Environmental Modelling & Software 63 (2015) 32–44

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Streamflow rating uncertainty: Characterisation and impacts on model calibration and performance

Jorge L. Peña-Arancibia [a,*], Yongqiang Zhang [a], Daniel E. Pagendam [b], Neil R. Viney [a], Julien Lerat [d], Albert I.J.M. van Dijk [e,a], Jai Vaze [a], Andrew J. Frost [c]

[a] Land and Water Flagship, CSIRO, Black Mountain, GPO 1666, Canberra, ACT 2601, Australia
[b] Digital Productivity and Services Flagship, CSIRO, EcoSciences Precinct, PO Box 2583, Brisbane, QLD, Australia
[c] Bureau of Meteorology, PO Box 413, Darlinghurst, NSW 1300, Australia
[d] Bureau of Meteorology, GPO Box 2334, Canberra, ACT 2601, Australia
[e] Fenner School of Environment & Society, Australian National University, Canberra, ACT 0200, Australia

A R T I C L E I N F O

Article history:
Received 23 June 2014
Received in revised form
2 September 2014
Accepted 9 September 2014
Available online

Keywords:
Modelling
Calibration
Ensemble
Stage-discharge
Rating curve
Nadaraya-Watson
Heteroscedasticity
Uncertainty

A B S T R A C T

Common streamflow gauging procedures require assumptions about the stage-discharge relationship (the ‘rating curve’) that can introduce considerable uncertainties in streamflow records. These rating uncertainties are not usually considered fully in hydrological model calibration and evaluation yet can have potentially important impacts. We analysed streamflow gauge data and conducted two modelling experiments to assess rating uncertainty in operational rating curves, its impacts on modelling and possible ways to reduce those impacts. We found clear evidence of variance heterogeneity (heteroscedasticity) in streamflow estimates, with higher residual values at higher stage values. In addition, we confirmed the occurrence of streamflow extrapolation beyond the highest or lowest stage measurement in many operational rating curves, even when these were previously flagged as not extrapolated. The first experiment investigated the impact on regional calibration/evaluation of: (i) using two streamflow data transformations (logarithmic and square-root), compared to using non-transformed streamflow data, in an attempt to reduce heteroscedasticity and; (ii) censoring the extrapolated flows, compared to no censoring. Results of calibration/evaluation showed that using a square-root transformed streamflow (thus, compromising weight on high and low streamflow) performed better than using non-transformed and log-transformed streamflow. Also, surprisingly, censoring extrapolated streamflow reduced rather than improved model performance. The second experiment investigated the impact of rating curve uncertainty on catchment calibration/evaluation and parameter estimation. A Monte-Carlo approach and the nonparametric Weighted Nadaraya-Watson (WNW) estimator were used to derive streamflow uncertainty bounds. These were later used in calibration/evaluation using a standard Nash-Sutcliffe Efficiency (NSE) objective function (OBJ) and a modified NSE OBJ that penalised uncertain flows. Using square-root transformed flows and the modified NSE OBJ considerably improved calibration and predictions, particularly for mid and low flows, and there was an overall reduction in parameter uncertainty.

1. Introduction

Streamflow data are generally estimated from stage measurements through a stage–discharge relationship (the ‘rating curve’), developed through measurement of flow using manual methods (estimation of flow velocity combined with estimates of river width and height for subsections of the river) and relating that to measured flow height at various points in time; then interpolation/extrapolation of that relationship across all height-flow levels using regression techniques to produce a curve. Several sources of uncertainty can be accounted for in this procedure including measurements of flow height, width and shape of the river cross-section and inaccuracies in the measurement of the velocity–area relationship (Domeneghetti et al., 2012). Another source of uncertainty arises from the regression techniques used to derive the stage–discharge relationship. The classical approach for deriving

* Corresponding author. CSIRO Land and Water, GPO Box 1666, Canberra ACT 2601, Australia. Tel.: +61 (2)6246 5711; fax: +61 (2)6246 5800.
E-mail address: jorge.penaarancibia@csiro.au (J.L. Peña-Arancibia).

http://dx.doi.org/10.1016/j.envsoft.2014.09.011

stage-discharge (rating) relationship involves fitting a curve for (log-transformed) discrete rating measurements using (non)linear least squares. This implicitly assumes that the measurement residuals have a normal distribution and are unrelated to the expected discharge (Petersen-Øverleir, 2004). Residuals for existing curves often show non-normal distributions (e.g. Tomkins and Davidson, 2011) with higher residual values at higher stage values (heteroscedasticity). Scarce sampling and heteroscedasticity observed in streamflow residuals may introduce large uncertainty in streamflow estimates based on extrapolation of the rating curve (Westerberg et al., 2011). These streamflow observations are the core data used to calibrate hydrological models.

Objective functions (OBJs) are used in calibration to minimise the differences between observed and modelled streamflow and also to assess the model performance under prediction. Traditionally, the minimisation is performed against the sum-of-squared residuals under the assumption that these residuals are homoscedastic in nature (i.e. there is no variance heterogeneity in the streamflow data). This assumption is often not valid for streamflow data (Petersen-Øverleir, 2004) and its violation may overestimate goodness-of-fit metrics used in simulations (McMillan et al., 2010). Moreover, routinely used OBJs in calibration, for example the Nash-Sutcliffe Efficiency (NSE, Nash and Sutcliffe, 1970), place high weights on high flows which may be extrapolated, thus potentially biasing predictions (Croke, 2007).

In this paper, we investigate the impact of streamflow rating uncertainty on hydrological model calibration and performance (i.e. 'prediction' using streamflow data from catchments not used for model calibration or split-sample 'evaluation' using streamflow data from a period not used for model calibration). Firstly, we use a comprehensive hydrometric dataset of 65 streamflow gauges (described in Section 2) to assess the occurrence of heteroscedasticity and extrapolation in rating curves (Section 2.1). Secondly, we conduct two types of experiments:

(i) The first experiment makes use of the entire streamflow dataset (65 streamflow time-series) to assess the impacts of including uncertain extrapolated streamflow data in a regional calibration/prediction experiment (Section 2.2). Several methods were trialled to address this problem; from censoring all extrapolated high flows to using streamflow Box-Cox transformation (Box and Cox, 1964; Bennett et al., 2013) in an attempt to reduce heteroscedasticity. For this experiment, we calibrated a single parameter set ($n = 28$) of the process-based landscape water balance model Australian Water Balance Assessment system Landscape model (AWRA-L) (van Dijk, 2010; van Dijk and Renzullo, 2011; Vaze et al., , 2013) in 33 of the 65 stations and performance was independently evaluated for the remaining 32 stations. We assessed the impact of censoring high flows on the NSE compared to no censoring. We repeated the experiment using two streamflow transformations (logarithmic and square-root). The regional calibration (i.e. a single set of parameters to predict flows in a large geographical domain) was chosen for methodological and practical reasons. Firstly, predictions of streamflow and other fluxes and stores (e.g. evapotranspiration and soil water) are required in many ungauged basins with dissimilar climate and biophysical characteristics; a regional calibration using a large amount of catchment streamflow data might yield better results than parameter regionalisation techniques (Parajka et al., 2007; Vaze and Teng, 2011). Secondly, AWRA-L has been regionally calibrated (using a similar approach as in this paper) against Australian streamflow and evapotranspiration data; producing results that markedly improved compared to a previous non-calibrated version (version 1.0 vs. 0.5; Viney et al., 2011). The calibration results were also similar to results from locally calibrated conceptual models, showing that AWRA-L can capture the different climatic and biophysical characteristics that affect streamflow (Viney et al., 2011). Thirdly and finally, AWRA-L is currently used operationally to provide information on water fluxes and stores across Australia and its being continuously refined.

(ii) The second experiment investigates the impact of rating curve uncertainty on the NSE and parameter estimation in a local calibration/evaluation (Section 2.3) using a Monte-Carlo approach and the nonparametric Weighted Nadaraya-Watson (WNW) estimator. We use these methods for quantifying the error in the rating curve because they capture changes in the rating curve with time, they are nonparametric and make minimal assumptions about the probabilistic distribution of the data. We employed them to derive rating curve uncertainty bounds for 100 streamflow realisations. To interpret impacts on parameter space, we calibrated 4 parameters of the simpler conceptual rainfall-runoff model GR4J (compared to the 28-parameter AWRA-L) (Perrin et al., 2003). These were later used in split-sample calibration/evaluation in a single station using a standard NSE OBJ and a modified NSE OBJ, which used the uncertainty bounds to penalise uncertain flows. Again, we repeated the experiment using logarithmic and square-root streamflow transformations.

The data and methods are described in Section 2. The results of the experiments are presented and analysed in Section 3, the findings are discussed in Section 4 and finally conclusions are drawn (Section 5).

2. Data and methods

The New South Wales (NSW) Office of Water (NoW) in Australia regularly republishes the 'Pinneena' water database on DVD (http://waterinfo.nsw.gov.au/pinneena/gw.shtml). The version used here (December, 2009) includes 127,000 years of daily streamflow information from 1400 stations. The database includes records of hydraulic control type (including concrete structures, rocky river bed not reinforced with concrete, gravel or sand river bed), stage height, rating tables, interpolation method, gauging measurements and percentage deviations of gaugings from the rating curve. This detailed database can be used to infer uncertainty due to extrapolation, the occurrence of heteroscedasticity and with the use of statistical techniques, uncertainty in streamflow data.

The catchments used in the experiments performed here (Fig. 1) were chosen from a subset in the 'Pinneena' database that was selected for previous modelling studies (Zhang et al., 2013) because they are headwater catchments without significant influence from river regulation, urbanisation and irrigation. From this subset, catchments with >15 years of daily streamflow data (all in ML d^{-1}) during 1980–2008 and >70 rating measurements were selected, resulting in a total of 65 stations. Hourly streamflow data was extracted from the 'Pinneena' database and averaged from 9:00 a.m. to 9:00 a.m. the following day, to coincide with the time of daily rainfall recording. The streamflow volumes were converted to areal average streamflow (mm d^{-1}).

Climate data used as model forcing included rainfall, Priestley-Taylor potential evapotranspiration, minimum and maximum temperature and incoming shortwave radiation. These were sourced or derived from the Specialised Information for Land Owners (SILO) dataset (Jeffrey et al., 2001) available from the Queensland Department of Environment and Resource