ANU

THE AUSTRALIAN NATIONAL UNIVERSITY

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

Barry Croke

The Fenner School of Environment and Society and Department of Mathematics

Objective functions

- Typically, objective functions measure how well a model matches the data – with different objective functions measuring different aspects of the model performance
  - e.g. NSE and log transformed NSE
- For PUB, really interested in how well the model matches reality – need to consider how well the data represents what is happening in the real world
- Also need to maximise information being used by the objective function

Measurements and uncertainties

- All measurements have an associated uncertainty. Without the uncertainty, the measurement is almost meaningless

- Any use of the measurement must take into consideration the uncertainty

Objective functions and uncertainty

- Any objective function should allow for the uncertainty in the quantities being compared (both observed and modelled values)
- For streamflow, uncertainty in the observed flows dominated by uncertainty in the rating curve
  - Need information on rating curve – in absence of this, need to make an estimate of the uncertainty
  - Uncertainty in shape of hydrograph considerably less than the uncertainty in the flow at any point – Monte Carlo approach needed?

Optimum weighted NSE

- Optimum weighted NSE:

$$R^2 = 1 - \frac{\sum_i s_i \omega_i \left(Q_{o,i} - Q_{m,i}\right)^2}{\sum_i s_i \lambda_i \left(Q_{o,i} - \overline{Q}_o\right)^2}$$

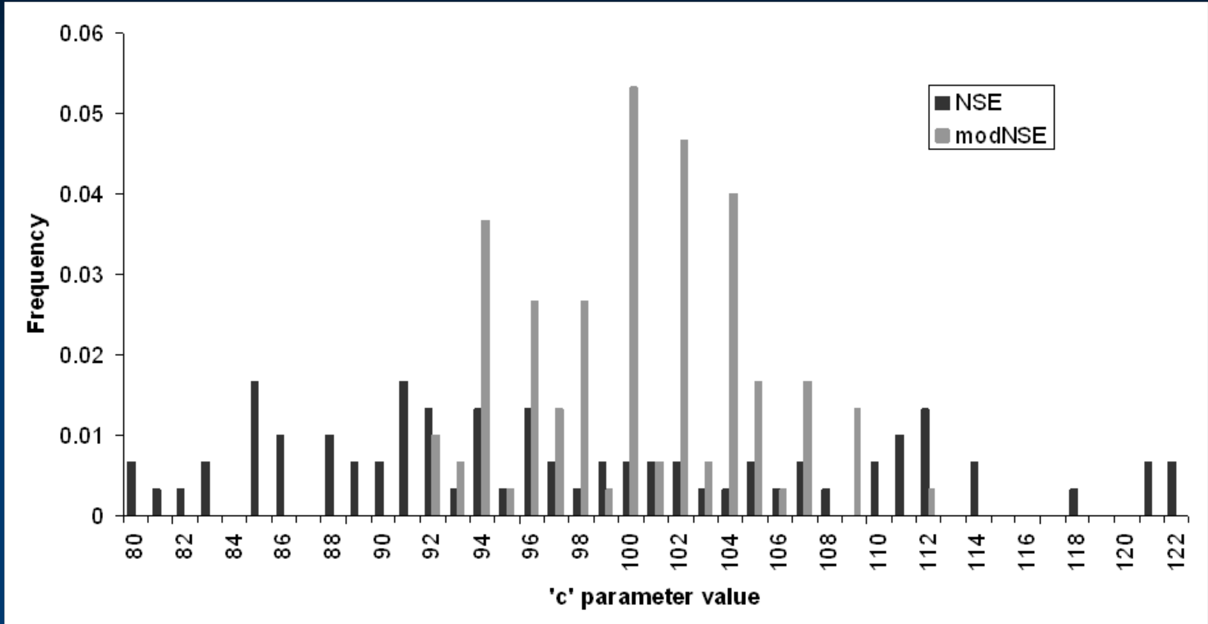
$$\omega_i = 1 \Big/ \left(\left(\Delta Q_{o,i}\right)^2 + \left(\Delta Q_{m,i}\right)^2\right), \quad \lambda_i = 1 \Big/ \left(\Delta Q_{o,i}\right)^2,$$

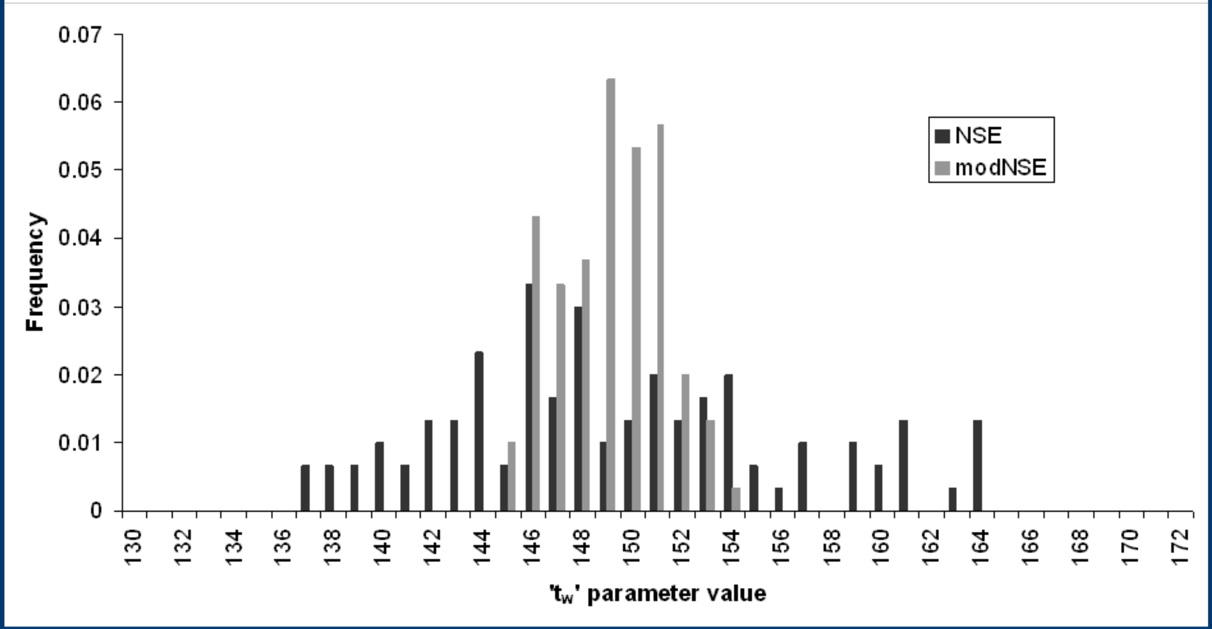
- If s_i <> 1, then report 2 measures, one with the adopted s_i, the other with $s_i = 1$

Sources of uncertainty

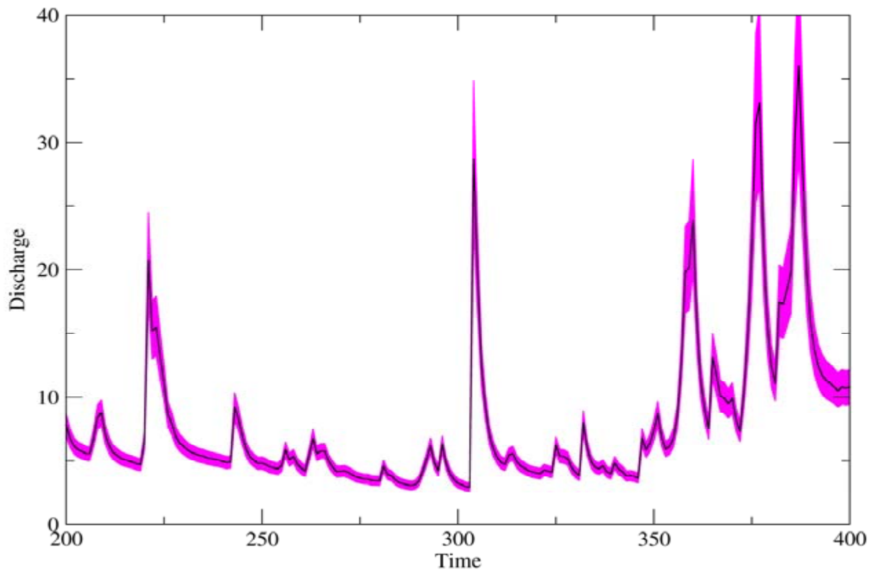
- Important sources of uncertainty
  - Areal rainfall estimates – as well as the temporal resolution
  - Streamflow (predominantly uncertainty in the rating curve?)
  - Additional data increases the information available, at the cost of increased noise (I/N)
- Other sources of uncertainty
  - Appropriateness of model structure
    - Not included in weighting – reflected in model performance
  - Parameter uncertainty
    - Simulation only

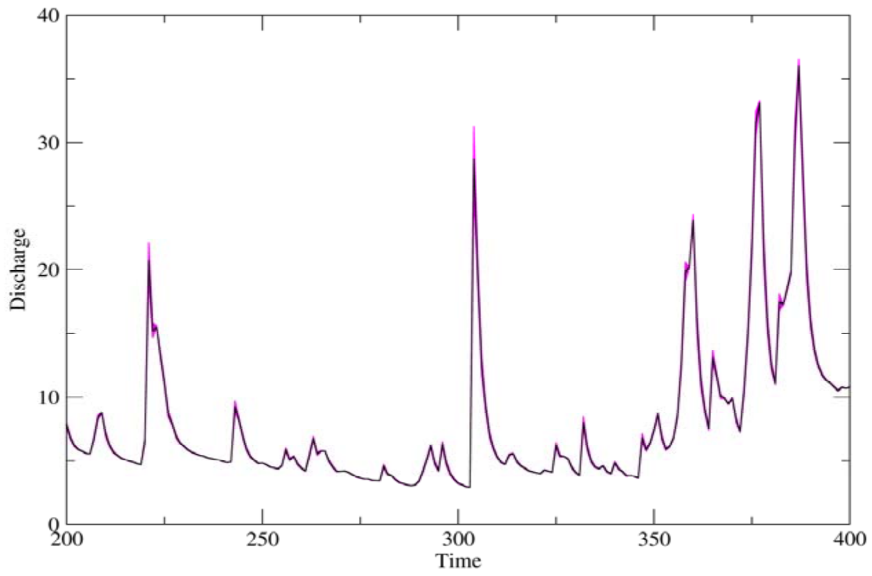
Uncertainty and R^2





Serial correlation in observed flow





Including serial correlation

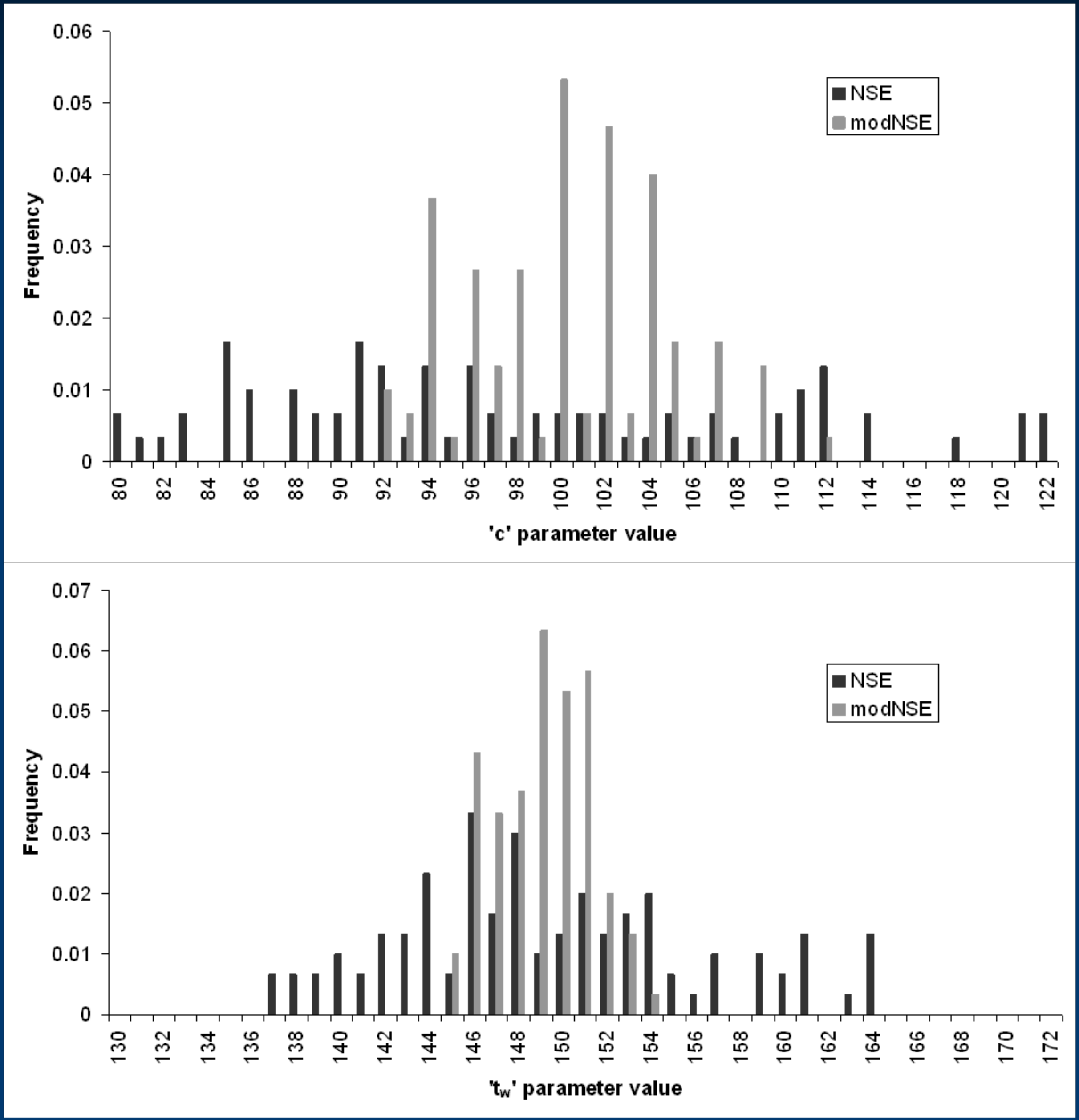
$$R^2 = 1 - \frac{\sum_{i=1}^{n} s_i \left[\omega_s \left(q_{o,i} - q_{m,i} \right)^2 + \omega_{s,i,i-1} \left(d_{o,i,i-1} - d_{m,i,i-1} \right)^2 \right]}{\sum_{i=1}^{n} s_i \left[\lambda_i \left(q_{o,i} - \bar{q}_o \right)^2 + \lambda_{s,i,i-1} \left(d_{o,i,i-1} - \left(\frac{q_{o,1} - q_{o,n}}{n}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

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

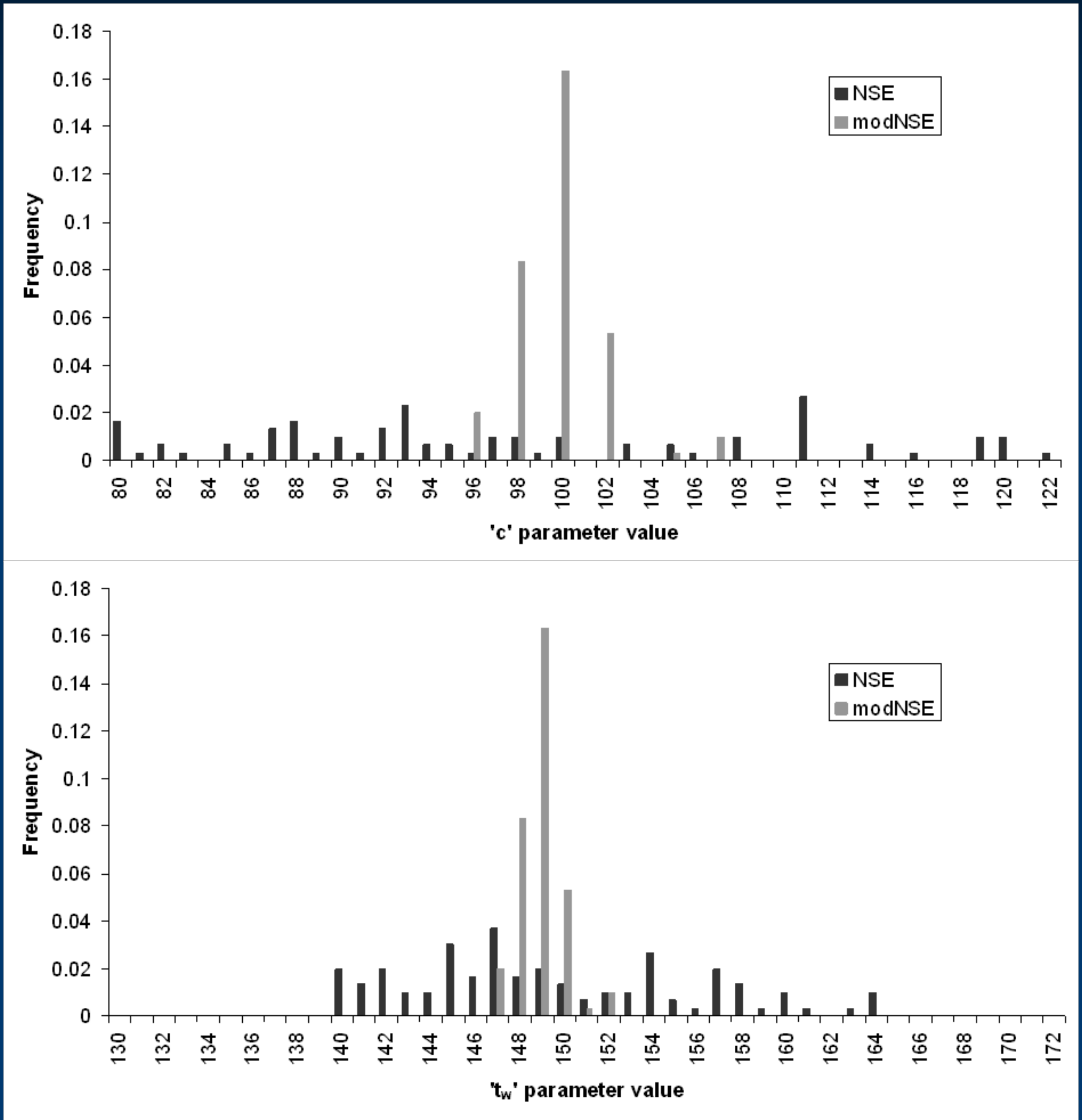
$$\left(\Delta d_{o,i,i-1} \right)^2 = \left(\frac{\partial d_{o,i,i-1}}{\partial a} \Delta a \right)^2 + \left(\frac{\partial d_{o,i,i-1}}{\partial b} \Delta b \right)^2 + \left(\frac{\partial d_{o,i,i-1}}{\partial h_0} \Delta h_0 \right)^2 + \left(\frac{\partial d_{o,i,i-1}}{\partial h_i} \Delta h_i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

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

Uncertainty and R^2



Including serial correlation



Long-term serial correlation

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

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

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

Conclusions

- Due to serial correlation (particularly in observed flow), there is considerable information in the model residuals that is not considered in the NSE
- Inclusion of serial correlation in NSE can reduce parameter uncertainty
- Serial correlation in observed flow exists as long as the rating curve is stable
- Lack of long-term serial correlation in modelled flow limits the effectiveness of the long-term serial correlation in the observed flow, but does not prevent use of this (depends on relative uncertainty)