krzysztofowicz, R. (1999) Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resour. Res.* **35**(9), 2739–2750.

Lall, U. & Sharma, A. (1996) A nearest neighbor bootstrap for resampling hydrologic time series. *Water Resour. Res*. **32**(3), 679–693.

Marquardt, D. W. (1963) An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. for Industrial and Applied Mathematics* **11**(2), 431–441.

Montanari, A. & Brath, A. (2004) A stochastic approach for assessing the uncertainty of rainfall–runoff simulations. *Water Resour. Res*. **40**(1), W01106.

Perrin, C., Michel, C. & Andréassian, V. (2003) Improvement of a parsimonious model for streamflow simulation. *J. Hydrol*. **279**(1–4), 275–289.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M. & Franks, S. (2010) Understanding predictive uncertainty in hydrologic modeling: the challenge of identifying input and structural errors. *Water Resour. Res*. **46**(5), W05521.

Simons, M., Podger, G. & Cooke, R. (1996) IQQM–A hydrologic modelling tool for water resource and salinity management. *Environmental Software* **11**(1–3), 185–192.

Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S. W. & Srikanthan, S. (2009) Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using bayesian total error analysis. *Water Resour. Res*. **45**(12), W00B14.

Vaze, J., Post, D., Chiew, F., Perraud, J., Viney, N. & Teng, J. (2010) Climate nonstationarity – validity of calibrated rainfall–runoff models for use in climate change studies. *J. Hydrol*. **394**(3–4), 447–457.

Vrugt, J., Gupta, H., Bouten, W. & Sorooshian, S. (2003). A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resour. Res*. **39**(8), 1201.