

# Influential point detection diagnostics in hydrological model calibration

D. Wright, M. Thyer and S. Westra

Email: david.p.wright@adelaide.edu.au

University of Adelaide, School of Civil, Environmental & Mining Engineering, Australia



## Influence diagnostics in hydrological modelling

Influence diagnostics not widely used in hydrology despite being a standard tool in statistical literature.

### Aims of this study:

1. **Quantify** the influence of individual days on model parameters and predictions
2. **Explore** the relationship between streamflow and influence
3. **Compare** numerical and analytical influence diagnostics

## Measuring influence using two classes of diagnostics:

We evaluate two classes of diagnostics that identify influential data for hydrological model calibration:

### 1. Numerical case-deletion diagnostics:

Directly measure the influence of deleting each data point on the:

1. **Relative change in predictions**
2. **Relative change in parameters**
3. **Objective Function Displacement**

### 2. Analytical Cook's distance [Cook and Weisberg, 1982]

Calculate influence from matrix multiplication:

$$CD_i = \frac{(r_{Si})^2}{p} \times \frac{h_i}{1-h_i} \quad (1)$$

Where  $r_{Si}$  is standardised residual,  $p$  is the number of parameters and  $h_i$  is the leverage

We calculate Cook's distance using linear and nonlinear leverage [St. Laurent and Cook, 1993].

## References:

Cook, R. D., and S. Weisberg (1982), *Residuals and influence in linear regression*, Chapman and Hall, New York.

Perrin, C., C. Michel, and V. Andreassian (2003), Improvement of a parsimonious model for streamflow simulation, *Journal of Hydrology*, 279(1-4), 275-289.

St. Laurent, R. T., and R. D. Cook (1993), Leverage, local influence and curvature in nonlinear regression, *Biometrika Trust*, 80(1), 99-106

Westra, S., M. Thyer, M. Leonard, D. Kavetski, and M. Lambert (2014), A strategy for diagnosing and interpreting hydrological model nonstationarity, *Water Resources Research*, 50(6), 5090-5113

Woods, R. A., R. B. Grayson, A. W. Western, M. J. Duncan, D. J. Wilson, R. I. Young, R. P. Ibbitt, R. D. Henderson, and T. A. McMahon (2001), Experimental Design and Initial Results from the Mahurangi River Variability Experiment: MARVEX, *Observations And Modeling Of Land Surface Hydrological Processes*, pp. 201-213.

## How influential are individual days to model calibration?

### Case study 1: Ephemeral catchment



Scott Creek  
(SE Australia)

Model: GR4J [Perrin et al., 2003]

Figure 1: Scott Creek catchment [Westra et al. 2014]

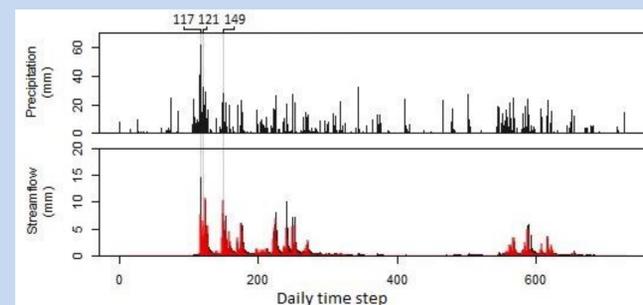


Figure 2: Scott Creek hydrological time series

### Case study 2: Humid catchment

French Broad River  
(North Carolina, USA)  
[Woods et al. 2001]

Model: GR4J [Perrin et al., 2003]



Figure 3: French Broad River catchment [Google, 2014]

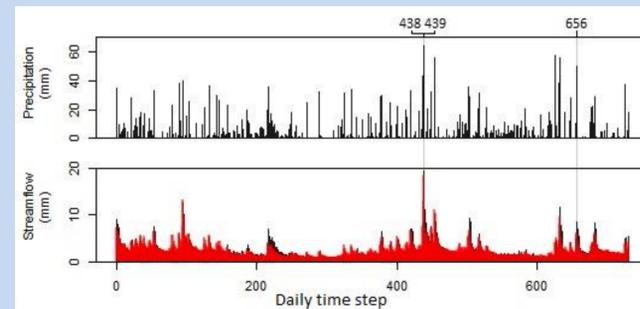


Figure 4: French Broad River hydrological time series

## Individual points have a influence on predictions

Relative change in the mean and max. model predictions for the top 10 most influential points in each case study in Figure 5.

### Ephemeral catchment:

- 25% change in predicted maximum
- 13% change in predicted mean

### Humid catchment:

- 2.3% change in predicted maximum
- 0.2% change in predicted mean

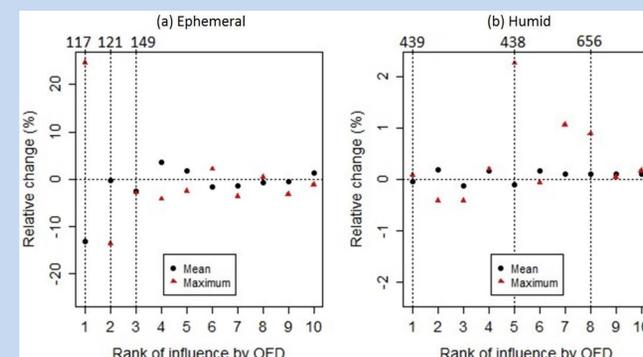


Figure 5: Relative change in removing the top 10 most influential data on model predictions

## Individual points have a influence on parameters

Example bivariate scatterplots are shown in Figure 6 for each case study with equal Mahalanobis distance contours.

### Ephemeral catchment:

- 35% change in  $\theta_1$  and  $\theta_2$
- Max. Mahalanobis distance of 26.9

### Humid catchment:

- 3% change in  $\theta_1$  and  $\theta_2$
- Max. Mahalanobis distance of 19.3

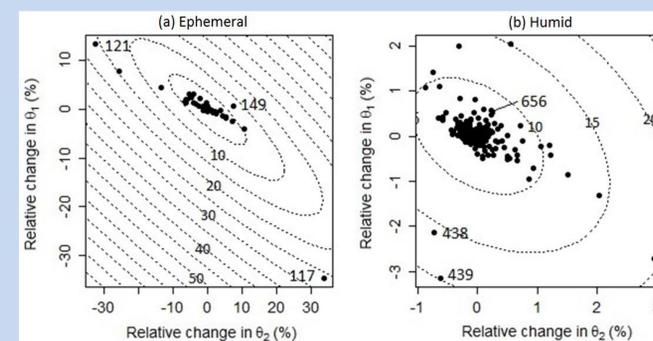


Figure 6: Example parameter scatter plots contributing to a four-dimensional Mahalanobis distance

## Peak flows are not always influential

There is not a clear one-to-one relationship between observed streamflow value and the influence diagnostics as illustrated in Figure 7.

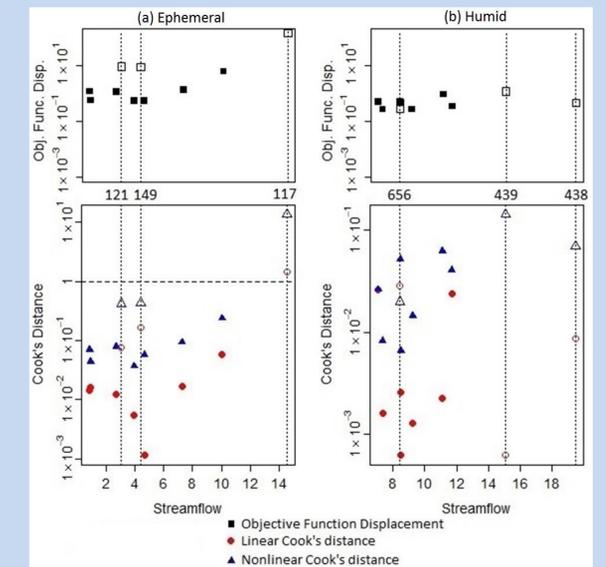


Figure 7: Observed streamflow against influence for the top 10 points ranked by case-deletion Objective Function Displacement

## Cook's distance combines the residual and the leverage of a point

Components of linear and nonlinear Cook's distance are shown in Figure 8. Including nonlinearity in Cook's distance formulation generally increases the magnitude of the identified points.

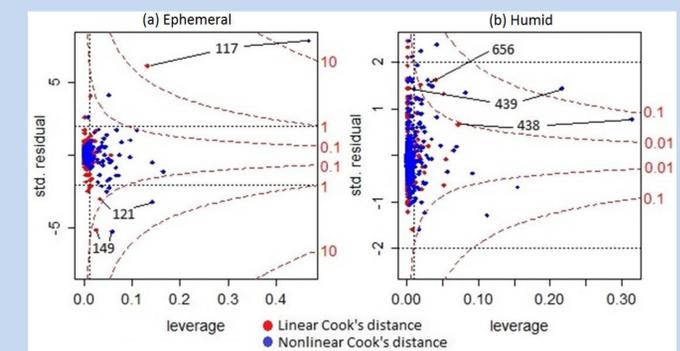


Figure 8: Components of linear and nonlinear Cook's distance. Contours represent regions with equal Cook's distance.

## Conclusions:

1. Single points can have a substantial influence on model predictions and parameter estimates in hydrological modeling
2. Influence of individual points is less in the humid catchment than the ephemeral catchment
3. Nonlinear Cook's distance is computationally cheaper than case-deletion (300-1000 times) and easily implemented

Paper currently in review: Wright et al. [2015] *Influential point detection diagnostics in the context of hydrological model calibration*

## Nonlinear Cook's distance

1. Provides almost identical influence rankings to case-deletion
2. 300-1000 times computationally cheaper than case deletion