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

Representing uncertainty in objective functions: extension to include the influence of serial correlation

Croke, B.F.W.

[1] *The Department of Mathematics, and Integrated Catchment Assessment and Management (iCAM) Centre, The Fenner School of Environment and Society, The Australian National University, Australian Capital Territory*
Email: barry.croke@anu.edu.au

Abstract: The role of performance indicators is to give an accurate indication of the fit between a model and the system being modelled. As all measurements have an associated uncertainty (determining the significance that should be given to the measurement), performance indicators should take into account uncertainties in the observed quantities being modelled as well as in the model predictions (due to uncertainties in inputs, both data and model parameters). In the presence of significant heteroscedasticity in the uncertainty in observed and modelled output of a system, failure to adequately account for variations in the uncertainties means that the objective function only gives a measure of how well the model fits the observations, not how well the model fits the system being modelled. Since in most cases, the interest lies in fitting the system response, it is vital that the objective function(s) be designed to account for any variations in the uncertainties.

Most objective functions (e.g. those based on the sum of squared residuals) assume homoscedastic uncertainties. If model contribution to the variations in residuals can be ignored, then transformations (e.g. Box-Cox) can be used to remove (or at least significantly reduce) heteroscedasticity. An alternative which is more generally applicable is to explicitly represent the uncertainties in the observed and modelled values in the objective function. Previous work on this topic addressed the modifications to standard objective functions (Nash-Sutcliffe efficiency - NSE, Root Mean Square Error - RMSE, chi-squared, coefficient of determination) using the optimal weighted averaging approach. A Monte Carlo trial using synthetic data with known uncertainty in the rating curve shows that the modified NSE gives significantly lower uncertainty in the estimated parameter values compared to those derived using the standard NSE.

In addition to the heteroscedasticity in the uncertainties, there may also be significant serial correlation in the uncertainties for different time steps as a result of the use of a rating curve in estimating the observed flows, and for modelled flows, the influence of system memory (propagation of input uncertainties through the model). This includes not only first-order serial correlation (correlation between the uncertainties of the values for neighbouring time steps), but also longer-term serial correlation induced through the uncertainty in the rating curve. The extent of the longer-term serial correlation will depend on the stability of the rating curve. Thus, the requirement for exploiting the long-term serial correlation is information on the uncertainty and stability in the rating curve, information which is not necessarily available at the current time. This requires a change in the way that streamflow databases are constructed. A modified form of the NSE has been proposed, and the effect of including first-order serial correlation tested against synthetic data. The result of a Monte Carlo trial shows significant reduction in parameter uncertainty when first-order serial correlation is included in the objective function.

Keywords: *uncertainty, objective function, calibration*

1. INTRODUCTION

Objective functions give a measure of the difference between observed and modelled values, and therefore can be used to both calibrate a model and test the performance under simulation. Most objective functions assume that the uncertainties in the observed and modelled values are homoscedastic (i.e. the uncertainty doesn't vary through the data set). Unfortunately, the assumption is often not valid, and is certainty invalid for hydrological datasets. The result is that the commonly used objective functions; for example, Nash Sutcliffe Efficiency (NSE) and Root Mean Square Error (RMSE) are poor measures of a models performance. Analysis of the impact of heteroscedasticity in the uncertainties using the optimal combinations of observations (Aitken, 1935) shows that the significance given to each data point is proportional to the inverse of the uncertainty squared. That is, very high weight is given to the most uncertain data points – independently of whether the uncertainty is known (Croke, 2007; Croke *et al.* 2008).

One approach for addressing the heteroscedasticity in the uncertainties is through the Box-Cox transformation (Box and Cox, 1964). This approach is based on the assumption that the scatter in the residuals (observed – modelled) is a good indicator of the uncertainties in the residuals. Providing this assumption is valid, this approach gives a very simple means of removing the heteroscedasticity, thereby ensuring that the objective functions give good measures of the model performance when the transformed values are used instead of the original values.

However, if error in the model structure introduces scatter in the model residuals, or if over-parameterisation reduces the scatter, then while the Box-Cox transformation will give homoscedastic residuals, there may still be significant heteroscedasticity in the uncertainty of the residuals (i.e. the scatter in the residuals is not only due to uncertainty in observed and modelled values). This is a result of the scatter in the model residuals not necessarily being a good indicator of the uncertainty in the residuals as any contribution from the model structure (if this does not adequately represent the system being modelled) is not included in the uncertainty in the modelled flows. This limits the effectiveness of the Box-Cox transformation to applications of parsimonious models which can be shown to have the correct structure. For other models, another approach is needed.

Sorooshian and Dracup (1980) proposed an objective function (Homoscedastic Maximum Likelihood Estimator - HMLE) which attempts to account for the uncertainty in the residuals, and is very similar in form to the modified NSE proposed by Croke (2007). In using the HMLE, the estimation of the uncertainties is usually done by assuming some relationship between the uncertainty and the observed values (e.g. a power law). The effectiveness of this approach depends on how accurately the assumed functional form matches the heteroscedasticity in the uncertainty of the residuals, and the accuracy in the fitted coefficients of that functional form (if the coefficients are estimated based on model residuals, then comments for Box-Cox transformation also apply here). Hill and Tiedeman (2007) give a detailed analysis of calibration of groundwater models, including derivation of weights based on the uncertainties.

In the case of observed stream flow, the uncertainty in the flow values depends primarily on the uncertainty in the rating curve (and to a lesser extent on the uncertainty in the measurement of the river level). The uncertainty in the rating curve results from uncertainty in the form of the functions used to fit the observed data, the discharge and stage measurement error during gauging, and potentially unaccounted drivers (e.g. the possible influence due to rate of change in stage on discharge). This paper will focus on the impact of uncertainty in the rating curve on the estimated streamflow, and consequently on the measurement of model performance. In the comparison of observed and modelled flows, it is also important to consider the uncertainty in the modelled values. This depends on the uncertainty in the model inputs and the model parameters. Uncertainty in the model structure (i.e. whether the model structure is appropriate) is not considered in the calculation of the performance indicator. Rather, this component of model uncertainty should be reflected in the value of the performance indicator (poor values indicate that the model structure could be improved). The expressions developed in this paper involve the uncertainty in both the observed and modelled values, though the synthetic examples given are focused on the uncertainty in the observed values.

2. GLOSSARY OF SYMBOLS AND TERMS

2.1. Terms

true value
the actual value (by definition unknown) that is being measured or calculated

uncertainty in a measurement or calculation
probability distribution of the true value. Multiple observations can be used to give an estimate of the random component of the uncertainty, but no information regarding any systematic error will be obtained. Consequently, the uncertainty derived from multiple observations is a lower limit to the uncertainty in the measurement. Uncertainty is represented by a leading Δ (e.g. Δx is the uncertainty in x).

error in a measurement or calculation
difference between the measured/calculated value and the true value

2.2. Symbols

- R^2: Nash Sutcliffe Efficiency
- $q_{o,i}$: observed flow at time step i
- $q_{m,i}$: modelled flow at time step i
- $\bar{q}_o$: mean observed flow
- $d_{o,i,j}$: $q_{o,i}$ - $q_{o,j}$,
- $d_{m,i,j}$: $q_{m,i}$ - $q_{m,j}$
- ω_i: weight applied to flow residual for time step i
- $\omega_{s,i,j}$: weight applied to numerator of long-term serial correlation term between time steps i and j
- λ_i: weight applied to observed flow for time step i
- $\lambda_{s,i,j}$: weight applied to denominator of long-term serial correlation term between time steps i and j
- s_i: significance (subjective) given to time step i

3. EXISTING OBJECTIVE FUNCTIONS

The Nash Sutcliffe Efficiency (NSE) is the most commonly used objective function in hydrology.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(q_{o,i} - q_{m,i}\right)^2}{\sum_{i=1}^{n} \left(q_{o,i} - \bar{q}_o\right)^2} \tag{1}$$

Using an optimal weighted average approach (Aitken, 1935), the NSE (and other objective functions) can be modified to account for uncertainties in modelled and observed values (Croke, 2007)

$$R^2 = 1 - \frac{\sum_{i=1}^{n} s_i \omega_i \left(q_{o,i} - q_{m,i}\right)^2}{\sum_{i=1}^{n} s_i \lambda_i \left(q_{o,i} - \bar{q}_o\right)^2} \tag{2}$$

where the weights ω_i and λ_i are obtained from the estimated uncertainties in the modelled ($\Delta q_{m,i}$) and observed ($\Delta q_{o,i}$) flows:

$$\omega_i = 1 \Big/ \left[\left(\Delta q_{o,i}\right)^2 + \left(\Delta q_{m,i}\right)^2\right], \qquad \lambda_i = 1 \Big/ \left(\Delta q_{o,i}\right)^2$$

and s_i is the significance given to each time step. Comparison of Eqns 1 and 2 show that the NSE:

- assumes negligible uncertainty in modelled values
- significance (s_i) given to each time step is proportional to the square of the uncertainty in the observed flow (significance is uniform across all time steps only if the uncertainty is homoscedastic).

4. SERIAL CORRELATION

Standard objective functions (e.g. NSE, RMSE) simply compare each observed and modelled value, producing an aggregated signal from the entire dataset. That is, they ignore serial correlation in the data. However, due to the use of a rating curve to derive the observed streamflow estimates, there is significant serial correlation in the observed flows and their associated uncertainty, not just between neighbouring time steps, but on much longer scales. The limit on the separation between any two observations in terms of the

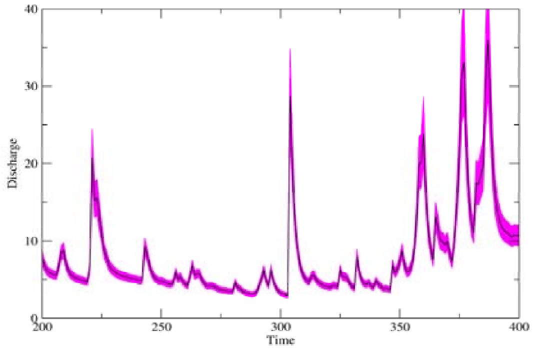

Figure 1: Observed flow, with uncertainty bounds in pink (uncertainty based on 10% error in exponent (b) and coefficient (a) of rating curve)

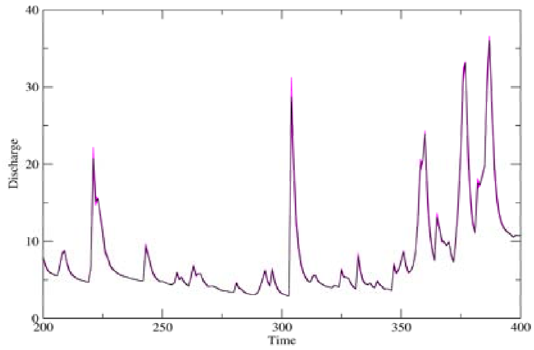

Figure 2: Observed flow, with uncertainty bounds in pink (Uncertainty bounds assuming previous time step error free – i.e. showing the significantly reduced uncertainty in the local slope of the hydrograph compared with the uncertainty in the flow in Figure 1)

serial correlation between them is determined by the stability of the rating curve.

Figures 1 and 2 show the impact of first-order serial correlation (serial correlation between nearest neighbours) on the uncertainties in observed streamflow measurements. Synthetic streamflow "data" have been generated, using a power law form of rating curve [$Q = a\ (h\text{-}h_0)^b$] , and including a 10% uncertainty (uniform distribution) in a and b. Figure 1 shows the uncertainty for each timestep, while Figure 2 shows the uncertainty in the flow assuming that the previous flow value is accurate (thus representing the error in the slope of the hydrograph). The significantly lower uncertainty in the relative flow (flow with respect to the flow for the previous time step) shows that there is a significant amount of information in the data that the standard objective functions do not exploit.

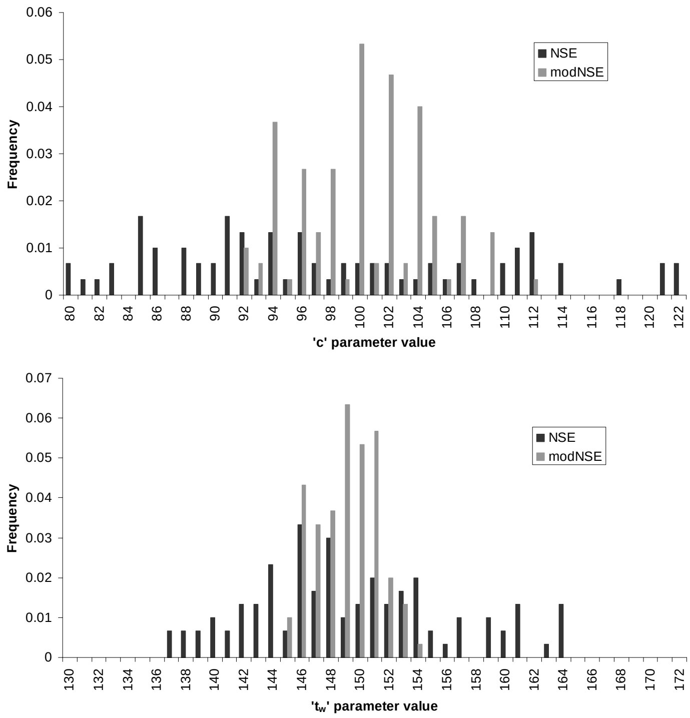

Figure 3. Calibrated parameter values for a Monte Carlo experiment on the impact of rating curve uncertainty on model parameters

A synthetic time series of data was generated using the IHACRES rainfall-runoff model, with the uncertainty in the synthetic data being introduced by perturbing an assumed rating curve of the form described above, with an uncertainty in b of 10%, and a 5mm uncertainty in h. The a parameter is ignored as this is a trivial case when applied to the IHACRES model as the uncertainty in a will map linearly to the c parameter (mass balance term) in IHACRES. Similarly, the uncertainty in the parameter h_0 is also not considered. The impact of the modified NSE on the calibration of two of the IHACRES model parameters is shown in Figure 3. Clearly, the modified NSE results in a significant decrease in the uncertainty in the estimated parameter values.

4.1. Short-range autocorrelation

The short-range autocorrelation (serial correlation between adjacent time steps, or first-order serial correlation) in uncertainty can be handled by comparing the rate of change (or slope) of the observed and modelled values. Bai et al. (2009) used a slope-based version of the NSE as one criterion in a multi-criteria calibration and testing of models. This can easily be achieved by using the change in value compared with the previous time step. The influence of short-range autocorrelation can be combined with the weighted NSE (Eqn 2) to give:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} s_i \left[\omega_s \left(q_{o,i} - q_{m,i}\right)^2 + \omega_{s,i,i-1}\left(d_{o,i,i-1} - d_{m,i,i-1}\right)^2\right]}{\sum_{i=1}^{n} s_i \left[\lambda_i \left(q_{o,i} - \bar{q}_o\right)^2 + \lambda_{s,i,i-1}\left(d_{o,i,i-1} - \left(\frac{q_{o,1} - q_{o,n}}{n}\right)\right)^2\right]} \tag{3}$$

where

$$d_{o,i,i-1} = q_{o,i} - q_{o,i-1}, \quad \omega_{s,i,i-1} = \left(\Delta d_{o,i,i-1}\right)^2 + \left(\Delta d_{m,i,i-1}\right)^2, \qquad \lambda_{s,i,i-1} = \left(\Delta d_{o,i,i-1}\right)^2$$

and $(q_{o,1}-q_{o,n})/n$ is the mean value of $d_{o,i,i-1}$. The uncertainty in the value of $d_{o,i,i-1}$ cannot be calculated from the uncertainties in the two flow values $q_{o,i}$ and $q_{o,i-1}$. Rather, the uncertainty is obtained by considering the uncertainty in the rating curve. Ignoring correlation between the rating curve's parameter values, we get

$$\left(\Delta d_{o,i,i-1}\right)^2 = \left(\frac{\partial d_{o,i,i-1}}{\partial a}\Delta a\right)^2 + \left(\frac{\partial d_{o,i,i-1}}{\partial b}\Delta b\right)^2 + \left(\frac{\partial d_{o,i,i-1}}{\partial h_0}\Delta h_0\right)^2 + \left(\frac{\partial d_{o,i,i-1}}{\partial h_i}\Delta h_i\right)^2 + \left(\frac{\partial d_{o,i,i-1}}{\partial h_{i-1}}\Delta h_{i-1}\right)^2 \tag{4}$$

Similarly, for the modelled flow, the uncertainty in the slope cannot be derived from the uncertainty in the two modelled flow values, but must be calculated independently, taking into consideration the propagation of uncertainty through the model, and the influence of the system memory.

The objective function could also be formulated as a separate term added to eqn (2)

$$R^2 = 1 - \frac{\sum_{i=1}^{n} s_i \left[\omega_s \left(q_{o,i} - q_{m,i}\right)^2\right]}{\sum_{i=1}^{n} s_i \left[\lambda_i \left(q_{o,i} - \bar{q}_o\right)^2\right]} - \frac{\sum_{i=1}^{n} s_i \left[\omega_{s,i,i-1}\left(d_{o,i,i-1} - d_{m,i,i-1}\right)^2\right]}{\sum_{i=1}^{n} s_i \left[\lambda_{s,i,i-1}\left(d_{o,i,i-1} - \left(\frac{q_{o,1} - q_{o,n}}{n}\right)\right)^2\right]} \tag{5}$$

Note that in both forms, the relative weighting between the value and slope terms is automatically handled by uncertainty in the observed and modelled values/slopes. The adopted form used here is that of equation 3.

Figure 4 shows the impact of including the short-term serial correlation on the calibrated model parameters. For both parameters, there is a significant decrease in the uncertainty in the estimated values when the first-order serial correlation is included in the objective function.

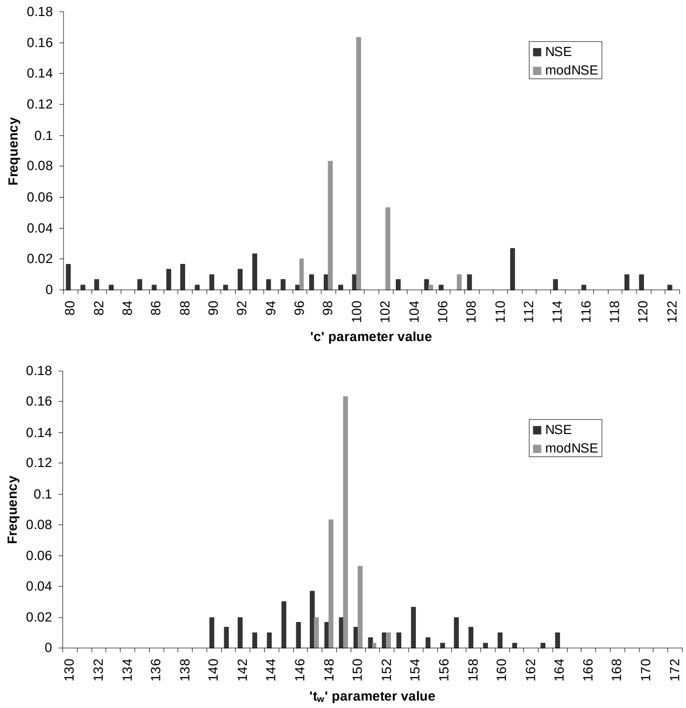

Figure 4. Calibrated parameter values when first-order serial correlation is taken into account

4.2. Long-term autocorrelation

There can also be long-term autocorrelation (correlation between data points more than 1 timestep apart) in the uncertainty in the observed data. An example is streamflow data, where the observation of the water level is converted into a discharge using a rating curve (mostly empirically derived). Providing that the rating curve is stable over the period in question, then the relative uncertainty of 2 observations with comparable flow depths is significantly smaller than the uncertainty in either observation. This means that there is additional information in the data that can be exploited to assess the model's performance, potentially leading to better discrimination between models, and more accurate estimation of model parameter values. Extending equation (3) to account for long-term autocorrelation where the rating curve is stable over the entire period leads to:

$$R^2 = 1 - \frac{\sum_i^n s_i \left\{ \omega_s \left(q_{o,i} - q_{m,i} \right)^2 + \sum_{j>i} \left[\omega_{s,i,j} \left(d_{o,j,j-i} - d_{m,j,j-i} \right)^2 \right] \right\}}{\sum_i^n s_i \left\{ \lambda_s \left(q_{o,i} - \bar{q}_o \right)^2 + \sum_{j>i} \left[\lambda_{s,i,j} \left(d_{o,j,j-i} - \overline{d_{o,j}} \right)^2 \right] \right\}} \tag{5}$$

This means that not only is the uncertainty in the rating curve needed, but also information on the stability of the rating curve. This additional information should result in a more stringent test of a models performance, and hence better model parameters and discrimination between models.

5. APPLICATION OF MODIFIED NSE

The calculation of the weights in the modified NSE can vary between calibration and simulation, depending on the calibration technique used. In simulation, the uncertainties in the model input data and parameter values each contribute to the uncertainty in the modelled output. Uncertainty due to how well the model structure represents the system being modelled is not considered in calculating the weights. Strictly speaking, there is no uncertainty in the model structure, unless there are coding errors present. Rather, this aspect of model uncertainty is reflected in the value of the objective function (a poor model structure should result in a poor result from the objective function).

Most calibration techniques involve testing the model performance for a given set of parameter values, and on the basis of the results, attempting to find either a local or global minimum. Consequently, there is no uncertainty in the model parameter values used in this process, and so the uncertainty in the modelled values is derived from the uncertainties in the input data only. If during the calibration process, the parameter values are estimated rather than set (e.g. when using the Simple Refined Instrumental Variable algorithm in IHACRES, the parameter values of the linear module are estimated for a given set of values for the non-linear module parameters), then the uncertainty in the estimated parameter values needs to be considered. Note that the resulting error accumulation may limit the usefulness of the long-term serial correlation in the observed data.

6. CONCLUSIONS

Most objective functions compare observed and modelled values under the assumption of homoscedastic uncertainties, and negligible serial correlation. Both these conditions are not met when modelling streamflow (and in many other circumstances) due to the use of a rating curve to estimate the observed streamflow, and the influence of system memory. Where there is significant variation in the uncertainties through a dataset, the variation in the uncertainty needs to be taken into account when measuring model performance. Since the uncertainty estimates are being used in calculating the value of the objective function, the accuracy of the uncertainty estimate can influence the result. Depending on the type of performance indicator being used, this may require just the relative uncertainty (magnitude of uncertainty may not be known, but the variability needs to be known), though some indicators may need the absolute uncertainty.

Further improvements in the ability of an objective function to discriminate between good and poor models can be made by considering the influence of serial correlation where this is significant. This may include just first-order serial correlation (correlation between nearest neighbours), or in some cases (e.g. rainfall-runoff models) can include longer-term serial correlation (correlation between well separated observations). In the case of rainfall-runoff models, this requires information on the stability of the rating curve through time, as well as the uncertainty in the fitted rating curve.

REFERENCES

Aitken, A.C. (1935), On Least Squares and Linear Combination of Observations, *Proceedings of the Royal Society of Edinburgh*, 55, 42-48.

Bai, Y., Wagener, T. and Reed, P. (2009), A top-down framework for watershed model evaluation and selection under uncertainty, *Environmental Modelling and Software*, doi:10.1016/j.envsoft.2008.12.012.

Box, G. E. P., and Cox, D.R. (1964), An analysis of transformations, *Journal of Royal Statistical Society, Series B*, 26, 211-246.

Croke, B.F.W. (2007), The role of uncertainty in design of objective functions, In Oxley, L. and Kulasiri, D. (eds) *MODSIM 2007 International Congress on Modelling and Simulation*. Modelling and Simulation Society of Australia and New Zealand, December 2007, ISBN: 978-0-9758400-4-7, 2541-2547, 2007.

Croke, B., Wagener, T., Post, D., Freer, J. and Littlewood, I. (2008), Evaluating the information content of data for uncertainty reduction in hydrological modelling, Position paper for Workshop 2, iEMSs2008. In Miquel Sànchez-Marrè, Javier Béjar, Joaquim Comas, Andrea E. Rizzoli, Giorgio Guariso (Eds.) *Proceedings of the iEMSs Fourth Biennial Meeting: International Congress on Environmental Modelling and Software (iEMSs 2008)*. International Environmental Modelling and Software Society, Barcelona, Catalonia, July 2008. ISBN: 978-84-7653-074-0, Vol 3, pp1912-1926.

Hill, M.C. and Tiedeman, C.R., (2007), Effective Groundwater Model Calibration, With Analysis of Data, Sensitivities, Predictions, and Uncertainty, Wiley-Interscience, 455pp, ISBN 0470041072, 9780470041079.

Sorooshian, S., and Dracup, J.A. (1980), Stochastic Parameter Estimation Procedures for Hydrological Rainfall-Runoff Models: Correlated and Heteroscedastic Error Cases, *Water Resources Research*, 16 (2), 430-442.