Journal of Hydrology 534 (2016) 164–177

Contents lists available at ScienceDirect

Journal of Hydrology

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

One-day offset in daily hydrologic modeling: An exploration of the issue in automatic model calibration

Masoud Asadzadeh [a,b,*], Luis Leon [b], Wanhong Yang [c], David Bosch [d]

[a] Civil Engineering Department, University of Manitoba, EITC E1-306, 15 Gillson Street, Winnipeg, MB R3T 5V6, Canada
[b] Environment Canada, National Water Research Institute, Burlington, Ontario L7R 4A6, Canada
[c] Department of Geography, University of Guelph, 50 Stone Road East, Guelph, Ontario N1G 2W1, Canada
[d] USDA-ARS, Southeast Watershed Research Lab, 2375 Rainwater Road, Tifton, GA 31794, USA

ARTICLE INFO

Article history:
Received 1 September 2015
Received in revised form 23 December 2015
Accepted 28 December 2015
Available online 4 January 2016
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Fi-John Chang, Associate Editor

Keywords:
Hydrologic modeling
Daily simulation
Automatic calibration
Offset in daily precipitation–runoff
Nash–Sutcliffe efficiency
Kling–Gupta efficiency

SUMMARY

Hydrologic modeling literature illustrates that daily simulation models are incapable of accurately representing hydrograph timing due to relationships between precipitation and watershed hydrologic response that happen with a sub-daily time step in the real world. For watersheds with a time of concentration less than 24 h and a late day precipitation event, the observed hydrographic response frequently occurs one day after the precipitation peak while the model simulates a same day event. The analysis of sub-daily precipitation and runoff in this study suggests that, this one-day offset is inevitable in daily analysis of the precipitation–runoff relationship when the same 24-h time interval, e.g. the calendar day, is used to prepare daily precipitation and runoff datasets. Under these conditions, daily simulation models will fail to emulate this one-day offset issue (1dOI) and result in significant daily residuals between simulated and measured hydrographs. Results of this study show that the automatic calibration of such daily models will be misled by model performance metrics that are based on the aggregation of daily residuals to a solution that systematically underestimate the peak flow rates while trying to emulate the one-day lags. In this study, a novel algorithm called Shifting Hydrograph In order to Fix Timing (SHIFT) is developed to reduce the impact of this one-day offset issue (1dOI) on the parameter estimation of daily simulation models. Results show that with SHIFT the aforementioned automatic calibration finds a solution that accurately estimates the magnitude of daily peak flow rates and the shape of the rising and falling limbs of the daily hydrograph. Moreover, it is shown that this daily calibrated model performs quite well with an alternative daily precipitation dataset that has a minimal number of 1dOIs, concluding that SHIFT can minimize the impact of 1dOI on parameter estimation of daily simulation models.

1. Introduction

The time step in actual precipitation–runoff relationship does not necessarily match well with the simulation time step in hydrologic models. Bosch et al. (2004) noticed that many major summer precipitation events in south-central Georgia occurred in the later part of the day, and that the hydrographic response of a relatively small sub-basin "*J*" in the Little River watershed was not observed until the subsequent day. A similar one-day gap was reported in daily precipitation–runoff analysis of studies by Van Liew et al. (2005), Uzeika et al. (2012) and Cho et al. (2013). Obviously, a short and intensive precipitation event that occurs early in a day can cause the same day hydrographic response in a small watershed with a sub-daily time of concentration (e.g., 18 h for sub-watershed "*J*" as estimated in Sheridan (1994)).

In the analysis of daily precipitation–runoff relationship at the outlet of a large-scale watershed, the (one day or even longer) gap should be expected between most peaks in runoff time series and the corresponding peaks in the precipitation time series mainly due to the large time of concentration. Daily simulation models are expected to emulate this gap and therefore greater model accuracy is expected in the estimation of runoff at the outlets of these watersheds. This expectation is consistent with the literature review by Gassman et al. (2007) that shows better daily model performance metric values achieved for larger basins, in general. In small-scale watersheds with sub-daily time of

* Corresponding author at: Civil Engineering Department, University of Manitoba, EITC E1-306, 15 Gillson Street, Winnipeg, MB R3T 5V6, Canada. Tel.: +1 204 474 9535.

E-mail addresses: Masoud.Asadzadeh@umanitoba.ca (M. Asadzadeh), luis.leon@ec.gc.ca (L. Leon), wayang@uoguelph.ca (W. Yang), david.bosch@ars.usda.gov (D. Bosch).

http://dx.doi.org/10.1016/j.jhydrol.2015.12.056
0022-1694/

concentration however, the one-day gap is less consistent. This gap is called the *one-day offset issue* (1dOI) in this study and appears to be impossible/irrational to be emulated by hydrologic models with daily simulation time step (Uzeika et al., 2012). 1dOI is investigated in this research study by analyzing sub-daily precipitation and runoff datasets and the corresponding aggregated daily datasets in an example watershed with an estimated sub-daily time of concentration.

Hydrologic models with sub-daily simulation time step such as WATFLOOD (Kouwen, 1988), MIKE SHE (Refsgaard and Storm, 1995), and WetSpa (Liu et al., 2003) should be expected to emulate 1dOI if sub-daily input forcing datasets are provided. However, models with daily simulation time step such as HSPF (Bicknell et al., 1997), Soil and Water Assessment Tool (SWAT) (Arnold et al., 1998), and EPIC and APEX models (Wang et al., 2012) should not be forced through model calibration to emulate 1dOI. Instead, the impact of 1dOI should be eliminated or minimized in the calibration of daily simulation models; otherwise, it can negatively impact the model evaluation metrics and ultimately mislead the parameter estimation process.

The model parameter estimation can be performed in a trial-and-error approach (manual calibration) or in an automatic simulation–optimization approach. In both approaches, the model performance needs to be subjectively evaluated through a visual result assessment and/or objectively measured by quantitative model performance metrics (Boyle et al., 2000; Eckhardt and Arnold, 2001). Visual comparison of measured versus simulated output fluxes is the most fundamental hydrologic model evaluation method (ASCE, 1993; Legates and McCabe, 1999; Krause et al., 2005; Moriasi et al., 2007). The most common visual comparison tool is the time series plot that clearly shows the model performance in estimating timing and magnitude of events. This method can easily detect and tackle 1dOI. For example, Bosch et al. (2004) shifted the whole simulated hydrograph one day forward to reduce the impact of 1dOI on the evaluation of the daily simulation model. While the visual comparison can be very effective in manual calibration, specific metrics are required to quantitatively evaluate the model performance during automatic calibration. A similar hydrograph shifting approach should be taken in automatic calibration of daily simulation models to minimize the impact of 1dOI on the parameter estimation results.

The main goal of this study was to introduce a methodology for eliminating the misleading impact of 1dOI on automatic calibration of daily hydrologic simulation models. To this end, an algorithm called Shifting Hydrograph In order to Fix Timing (SHIFT) was introduced and examined for automatic calibration of a suite of synthetic daily simulation models with a known perfect parameter setting and controlled number of 1dOI. SHIFT was also implemented in three real case studies that had been modeled in SWAT and showed symptoms of 1dOI. SWAT is a well-known and widely used watershed model for simulating the hydrology (e.g. Eckhardt and Arnold, 2001; Wu and Johnston, 2007; Easton et al., 2008; Li et al., 2010; Strauch et al., 2012; Meaurio et al., 2015; Grusson et al., 2015) and/or water quality (Abbaspour et al., 2007, 2015; Bekele and Nicklow, 2007; Tolson and Shoemaker, 2007; Boulange et al., 2014; Yen et al., 2015) at the sub-basin level with daily time step. While many studies utilize SWAT to build a computer model that can accurately simulate the current (historical) state of a watershed, some other studies use SWAT to analyze and predict the watershed behavior in new land-use, management practices and/or climate scenarios, e.g. see Bracmort et al. (2006), Arabi et al. (2007), Ficklin et al. (2009), Baker and Miller (2013), Faramarzi et al. (2013), Molina-Navarro et al. (2014), Awan and Ismaeel (2014) and Neupane and Kumar (2015). Also, the literature shows successful application of SWAT to a range of watershed sizes from a few square kilometers in Kannan et al. (2007) and Arabi et al. (2007) to thousands of square kilometers in Ficklin et al. (2009) and continental-scale in Schuol et al. (2008), Faramarzi et al. (2013) and Abbaspour et al. (2015). SWAT users also have the option to run a sub-daily simulation as in Di Luzio and Arnold (2004) if sub-daily datasets are available.

The daily calibrated SWAT models with and without the application of SHIFT were compared to assess the effectiveness of SHIFT on reducing the misleading impact of 1dOI on parameter estimation results. As an alternative approach, SWAT was calibrated with a weekly model evaluation time step which is supposed to be less sensitive to 1dOI.

2. Methods and materials

2.1. Model evaluation performance metrics

Various metrics have ever been introduced and utilized in the literature of hydrologic model calibration to evaluate the performance of daily simulation models. Many of these metrics are based on the aggregation of daily residuals between a measured flux, e.g. streamflow, and the corresponding model output flux. In order to avoid the compensating error, these metrics take the absolute or squared value of the residuals in each time step, so they are very sensitive even to small gaps between simulated and measured datasets (Beven, 2012). The impact of 1dOI on the daily value of the Nash–Sutcliffe efficiency (*NSE* in Nash and Sutcliffe (1970)) metric and its successor the Kling–Gupta efficiency (*KGE* in Gupta et al. (2009)) metric is assessed in this study.

NSE is arguably one of the most widely used performance metrics in hydrologic model calibration (Moriasi et al., 2007; Gupta et al., 2009). The mathematical form of *NSE* is shown in Eq. (1) where the numerator is the sum of squared residuals between each pair of measured M_t and simulated S_t output fluxes over the evaluation time period T with time step t that can be equal to or larger than the simulation time step. The denominator in Eq. (1) normalizes the metric by the total squared variation in the measured dataset from its average value $\overline{M}$ to make the metric dimensionless. *NSE* can range from $-\infty$ to 1, from the worst to the perfect match, respectively. For a general daily streamflow simulation model, *NSE* below zero is often deemed questionable (Schaefli and Gupta, 2007), while *NSE* higher than 0.5 can be deemed acceptable if percent bias and the ratio of the root mean squared error to the standard deviation of measured data are also acceptable (Moriasi et al., 2007). Daily *NSE* is very sensitive to 1dOI because it is based on the sum of squared residuals. So, it was adopted in this study to show how this sensitivity can mislead the automatic calibration of daily simulation models when 1dOI existed. For example, Bosch et al. (2004) showed that shifting the simulated hydrograph one day forward can increase daily *NSE* from −0.03 to 0.53 and change our assessment about the calibrated model performance from questionable to acceptable.

$$NSE = 1 - \frac{\sum_{t=1}^{T}(M_t - S_t)^2}{\sum_{t=1}^{T}(M_t - \overline{M})^2} \quad (1)$$

Gupta et al. (2009) decomposed *NSE* into three components that compare the simulated and measured datasets based on the linear correlation coefficient, bias, and relative variability and showed that maximizing *NSE* can lead the calibration toward underestimating variability. To solve this issue, Gupta et al. (2009) introduced *KGE* shown in Eq. (2) that measures the Euclidean distance in a three dimensional space between the ideal point (1, 1, 1) and the linear correlation coefficient (r), relative variability (α), and ratio of mean (β) calculated for a time series of simulated and measured variables. It appears that, *KGE* is sensitive to 1dOI