

Time-series model calibrations – influence and estimation of different error types

S.S.H. Kim^a, J.D. Hughes^a, D. Dutta^a and J. Vaze^a

^a *CSIRO Land and Water*
Email: shaun.kim@csiro.au

Abstract: Calibration of time-series models traditionally attempts to reduce the total model prediction error by changing parameters so that simulations match or closely resemble observations. However, a shortcoming of this approach is the assumption that the observational error in the forcing input and output response data is insignificant. In most practical applications this is not the case. In addition, inappropriate model structures may provide good results even though they are actually not suitable representations of the true system of interest. Consequently, calibrated parameters can account for error, resulting in misleading performance and underestimated uncertainty. It is clearly necessary to validate calibrated parameters based on whether they produce realistic errors. Is there a way to estimate the “true” uncertainty in time-series models? Typically, the errors of model simulations are examined as the difference between the simulation output and the corresponding measured values, commonly referred to as the residual error time-series. These residual error time-series encapsulate error from multiple sources, including error in the observed input and output data, and error due to model structure problems and/or poor parameterisation. Estimation of the uncertainty within each of the potential error sources is ideally informed *a priori* but this requires detailed knowledge of the measurements or estimation techniques. In the absence of such information, the residual error time-series can be separated into different source components using data-based techniques. However, this requires the assumption that the nature of the errors are distinct for each of the different components. This study describes and tests a simple method that estimates separately the forcing input error and the output response error in state-dependent models. A major assumption of the method is that errors in the input data produce autocorrelated errors in the output of state-dependent models. In this paper, the assumptions of the method are tested using a simple state-dependent model. The method is outlined and a proof of concept performed using controlled conditions.

Keywords: *Autocorrelation, over-fitting, residual error, state-dependent models*

1. INTRODUCTION

For many environmental managers, computer modelling is a necessary tool for decision making and planning. However, relying on modelling results without consideration of uncertainty is a risky management strategy. Understanding the limitations of the predictive capability of models is essential. Environmental models are usually simplifications of the “real” system. It is generally impractical to model at scales that incorporate every process within a system, which would require unrealistic amounts of data, computer power and system knowledge. Model predictions should therefore be accompanied with uncertainty analyses to present the actual confidence in the results that are presented.

There are multiple sources of uncertainty in the context of environmental modelling. Already mentioned is the inherent uncertainty in the model hypothesis but another is in the observed system inputs and responses (Kavetski *et al.*, 2002). The observational uncertainties can be split into two categories: a) uncertainty in forcing inputs (e.g. rainfall); and b) uncertainty in output responses (e.g. streamflow). It has been shown that observational errors can cause major differences to parameterisations during calibrations (Kavetski *et al.*, 2002; Oudin *et al.*, 2006).

The simplest environmental model calibrations involve optimisation of parameter values so that the output resembles observed data. These include variants of the single-objective (e.g. standard least squares (SLS)) and multi-objective schemes. However, these generally assume appropriate model structures and disregard observational uncertainty. Kavetski *et al.* (2002) performed a demonstration to illustrate the problems. Using the SLS method, they tried to match a model’s simulation output that was run with a pre-determined parameter set. They purposefully corrupted the inputs with log-normally distributed error multipliers, resulting in calibrated parameters that were very different to the “true” values. This could also be referred to as over-fitting, and is when the model is parameterised in such a way that compensates for errors in the data. As a consequence, the model processes might display biases which do not exist in reality, potentially compromising the model’s predictive performance.

A rigorous way of dealing with uncertainty in environmental models is to explicitly articulate it in error models at their respective sources during simulation (Kavetski *et al.*, 2002; Renard *et al.*, 2010). Besides adhering to uncertainty transparency, this would allow more appropriate parameterisations to be identified by preventing over-fitting. However, this would generally require prior knowledge of the nature of the errors for each of the potential sources. The error models might be informed by an analysis of the data acquisition and estimation techniques, for example, the accuracy and precision of measurement devices, or streamflow rating curve estimations (Shao *et al.*, 2014).

Methods within Bayesian frameworks have been developed that allow identification of the various sources of error and their estimation. Kavetski *et al.* (2002) proposes the Bayesian Total Error Analysis (BATEA) framework that requires explicit specification of uncertainty models. Vrugt *et al.* (2005) presents the Simultaneous Optimization and Data Assimilation (SODA) method which calibrates using the ensemble Kalman filter and provides better treatment of input, output, parameter and model structural uncertainties in hydrological modelling. While Bayesian frameworks are theoretically robust for estimating uncertainty in modelling applications, if they are not constrained by any knowledge about the error structure at the different sources, then estimates of uncertainty may become prohibitively large.

In the absence of information about input and output data collection and processing, there are data-based techniques which might allow the estimation of errors from each source. For example, previous studies have explored this for output response error. These methods usually involved differencing the residual error time-series to give estimates of the error deviation, and assume that the output errors are not autocorrelated (Hall *et al.*, 1990; Vrugt *et al.*, 2005).

In this paper, a simple, efficient and systematic way of estimating the error at their sources is described and tested. This procedure makes use of basic data-based techniques. The method is designed to be used in existing Bayesian frameworks. If prior uncertainty information is available, then the new method could be used to validate existing error models or possibly used to improve them. The method is intuitive and could be applied in a wide range of modelling applications.

The proposed method relies on some critical assumptions. These are explained in section 2, followed by a description of the new procedure. A proof of concept is given using a simple state-dependent model in section 3. Finally, the results are interpreted, and limitations and benefits of the new method discussed in section 4 with concluding remarks.

2. METHODS

2.1. Assumptions

In the current study, the assumptions are based upon conceptual hydrological models, however, these could apply more generally to state-dependent time-series models. For the purposes of demonstration, it is assumed that the given model structure, unless otherwise stated, is perfect. That is, the model is an appropriate representation of the system of interest. Also assumed is the notion of “true” parameters, inputs and outputs. Meaning that a simulation using true parameters and true input data will produce the true output. In practice, the true values are not known in calibrations. The input used for modelling follows:

$$I = I_{true} + \varepsilon_I \quad (1)$$

where I is the input data, I_{true} is the true input data and ε_I is the input error. Similarly, the observed output data is given by:

$$O = O_{true} + \varepsilon_O \quad (2)$$

where O is the observed output data, O_{true} is the true output and ε_O is the observed output error. A simulation of the true model can be represented by:

$$O_{true} = \eta(\theta_{true}, I_{true}) \quad (3)$$

where θ_{true} is the true parameters and η represents the simulation of the model. I and O are often used in the place of I_{true} and O_{true} during calibrations, however, this is known to be incorrect. When we incorporate equations 1 and 2 into equation 3, the resulting equation becomes:

$$O - \varepsilon_O = \eta(\theta_{true}, I - \varepsilon_I) \quad (4)$$

Finally, the equation contains three unknowns that need to be estimated: ε_I , ε_O and θ_{true} . In reality, there is also a high degree of uncertainty surrounding the correctness of the model structure or due to stochasticity of model processes. These can potentially be grouped together with input uncertainty and is discussed in section 4. The input uncertainty is not explicitly distinguished from model uncertainty in the current study, and so for the sake of simplicity, the model uncertainty is assumed to be negligible.

Observed output response error

The output error is assumed to be additive, randomly distributed and homoscedastic. The output errors are assumed to have a mean of zero, however, the standard deviation is not assumed. The output error is directly revealed in the residual error time-series, that is, the error isn't processed before it can be seen in the output. Since the response error signal is considered random and is unprocessed, it can be assumed that it will appear as a component that is not autocorrelated in the residual error time-series.

The assumptions of error are relatively simple but are considered appropriate by the authors for the purposes of demonstration. Assumptions in the current study may not be relevant to many systems, and so should be carefully considered. For example, in hydrology, it is often observed that response errors are heteroscedastic, which is an artefact of the rating curve function and extrapolation (Vrugt *et al.*, 2005). Heteroscedasticity is not considered in the current study, but the method could incorporate it in more sophisticated applications. It should be noted that the assumptions in the method could be tailored to different systems.

Forcing input error

The forcing input error is assumed to have similar characteristics as the output response error, that is, additive, randomly distributed and homoscedastic. The input error is also assumed to have a mean of zero with unknown standard deviation. There is, however, an assumption that the input error has a Gaussian distribution. In addition, the input error is processed by the model before seen in the residual error time-series. This results in input errors appearing as autocorrelated signals in state-dependent models even when the input isn't autocorrelated itself (Kavetski *et al.*, 2002). This is due to the “memory” of the model states that propagate input errors and cause them to be spread over multiple time-steps. For example, in a hydrological model, the over-estimation of rainfall in a time-step may cause an over-estimation of volume within a soil moisture store, in turn causing over-estimation of outflow in subsequent time-steps.

The degree of autocorrelation of the residual time-series caused by the input error will depend on the quickness of the model's response to the input. If the effects of input errors are confined to one time-step, then any input noise will not appear as autocorrelated in the residual error time-series. This might be seen in hydrological models of steep, smooth and hard terrain environments where input produces flashy responses.

In previous studies input errors have been assumed to be multiplicative (Kavetski *et al.*, 2002; McMillan *et al.*, 2011). This is a reasonable assumption, however, modelling additive errors were chosen in the current study due to its ease of interpretation. That being said, it may be possible to incorporate multiplicative input errors in future applications.

2.2 Input error estimation (IEE)

The IEE method backward estimates the potential input error from the residual error time-series. The output response error needs to be calculated before input error can be estimated. The step-wise procedure relies on the assumptions outlined in section 2.1. The steps are:

1. Simulate the model and compute the residual error time-series;
2. From the residual error time-series, estimate a range of plausible values for output response error standard deviation using a block bootstrap method;
3. Optimise a smoothing algorithm performed on the residual error time-series so that the standard deviation of original minus smoothed residuals match the plausible standard deviations calculated in step 2;
4. For each match in step 3, compute the standard deviation of the smoothed residual. This gives the range of impact that the input error has on the residual error time-series;
5. Add an error model to the input of the original model and set it up so that the original simulation is taken as the observed output response data;
6. Use the set-up from step 5 to search for input error standard deviations that give residual error standard deviations within the range given in step 4 (i.e. the range of standard deviations of the original residual after each smoothing).

To estimate the output response error standard deviation in Step 2 the following equation from Vrugt *et al.* (2005) was used:

$$\sigma_{\epsilon_o} = \sqrt{\frac{1}{2 \cdot (n-1)} \sum_{t=2}^n (\epsilon_{O,t} - \epsilon_{O,t-1})^2} \quad (5)$$

where n is the total number of time-steps. This equation was used in a block bootstrap that randomly samples blocks of data with replacement. The block bootstrap allows estimation of the distribution of the statistic while preserving autocorrelation in the data. A local polynomial regression fitting (LOESS) method was used for smoothing in step 3, where the span parameter was adjusted to control the degree of smoothing. A simple sweep search was used in step 6 to thoroughly sample a large range to find the potential values for input error standard deviation.

3. PROOF OF CONCEPT

3.1 Simple state-dependent model (SimpMod)

A simple two-parameter time-series model was used to illustrate the basic workings of the proposed method. The structure of the model is shown in Figure 1. The model represents a very basic conceptual rainfall-runoff model. Mass is conserved and the model does not allow negative values for input, output and stores.

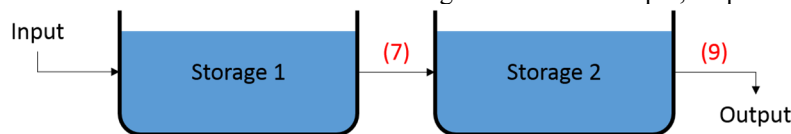


Figure 1. A simple conceptual model structure (SimMod) designed for demonstration purposes. The text in red refer to the equation numbers.

The model is controlled by two state stores that are update at each time-step. There is a single input for each time-step, which is added to the first store:

$$V_{S_1,t} = V_{S_1,t-1} + I_t \quad (6)$$

where $V_{S_1,t}$ is the volume of the first store S_1 at time-step, t . I represents the input time-series. The outflow from the first store is then calculated:

$$Q_{S_1,t} = \alpha \cdot V_{S_1,t} \quad (7)$$

where Q_{S_1} denotes the outflow for the first store. α is a parameter between 0 and 1 that controls the outflow from the first store. The outflow from the first store is added to the second:

$$V_{S_2,t} = V_{S_2,t-1} + Q_{S_1,t} \quad (8)$$

where S_2 denotes the second store. The calculation of the final outflow is similar to equation 7, given by:

$$Q_{S_2,t} = \beta \cdot V_{S_2,t} \quad (9)$$

where β is a parameter between 0 and 1 that controls the outflow from the second store. Equations 6 to 9 are performed in sequence for each time-step.

3.2 Testing method

A synthetic input time-series (I_{true}) of length 200 was created by random sampling from a normal distribution with a mean of 1000 and standard deviation of 6. Simulating SimpNet with this input and predetermined parameters ($\alpha = 0.2$, $\beta = 0.3$), an output time-series was created, to give the true output time-series (O_{true}). The input error time-series was generated by random sampling from a normal distribution:

$$\varepsilon_I = N(\mu_{\varepsilon_I}, \sigma_{\varepsilon_I}^2) \quad (10)$$

where μ_{ε_I} and σ_{ε_I} are the mean and standard deviation of the input error, respectively. Similarly, the output error time-series was sampled from a normal distribution, given by:

$$\varepsilon_O = N(\mu_{\varepsilon_O}, \sigma_{\varepsilon_O}^2) \quad (11)$$

where μ_{ε_O} and σ_{ε_O} are the mean and standard deviation of the output error, respectively. For the proof of concept, $\mu_{\varepsilon_I} = 0$, $\sigma_{\varepsilon_I} = 2$, $\mu_{\varepsilon_O} = 0$ and $\sigma_{\varepsilon_O} = 0.5$.

Equations 10 and 11 were used to corrupt the true input and output data. Using these and the IEE method, an attempt was made to obtain accurate estimates of the standard deviations for input and output error. This was performed using the true parameters and using standard least squares (SLS) calibrated parameters based on corrupted data. The testing method was repeated 100 times to ensure positive results weren't fortuitous.

3.3 Results

The IEE analysis results are shown in Table 1. The method estimates the range of both input and output error standard deviations when using the true parameters. To clarify, error standard deviations associated with the word "true" are error standard deviations using true parameters. The SLS calibrated parameters were very similar to the true parameters (Figure 2). Despite this, the estimated input error standard deviations were lower than that given by the true parameters, but in general, the actual standard deviation was still within the estimated range (Figure 3 and 4). A one-sided paired Wilcoxon signed-rank test on the means of the input error standard deviations gave a p -value of 1.4×10^{-9} , indicating that the calibrated parameters gave significantly lower estimates for input error standard deviation than the true parameters.

Time-series simulations using the input error models are shown in Figure 5. These are overlaid by the observed output response time-series and their estimated error ranges. The underestimation of input error for calibrated parameters can be seen where the output response time-series diverge from the simulations and exceeds its bounds (seen in red). In general, the observed output covers a larger area of the simulation for the calibrated parameters than the true parameters.

4. DISCUSSION AND CONCLUSIONS

When calibrated parameters were used in the IEE method, input errors were underestimated compared to the true parameters. This is a result of over-fitting. If model structure and parameters are correct, then the output response should be contained within the model results simulated with forcing input errors. If it is not, then this suggests that the model is unable to adequately reproduce all possible observed output responses. This can be seen in Figure 5, where simulations better contain the output responses for the true parameters than the calibrated parameters (seen as more red in Figure 5b than Figure 5a). The IEE method has potential to be a useful validation tool for identifying over-fitting of parameter sets and model structures. It also shows promise to be useful during calibration, however, practicality is a challenge since performing the method many times will require considerable computation.

Table 1. Error standard deviation estimates for true and calibrated (calib) parameters.

	Standard deviation mean		Standard deviation 1st percentile		Standard deviation 99th percentile		Standard deviation actual
	true	calib	true	calib	true	calib	
Output error	0.51	0.52	0.47	0.47	0.58	0.58	0.50
Input error	1.81	1.76	1.02	1.02	2.63	2.54	2.0

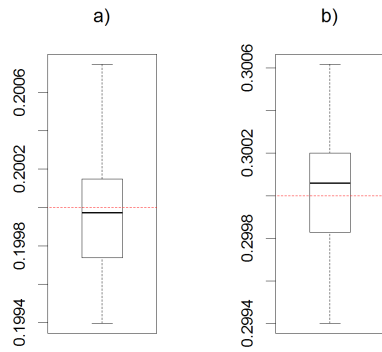


Figure 2. Calibrated parameters for a) α ; and b) β . The red dashed lines show the true parameter values.

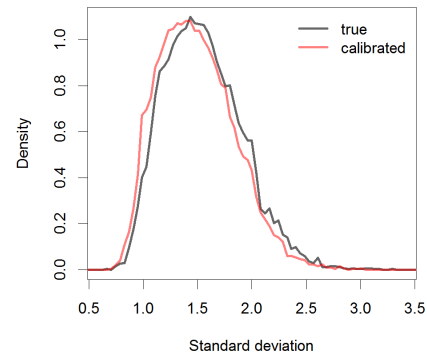


Figure 3. Probability distributions of input error standard deviations (one from the 100 repetitions).

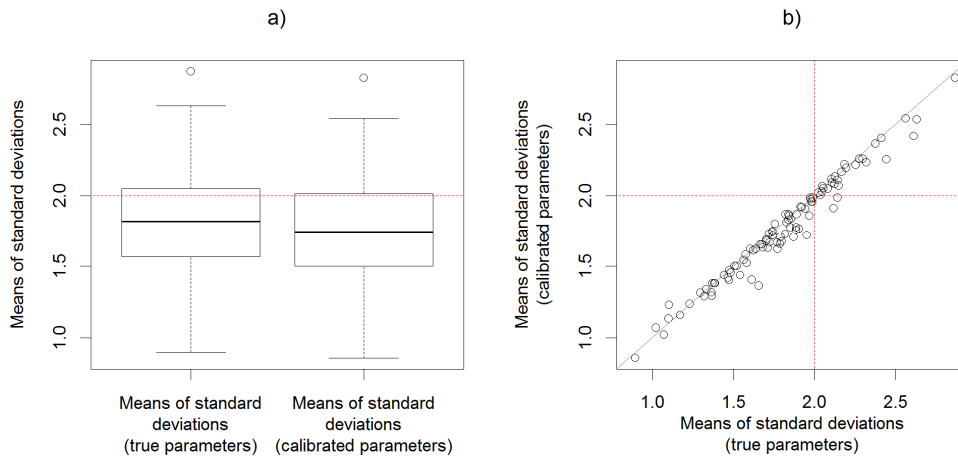


Figure 4. Means of input error standard deviations shown as a) boxplots; and b) a scatterplot. The red dashed lines indicate the actual standard deviation of the input error.

The method proposed in this paper relies on the model structure being an appropriate representation of the system being modelled. If this is not the case then the autocorrelated component will represent both the input error and the error due to inadequate model structure. The input error will be indistinguishable from model structure error. One of the benefits of the method is that it allows one the ability to verify any prior error estimates that may have been performed by, say, analysis of measurement procedures. If the IEE overestimates the input error compared to more direct estimates, then the excess error might be attributed to model structure problems.

While currently the IEE method is able to estimate error for only a limited number of variables, future work could involve differentiating the nature of the error sources in more detail. However, it would be difficult to develop a generic approach since intricate assumptions about the structure of errors are likely to be tied to specific model structures and/or data types. Also, it is likely that the results of the IEE method will depend on the ratio of real signal to noise, which needs to be explored in future research.

The IEE method currently requires knowledge of the error model structure and at this stage only estimates error standard deviation. An important assumption is that the error mean is known. This is assumed to be zero in the current study, however, in reality this might be difficult to determine. Inaccurate error mean estimates would have significant impact on the results, and so, in practice, should be determined through careful analyses of the observed data collection and processing methods.

The model structure of SimpMod is reminiscent of a river storage routing model, where both input and output are observed streamflow data. Also, the error models used in the current study are in line with typical observed streamflow data assumptions, that is, the streamflow errors are normally distributed and additive, albeit without heteroscedasticity. The method proposed in this paper would not require much adjustment to be implemented for river system models. However, further adjustment might be required to adapt it to rainfall-runoff models to account for inherent differences in rainfall input error compared to streamflow. For example, rainfall errors are commonly considered multiplicative rather than additive (McMillan *et al.*, 2011).

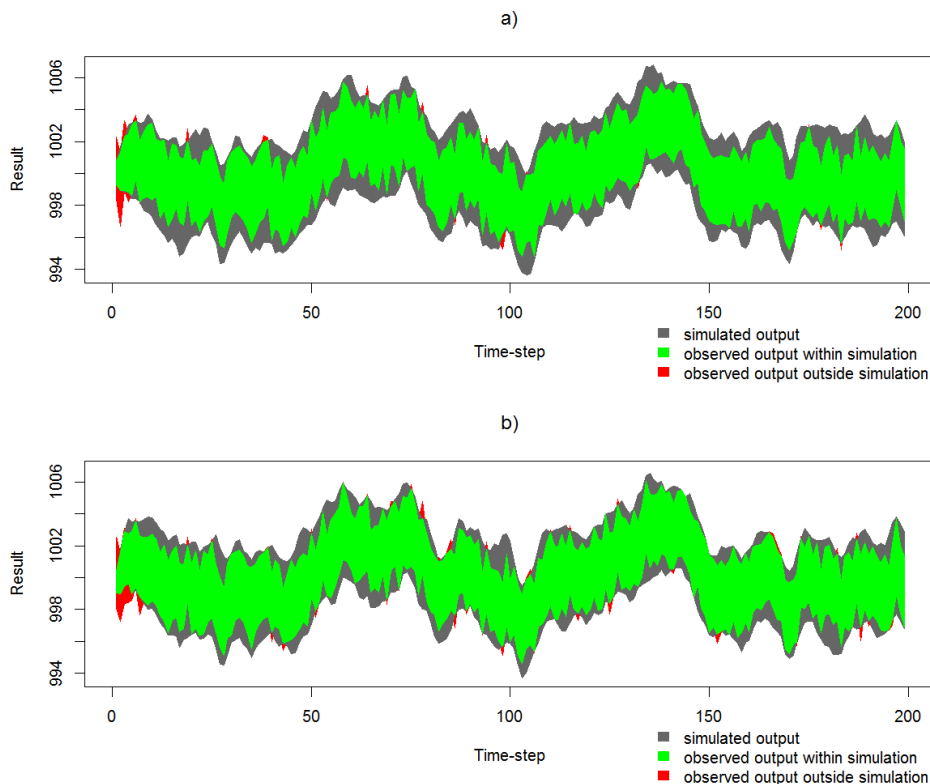


Figure 5. Time-series simulations and output responses using a) true parameters and b) calibrated parameters. Time-series are simulated with their corresponding input error model (max. standard deviation). The output response time-series are also displayed with their associated errors (max. standard deviation).

REFERENCES

- Hall, P., Kay, J.W. and Titterton D.M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression, *Biometrika*, 77, 521-528.
- Kavetski, D., Franks, S.W. and Kuczera, G. (2002). Confronting input uncertainty in environmental modelling. In *Calibration of watershed models*, edited by Q. Duan, H. Gupta, S. Sorooshian, A.N. Rousseau, and R. Turcotte, AGU Books, Washington, 49-68.
- McMillan, H., Jackson, B., Clark, M., Kavetski, D. and Woods, R. (2011). Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. *J. Hydrol.*, 400, 83-94.
- Oudin, L., Perrin, C., Mathevet, T., Andréassian, V. and Michel, C. (2006). Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. *J. Hydrol.*, 320, 62-83.
- Renard, B., Kavetski, D., Kuczera, G., Thyer, M. and Franks, S.W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, 46(5).
- Shao, Q., Lerat, J., Podger, G. and Dutta, D. (2014). Uncertainty estimation with bias-correction for flow series based on rating curve. *J. Hydrol.*, 510, 137-152.
- Vrugt, J.A., Diks, C.G., 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).