Journal of Hydrology 527 (2015) 1161–1172

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Influential point detection diagnostics in the context of hydrological model calibration

David P. Wright *, Mark Thyer, Seth Westra

School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide 5005, Australia

A R T I C L E I N F O

Article history:
Received 5 November 2014
Received in revised form 22 April 2015
Accepted 23 May 2015
Available online 30 May 2015
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Vazken Andréassian, Associate Editor

Keywords:
Hydrologic calibration
Case-deletion
Cook's distance
Influence diagnostics
Mahalanobis distance

S U M M A R Y

Influential data are those that have a disproportionate impact on model performance, parameters and/or predictions. This paper evaluates two classes of diagnostics that identify influential data for hydrological model calibration: (1) numerical "case-deletion" diagnostics, which directly measure the influence of each data point on the calibrated model; and (2) analytical diagnostics based on Cook's distance, which combine information on the model residuals with a measure of the distance of each input point from the centre of the range of the input data (i.e., the leverage). Case-deletion methods rank influence by changes in the model parameters (measured through the Mahalanobis distance), performance (using objective function displacement) and predictions (e.g. mean and maximum streamflow). For the analytical methods, both linear and nonlinear estimates of leverage are used to calculate Cook's distance, which is used to rank influential data. We apply these diagnostics to three case studies and show that a single point could change mean/maximum streamflow predictions by 7%/9% for a rating curve model, and 13%/25%, for a hydrological model (GR4J) in an ephemeral catchment. In contrast, the influence was far less for GR4J in a humid catchment (0.2%/2.3%). Assuming the data are of high quality this indicates deficiencies in the ability of the GR4J model structure to reproduce the flow regime in the ephemeral catchment. The linear Cook's distance-based metric produced reasonably similar rankings to the case-deletion metrics at a fraction of the computational cost (300–1000 times faster), but with less flexibility to rank influence using specific aspects of model behaviour. The nonlinear distance produced rankings that were virtually the same as the case-deletion metrics for all case studies – this highlights the importance of its use for nonlinear hydrological models. Visual assessment was not a reliable method of influence analysis as there was no direct relationship between the most influential data and the highest observed streamflows. The findings establish the feasibility and importance of including influence detection diagnostics as a standard tool in hydrological model calibration.

1. Introduction

The process of hydrological model calibration involves the estimation of parameters that maximise the similarity between observed and simulated hydrological response time series such as streamflow. This process requires the optimisation of one or several objective functions (Duan et al., 1992), which provide a summary measure of overall model performance. However in doing so, information on the influence of individual data points in determining the calibrated parameter set (and hence the model predictions) is often ignored.

Identifying data points that have a large influence on hydrological predictions is of particular importance when those data points are erroneous, as this is likely to lead to sub-optimal model performance when applied to an independent dataset. The importance of such "disinformative" data has been highlighted by Beven and Westerberg (2011), who identify the need for more formal methods to identify and remove erroneous data prior to model calibration. They suggest two strategies: firstly that the discrepancies of a water balance time series are evaluated for values outside some acceptable limits of uncertainty, and secondly that likelihood measures are developed that are robust with respect to disinformation. However, examining all of the high residual data can be labour intensive, and focusing only on a smaller subset of influential data is likely to be more feasible in practice. Furthermore, not all influential points are erroneous; in fact, in certain situations it may even be desirable that some data points are more influential than others. For example, objective functions that place a larger weight on high flows maybe more desirable if the application is for peak flow prediction (e.g. Duan et al., 2007). This paper aims to provide

* Corresponding author.
E-mail addresses: david.p.wright@adelaide.edu.au (D.P. Wright), mark.thyer@adelaide.edu.au (M. Thyer), seth.westra@adelaide.edu.au (S. Westra).

http://dx.doi.org/10.1016/j.jhydrol.2015.05.047
0022-1694/© 2015 Elsevier B.V. All rights reserved.

hydrological modellers with the tools to assess relative influence of data points on model calibration.

Influential data points are defined as points that exert a disproportionate impact on the calibrated parameters, performance and/or predictions. Formal influential point detection methods are widely used both for the detection of erroneous points and for identifying possible model deficiencies, with common applications in linear regression (Cook, 1979), generalised linear regression (Thomas and Cook, 1989), generalised additive models (Hastie and Tibshirani, 1990) and various other regression-based approaches (Chen et al., 2012; Russo et al., 2009). The diagnostics can be grouped into two classes: case-deletion approaches and analytical leverage-based approaches.

Case-deletion methods were first developed by Cook (1977) and involve removing ("deleting") a data point ("case") from the set of calibration points, and then recalibrating the model. Parameter estimates and model predictions from the recalibration are compared to the results from the full calibration, and this is repeated for all data points in the calibration set. A recent example in the context of flood frequency analysis used case-deletion to show that low flow outliers can have a disproportionate influence on extreme flood quantile estimates (Lamontagne et al., 2013). Their technique was based on a generalised Grubbs-Beck test statistic developed by Cohn et al. (2013) that is designed to identify potentially influential low flows.

Case-deletion approaches can be computationally intensive, as they require the re-estimation of the parameters after deleting each point from the calibration data set. Furthermore, case -deletion involves comparing the optimal parameter sets from each calibrated model run, and thus anomalous results are possible for models with complex response surfaces that are prone to local optima (Duan et al., 1992). As an alternative, Cook's distance (Cook, 1977) provides an analytical measure of the influence of points, and thus does not require multiple re-calibrations. It combines measures of the distance between each observed data point and the fitted model (the residual) and the distance of each data point from the centre of the input space (the leverage). Cook's distance was originally developed for linear regression models, but may also be applied to nonlinear models if the models are approximately linear in the vicinity of the optimum parameter set (Cook and Weisberg, 1982). Alternatively, nonlinear formulations of the leverage are also available (St. Laurent and Cook, 1992), and may be better suited to the highly nonlinear behaviour of many hydrological models (e.g. see discussion in Kavetski and Kuczera, 2007).

The influence concepts are illustrated in Fig. 1 by applying case-deletion to a linear regression model. Point A is highly influential, with a significant difference in calibrated parameters when including this point ($\beta_0 = 2.0, \beta_1 = 2.3$, compared to $\beta_0 = 3.4$, $\beta_1 = 1.9$). The influence on predictions is also evident by comparing the fitted regression lines, with the greatest differences apparent towards the high and low extremes of the input data. In contrast, although point B has a similar residual to A (i.e. the difference between the data point and the fitted curve is similar), it exerts a much smaller influence on both the parameters ($\beta_0 = 3.8, \beta_1 = 1.9$) and the fitted regression line. Although the application of influence diagnostics may appear trivial in this example, the complex mapping from input to output space in hydrological models often precludes visual techniques, so that more formal approaches for the detection of influential points are required.

The prospect that a small number of data points can exert a very high influence on model performance motivates the more widespread implementation of influence diagnostics in hydrology, however applications have been few and recent. In the context of groundwater modelling, Yager (2004) found that models were highly sensitive to small changes in influential data. Foglia et al. (2007) used a series of case-deletion metrics and Cook's distance

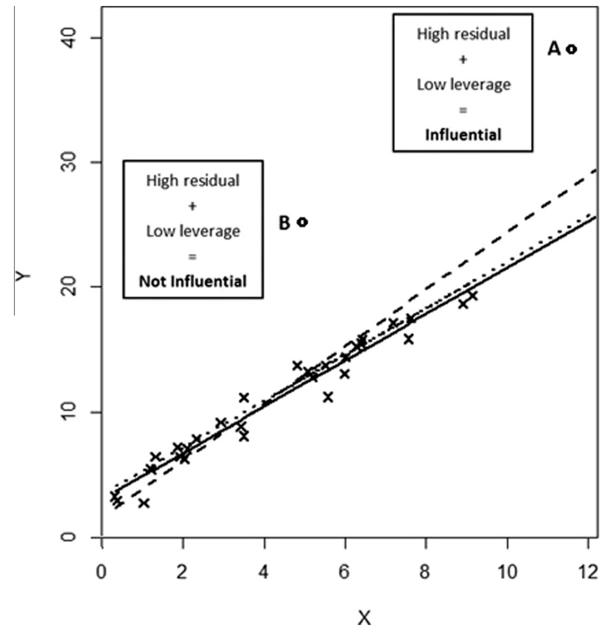

Fig. 1. A simple linear regression scatter plot illustrates the impact of a highly influential data point on the fitted model. The solid line is the prediction curve without point A or B in the calibration data; the broken prediction curve is with point B only excluded, as A is an observation that is both an outlier and a high leverage point; the dotted prediction curve is with point A only excluded, as B is an observation with the same residual as point A but with low leverage.

approaches on a groundwater model and found similar performance between the two metrics. Foglia et al. (2009) applied linear Cook's distance as part of a suite of diagnostics to a short time series of 37 daily observations in the rainfall-runoff model TOPKAPI and found that some of the low flow observations during small precipitation events were more important than anticipated. Legates and McCabe (1999) discuss the oversensitivity to outliers of correlation based goodness-of-fit measures used in hydrological models and recommend that additional evaluation measures should supplement calibration. Berthet et al. (2010) found a quadratic criterion to be influenced by a very small number of time steps characterised with high runoff variation. Perrin et al. (2007) assess the impact of the quantity and quality of streamflow data on parameter calibration and model robustness and show that a subset of influential points from a larger dataset are sufficient to obtain robust estimates. Singh and Bárdossy (2012) pre-process hydrological data using depth functions to identify unusual events and investigate the calibration of the model with only this set of critical data to assess if the subset has enough information to identify model parameters. Each of these studies contributes towards the more widespread use of influence assessment, however a comprehensive assessment of the influence of individual data points in the context of hydrological model predictions and parameters is still lacking.

The goal of this paper is to evaluate the use of influence diagnostics in the context of common hydrological modelling case studies: stage/discharge rating curve model and a conceptual hydrological model. Case-deletion, linear and nonlinear Cook's distance will be compared in terms of performance and computational run times. Tailored statistics that are suitable for hydrological model applications will be developed for measuring the effect of data points on the model parameters, performance and/or predictions. This analysis will identify the extent to which the model predictions are influenced by a small number of data points – thereby evaluating the information content of data points