Journal of Hydrology 510 (2014) 137–152

Contents lists available at ScienceDirect

Journal of Hydrology

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

Uncertainty estimation with bias-correction for flow series based on rating curve

Quanxi Shao [a,c,*], Julien Lerat [b,c], Geoff Podger [b,c], Dushmanta Dutta [b,c]

[a] CSIRO Computational Informatics, Private Bag No. 5, Wembley, WA 6913, Australia
[b] CSIRO Land and Water, GPO Box 1666, Canberra, ACT 2601, Australia
[c] Water for a Healthy Country Flagship Program, CSIRO, Canberra, ACT, Australia

ARTICLE INFO

Article history:
Received 15 October 2013
Received in revised form 9 December 2013
Accepted 12 December 2013
Available online 20 December 2013
This manuscript was handled by Geoff Syme

Keywords:
Discharge estimation
Rating curve
Uncertainty
Hydrological modelling

SUMMARY

Streamflow discharge constitutes one of the fundamental data required to perform water balance studies and develop hydrological models. A rating curve, designed based on a series of concurrent stage and discharge measurements at a gauging location, provides a way to generate complete discharge time series with a reasonable quality if sufficient measurement points are available. However, the associated uncertainty is frequently not available even though it has a significant impact on hydrological modelling. In this paper, we identify the discrepancy of the hydrographers' rating curves used to derive the historical discharge data series and proposed a modification by bias correction which is also in the form of power function as the traditional rating curve. In order to obtain the uncertainty estimation, we propose a further both-side Box–Cox transformation to stabilize the regression residuals as close to the normal distribution as possible, so that a proper uncertainty can be attached for the whole discharge series in the ensemble generation. We demonstrate the proposed method by applying it to the gauging stations in the Flinders and Gilbert rivers in north-west Queensland, Australia.

1. Introduction

Streamflow discharge constitutes one of the fundamental data required to perform water balance studies and develop hydrological models. The reliability of hydrological models depends highly on the quality of streamflow discharge. In addition, the use of hydrological model to predict future streamflow requires the consideration of the data quality in calibration. Therefore, as the input data in hydrological models, the streamflow discharge needs to be as accurate as possible and to have proper uncertainty quantification attached in model calibration. However, continuous measurement of discharge relies on costly equipments such as Doppler current profilers (Nihei and Kimizu, 2008), that can only be used in a limited range of conditions only (e.g. deep river channels), and therefore regularly intensive measurements are rare except at some important locations. In practice, a relationship is established between the discharge and other observable variables, in particular the water level (stage), after a series of concurrent stage and discharge measurement are performed. The stage–discharge relationship is referred as the rating curve. The discharge can then be routinely estimated by the stage measurement via the rating curve. An extensive overview of the use of rating curves in estimating stream flow for hydrological data production can be found in WMO (2008).

The simplest rating curve, which is also hydraulic motivated, is given by the mathematical form of

$$Q = A(w + c)^b, \qquad (1)$$

where Q is the discharge, w the water level, and A, b and c are the parameters (Lambie, 1978). (It should be noted that the simplest form is the case with $c = 0$ but often that c is added for better fitting). The background theory of this kind of rating curves can be found in Ackers et al. (1978), ISO (1998).

The above rating curve works well for regular channel (e.g., smoothly changed topography of cross section) under normal condition (e.g., stable flow condition). However, the discharge relationship is in fact affected by not only stage but also many other factors such as slope of water surface and shape of the channel (Herschy, 1999). Although for some important stations with collection of more factors, more complicated and actual discharge estimation can be achieved by using more factors such as Manning's equation (Manning, 1891) which still needs approximation for its parameters, the use of rating curve in the form of Eq. (1) is still a popular method for discharge estimation. In fact, the rating curve (Eq. (1)) is based on Manning's equation. More efforts have been made to improve the discharge estimation by generalizing the rating curve. For example, Petersen-Øverleir and Reitan (2005) used a two

* Corresponding author at: CSIRO Computational Informatics, Private Bag No. 5, Wembley, WA 6913, Australia. Tel.: +61 8 9333 6613; fax: +61 8 9333 6121.
E-mail address: Quanxi.shao@csiro.au (Q. Shao).

0022-1694/$ - see front matter Crown Copyright © 2013 Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhydrol.2013.12.025

segmented model and then generalized to a multiple segmented model (Reitan and Petersen-Øverleir, 2009). This setting copes with change(s) related to factors at certain water level such as flow control or change of channel shape. To cope with the temporal variation of the stage–discharge relationship, hydrographers normally produce a set of rating curves for different periods of data.

In practice, all flow discharges used in hydrological modelling are generated based on various approaches with a very limited number of measured discharges available. It is rare to have uncertainty estimation attached to the discharge data series although there is surely uncertainty in the data. As data uncertainty has been emphasized recently in hydrological modelling (Pappenberger and Beven, 2006), it is an urgent task to understand and finally provide the estimation of uncertainty in discharge data.

In this paper, we propose a framework to estimate discharge data and its associated uncertainty based on a preliminary discharge estimate performed by the data provider via a rating curve. We view this approach as a general framework as it can be applied widely to existing discharge data series. The framework is motivated by our current research project in Australia. As an application, we demonstrate the framework by the data from the research region.

The paper is organized as follows. The proposed framework is given in the next section. The study area and data are described in Section 3, following by the results and discussion. The conclusions are presented in Section 4.

2. Proposed framework

The objective of the framework is to provide an estimation of the uncertainty associated with a long series of modelled flows (e.g. several decades) produced with a rating curve model from observed water levels at a gauging station. The flows are denoted by $q_{m,j}$ $(j = 1, \cdots, T)$.

Assume that the rating curve model is too complex to be described with simple parametric relationships. This situation can occur when the rating curve is fitted manually by the hydrographer or for unstable river beds, which call for frequent revision of the rating curve. As a result, the only information available to develop the uncertainty estimation is the collection of measured flow data at the site noted by $Q_{o,i}$ $(i = 1, 2, \cdots, n)$, also referred as gauging points, with corresponding modelled values $Q_{m,i}$ based on rating curves. The number of gauging n is typically much smaller than T, ranging from less than ten up to hundred in magnitude for gauging stations that are closely monitored. For example, in the project motivating this study (see Section 3), the numbers of gauging n are usually less than a hundred but the estimated flow series are daily data over more than ten years, resulting in thousands of data. The main task here is to establish a technique to correct the possible discrepancy in the estimated flows and provide uncertainty estimation on the final flow estimation.

Note that all the estimated flows are available based on rating curves. One needs to correct its values. To do this, with the aid of measured flow only available at irregular time, one needs to find the way to model the measured flow data using corresponding estimated ones. Under an assumption that both the measured and modelled discharges have power relationship with the stages in the form of Eq. (1) with a common parameter c, one can propose the regression of measured flow as the function of modelled ones in the similar power function as

$$Q_{o,i} \approx \alpha Q_{m,i}^{\beta}. \tag{2}$$

(The error term will be specified later). It should be noted that the intercept coefficient is set to be zero based on the assumption that there is no systematic errors in the rating curves, which is supported by the real data. This setting includes the linear model as a special case by letting $\beta = 1$.

However, such an independent and identically distributed assumption imposed to the residuals is not always satisfactory in practice. However, it is frequently challenging to incorporate the error term in the above model due to the complicated error structure caused mainly by non-normality and heteroscedasticity in the residuals between the true and modelled values. Such a non-normality also makes the least squares approach inefficient (because underlying statistical assumptions for the least squares is that the residuals are independent and identically distributed as normal distribution). To solve this problem with strong assumption on the form of heteroscedasticity, the both-side Box–Cox regression is proposed by applying the Box–Cox transformation to both sides as

$$t(Q_{o,i}, \lambda) = t(\alpha Q_{m,i}^{\beta}, \lambda) + \varepsilon_{bc,i}. \tag{3}$$

where

$$t(x, \lambda) = \begin{cases} \frac{x^{\lambda}-1}{\lambda}, & \lambda \neq 0, \\ \log(x), & \lambda = 0 \end{cases}. \tag{4}$$

is the Box–Cox transformation with inverse transformation

$$t^{-1}(x, \lambda) = \begin{cases} (\lambda x + 1)^{1/\lambda}, & \lambda \neq 0, \\ \exp(x), & \lambda = 0. \end{cases} \tag{5}$$

This modelling idea was investigated by Wood (1974) and Carroll and Ruppert (1984) in statistics and was discussed in details Carroll and Ruppert (1988).

The advantage of both-side Box–Cox nonlinear regression is that it not only handles the nonlinear relationship but also tends to provide identically distributed residuals and make the resulted residuals as close to normal distribution as possible. By doing this, one does not need to pre-define the form of variation for the residuals. While the proposed bias-corrected flow adjustment works well if the data range of measured flows is comparable with that of the total modelled flows (that is, the correction for the modelled flow data does not subject to heavy extrapolation, especially for the high flows as the lower flows do not have significant problem due to the low bound of zero), it is not wise to implement bias correction if the data range of measured flows is too narrow in comparison with that of total modelled flow. In this case, the two-side Box–Cox regression becomes

$$t(Q_{o,i}, \lambda) = t(Q_{m,i}, \lambda) + \varepsilon_{bc,i}. \tag{6}$$

Which is to find a suitable Box–Cox transformation for both $Q_{o,i}$ and $Q_{m,i}$ $(i = 1, 2, \cdots, n)$ so that the differences between the transformed observed and modelled discharges is distributed as close to zero mean normal distribution as possible. As a result, our framework still works for stabilizing the variation in the residuals becomes to the estimation for the variation of the residuals.

It might be interesting to see the implication of the two-side Box–Cox transformation in the traditional regression framework where the response variable (measured flows) is a function of covariate(s) (modelled flows) with normal error and, in case of heterogeneity in residuals, the traditional regression needs to be calibrated by weighted least squares. By using the first-order Taylor series in term of the residual, our two-side Box–Cox transformation (Eq. (3)) is approximated as

$$Q_{o,i} \approx \alpha Q_{m,i}^{\beta} + \alpha^{1-\lambda} Q_{m,i}^{\beta(1-\lambda)} \varepsilon_{bc,i}, \tag{7}$$

which is in fact a weighted least squares with reparameterization $w = \beta(1-\lambda)$ and $\sigma_{\varepsilon} = \sigma_0/\alpha^{(1-\lambda)/2}$, where σ_0 being the standard variation of the residual in the original weighted least squares. However, in the traditional weighted least squares, the weight parameter w needs to be specified beforehand; otherwise an iterative computing procedure is needed. Therefore, our model (Eq. (3)) has clear computing advantage as well.