Does a fifteen-hour shift make much difference? – Influence of time lag between rainfall and discharge data on model calibration

Jie Jian[a], Justin Costelloe[a], Dongryeol Ryu[a] and Q. J. Wang[a]

[a]Department of Infrastructure Engineering, The University of Melbourne, Victoria 3010, Australia

Email: jjian@student.unimelb.edu.au

Abstract: Hydrological modelling requires accurate datasets that include input forcing data (precipitation and PET) and response data (mainly discharge data). For models based on daily data, ideally, both the input and response data are resampled in identical time intervals. However, there could be time lags between the rainfall and discharge data due to multiple time intervals used in real situations. The daily discharge data provided by the Australian Bureau of Meteorology Water Data Online (Q_WDO) has a fifteen-hour delay to the daily rainfall and PET data, while the discharge data from the Australian Bureau of Meteorology Hydrological Reference Stations (Q_HRS) can match the daily input forcing data. Both the Q_WDO and Q_HRS were applied in two conceptual rainfall-runoff models (GR4J and PDM) to investigate the influences of time lag between input and response data. The effects of different catchment characteristics on this issue were examined. The results suggested that the time lag between input and response data could significantly decrease the performance of hydrological models that are based on daily datasets, and the degradation of performance is more likely to happen in catchments with smaller areas, lower elevations, wetter climate conditions and higher interannual variability. The time lag issue can be significant and more attentions should be paid to resampling daily datasets.

Keywords: Hydrological modelling, time lag, daily data, catchment characteristics

1.   INTRODUCTION

Hydrological modelling provides a way for understanding water cycles, hydrological systems and associated physical processes (Gupta et al., 1998; Vörösmarty et al., 2001). The modelling process requires accurate datasets that include input forcing data (rainfall and PET) and response data (discharge). Particularly, for models based on daily data, the input and response data should be resampled in identical time intervals in which the daily data are modelled.

In practice, there can be different time conventions for resampling daily data. For example, the Australian Bureau of Meteorology provides daily rainfall and PET using the one-day time interval from 9:00 of the previous day to 9:00 of the current day, while daily discharge data are more commonly derived for a different 24-hour time interval, such as midnight yesterday to midnight today. The daily discharge published on Hydrological Reference Stations (http://www.bom.gov.au/water/hrs/) are calculated from 9:00 yesterday to 9:00 today, which is matched with the daily input data. But on Water Data Online (http://www.bom.gov.au/waterdata/), the daily discharge is calculated from 0:00 today to 0:00 tomorrow, and this is the most common time interval for State-based discharge data. Moreover, after checking the daily data, we found that only the states in eastern Australia (VIC, NSW, QLD and ACT) actually follow the nominated interval for both datasets, the other states in central and western Australia (SA, NT, WA) have other time intervals, even though they are marked as 9:00 to 9:00 or 0:00 to 0:00 (shown in Table 1).

The mismatch in resampled time intervals between rainfall and discharge leaves some important questions about its influence on hydrological analysis and modelling such as i) Would the time mismatch affect the hydrological model calibration accuracy? ii) How does the extent, by which the mismatch influences calibration accuracy, depend on catchment characteristics?

Table 1. Information of two discharge datasets from BoM

Source of daily Q data	Time interval to calculate daily data (local time)	Time lag between rainfall and discharge (hour)	State
Q_HRS (nominally 9:00 to 9:00)[1]	01:00 yesterday – 01:00 today	-8	WA
	23:30 the day before yesterday – 23:30 yesterday	-9.5	SA, NT
	09:00 yesterday – 09:00 today	0	VIC, QLD, NSW, ACT
Q_WDO (nominally 0:00 to 0:00)[2]	16:00 today – 16:00 tomorrow	31	WA
	14:30 today – 14:30 tomorrow	29.5	SA, NT
	0:00 today – 0:00 tomorrow	15	VIC, QLD, NSW, ACT

[1]Sourced from Hydrological References Stations.
[2]Sourced from Water Data Online. It is currently being updated to 9:00 to 9:00 after we informed them of the issue.

In this paper, the two discharge datasets (Q_HRS and Q_WDO) for the eastern states were used in hydrological calibration to examine the influences of the 15-hour time lag between the input forcing data and discharge data on calibration performance. The rainfall and PET data are derived from 9:00 yesterday to 9:00 today, and the discharge data from Hydrological Reference Stations (Q_HRS) and Water Data Online (Q_WDO) are applied separately. In addition, catchment characteristics are investigated to see if they are associated with changes in calibration performance due to the time lag.

2.   STUDY CATCHMENTS, MODELS AND DATASETS

Ideally, all hydrological reference stations (222 catchments) should be tested. However, since the discharge datasets in WA, SA and NT are not consistent with the datasets for eastern Australia (VIC, NSW, QLD and ACT) in either the HRS or WDO datasets, we only examine the eastern catchments in this study, and the total number of catchments is 160.

Two lumped conceptual models – GR4J and PDM – are employed to test if different model structures have effects on the time lag issues. The information on parameters is listed in Table 2.

Jian *et al*., Does a fifteen-hour shift make much difference? – Influence of time lag between rainfall and discharge data on model calibration

Table 2. Parameters of GR4J and PDM

GR4J (Perrin et al., 2003)		PDM (Moore, 2007)	
Parameter	Description	Parameter	Description
x_1	Maximum capacity of the production store	C_{max}	Maximum store capacity
x_2	Groundwater exchange coefficient	b	Exponent of the pareto distribution controlling spatial variability of store capacity
x_3	One day ahead maximum capacity of the routing store	C_{min} Ratio	Minimum store capacity ratio
x_4	Time base of unit hydrograph	b_e	Exponent on actual evaporation function
		b_g	Exponent of recharge function
		k_g	Groundwater recharge time constant
		St Ratio	Soil tension storage capacity ratio
		k_b	Baseflow time constant
		k_1	Time constant of cascade linear reservoir
		k_2 Ratio	Time constant of cascade linear reservoir

The models are calibrated using daily data from 1980 to 1995. This excludes the "Millennium Drought" period, which could affect model calibration performance in complicated ways in some catchments (Dijk et al., 2013; Saft et al., 2016).

Eight catchment characteristics are tested in this study. They are topographic variables (catchment area, mean elevation, elevation range, mean slope) and hydroclimatic variables (mean annual rainfall, mean daily discharge, coefficient of variation (C_v) of annual discharge and dryness index (PET/P)).

3. METHODS

The Nash-Sutcliffe Efficiency (NSE) is used to evaluate the calibration results (Nash & Sutcliffe, 1970). Both of the models are calibrated using the 9:00 to 9:00 input forcing data against two sets of discharge data:

(1) Q_HRS: calculated from 9:00 the previous day to 9:00 the current day (control case).
(2) Q_WDO: calculated from 0:00 the current day to 0:00 the next day (experimental case).

After the calibration, the percentage of changes in NSE values is calculated as:

$$Per.\,change\;of\;NSE = \frac{NSE_{Q\_WDO} - NSE_{Q\_HRS}}{NSE_{Q\_HRS}} \times 100 \qquad (1)$$

The principle component analysis (PCA) is applied to investigate the influence of different catchment variables on model performance, and the correlation matrix is calculated to show the linear correlation between the percent change of NSE and each variable.

Although the NSE differences did not show strong linear correlation with some of the variables (Section 4), some non-linear relationships may exist. In order to explore the underlying trends, the 160 catchments were divided into five groups according to the following steps:

- The 160 catchments are ranked based on the values of a selected variable in ascending order.
- The 160 catchments are divided into 5 groups evenly (each group contains 32 catchments). The first group contains the 32 catchments that have the variable values from the smallest to the 32th; and the fifth group contains the 129th to the largest.
- For each group, the figures showing the percent change of NSE versus the variable values are drawn to illustrate their relationship.

4. RESULTS AND DISCUSSIONS

The differences of calibration performance for the two datasets are shown in Figure 1. In most catchments, the NSE values decreased when Q_WDO was applied. The percentage changes of NSE values were mainly distributed in the ranges of 0 ~ -20% for GR4J and 0 ~ -40% for PDM, and it could be as much as -100%. Based on the result, it is obvious that the fifteen hours' time lag between input and response data could significantly hamper the calibration performance, leading to large inaccuracy.

Model parameters are related to hydrological physical processes, such as soil storage, routing processes, groundwater exchange, etc. To determine the potential links between the calibration change and physical processes, the correlation coefficients of the percentage change of NSE and each parameter (from the control

Jian *et al*., Does a fifteen-hour shift make much difference? – Influence of time lag between rainfall and discharge data on model calibration

case) were calculated (shown in Table 3). The influences of the time lag issue in GR4J were most highly correlated with parameter x_4 – "Time base of unit hydrograph", while for PDM it was mainly related to parameter k_1 – "Time constant of cascade linear reservoir". Both x_4 and k_1 are the key components that control the internal routing processes (Moore, 2007; Perrin et al., 2003), which indicate that the catchments with higher routing capacity (i.e. larger or with slower response times) are more likely to mitigate the influences of the time lag between input forcing data and discharge data.
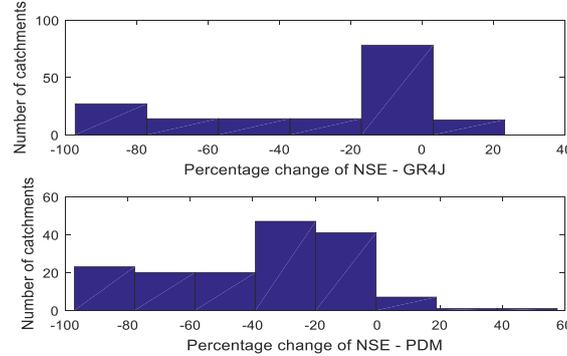

Figure 1. The distribution of the percentage change in NSE values

Table 3. Correlation between the percentage change of NSE and each parameter

Model	Parameters									
GR4J	x_1 0.001	x_2 0.076	x_3 0.133	x_4 0.412						
PDM	C_{max} 0.018	b -0.087	C_{min} Ratio 0.236	b_e -0.092	b_g -0.027	k_g 0.067	St Ratio 0.136	k_b 0.137	k_1 0.510	k_2 ratio 0.058

4.1. PCA and correlation matrix

Principle component analysis was applied to the percentage change of NSE and each catchment characteristics (Figure 2), and the corresponding correlation matrix is shown in Table 4.

Table 4. Correlation matrix of percentage change of NSE (dependent variables) and each catchment characteristics (independent variables)

	Per. change of NSE -GR4J	Per. change of NSE -PDM	area (km²)	mean elevation (m)	elevation range (m)	mean Slope	mean annual rainfall (mm)	mean daily Q (mm)	C_v annual Q	Dryness Index (PET/Rain)
Per. change of NSE -GR4J	1	0.860	0.196	0.202	0.146	0.003	-0.328	-0.203	-0.290	0.111
Per. change of NSE -PDM	0.860	1	0.249	0.231	0.194	0.080	-0.231	-0.129	-0.295	0.098
area (km²)	0.196	0.249	1	-0.154	-0.061	-0.253	-0.299	-0.185	0.188	0.717
mean elevation (m)	0.202	0.231	-0.154	1	0.447	0.449	0.003	-0.042	-0.213	-0.298
elevation range (m)	0.146	0.194	-0.061	0.447	1	0.745	0.228	0.116	-0.400	-0.385
mean Slope	0.003	0.080	-0.253	0.449	0.745	1	0.478	0.385	-0.574	-0.624
mean annual rainfall (mm)	-0.328	-0.231	-0.299	0.003	0.228	0.478	1	0.899	-0.468	-0.663
mean daily Q (mm)	-0.203	-0.129	-0.185	-0.042	0.116	0.385	0.899	1	-0.471	-0.504
C_v annual Q	-0.290	-0.295	0.188	-0.213	-0.400	-0.574	-0.468	-0.471	1	0.586
Dryness Index (PET/Rain)	0.111	0.098	0.717	-0.298	-0.385	-0.624	-0.663	-0.504	0.586	1

According to Table 4 and Figure 2, the NSE changes in GR4J and PDM are highly correlated, which means that the two models of different structures showed similar performance changes in response to the time lag

issue. Besides, mean annual rainfall and mean daily Q are highly correlated and closely located in the PCA plot, so that they can be combined and regarded as one variable in the following study.

Based on the correlation coefficients (*R* value) and the locations in PCA plot, the independent variables can be divided into three groups: weak positive correlation group (*R* value: 0.14~0.25), weak negative correlation group (*R* value: -0.12~-0.33) and no correlation group (*R* value: 0~0.11). Positive correlation group contains area, mean elevation and elevation range, while negative correlation group consists of mean annual rainfall, mean daily Q and C_v of annual Q. The reason that the positive and negative correlations are weak is that those variables are nearly orthogonal to the NSE changes (Figure 2), which indicates that these variables showed poor explanatory power. Compared to the positive correlation group, negative correlation group had stronger effects on the NSE changes. However, the no correlation group, which has the variables mean slope and dryness index, showed little influences on time lag issue.
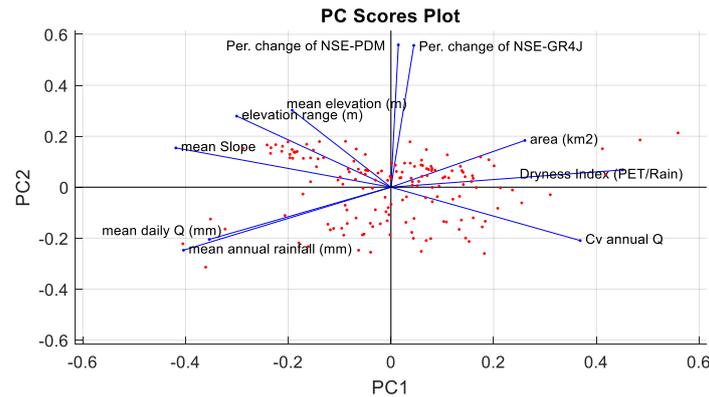

Figure 2. Principle component analysis results

4.2. The effects of each sensitive variable

To investigate possible non-linear relationships between the percentage change of NSE and each variable, the 160 study catchments were divided into five groups. The distributions of NSE change in each group are illustrated in Figure 3 and the mean values of each group are shown in Figure 4. The figures are drawn for the sensitive variables only (mean annual rainfall and mean daily Q are similar, so that only mean annual rainfall is plotted to represent both of them).

Generally, the groups with smaller NSE changes had narrower distributions, while the groups showed larger changes of NSE had larger uncertainty (Figure 3).

There is an increasing trend between catchment area and NSE changes (Figure 3a and Figure 4a). Catchments with smaller areas tended to have larger NSE changes, while larger catchments were less likely to be affected by the time lag issue. In addition, smaller catchments displayed larger uncertainty since the smaller sites had larger range of NSE changes (Figure 3a). It is understandable because the response time of runoff to catchment outlet in smaller catchments are shorter than those in larger ones, resulting in lower capacity to mitigate the time lag effects. This trend is more obvious for the GR4J model. The mean elevation variable showed similar patterns with catchment area (Figure 3b and Figure 4b). Catchments in higher locations may have longer pathways for water to reach the catchment outlet, thus these catchments could be more resilient to the mismatch between rainfall and discharge. However, this trend is more inconsistent across the range compared to the results of area variable (Figure 3b).

According to Figures 3c-3d and Figures 4c-4d, the hydroclimatic variables presented opposite trends to the above topographic variables. Catchments with higher rainfall (and Q) showed more degradation of performance in response to the time lag. This is probably because wetter catchments respond quicker to rainfall and the corresponding routing processes are shorter. As a result, the capacity to reduce the time lag effects is lower than for drier catchments. The similar trend is found in C_v of annual Q vs NSE change, which shows that the catchments with higher interannual variability feature larger NSE changes as a result of the time lag. In fact, the NSE of the calibrated model without of lag show declining trend with increasing C_v over the study catchments, however the high-C_v catchments with lower NSE seem to be more vulnerable to the time lag between the rainfall and the discharge.

Jian *et al*., Does a fifteen-hour shift make much difference? – Influence of time lag between rainfall and discharge data on model calibration
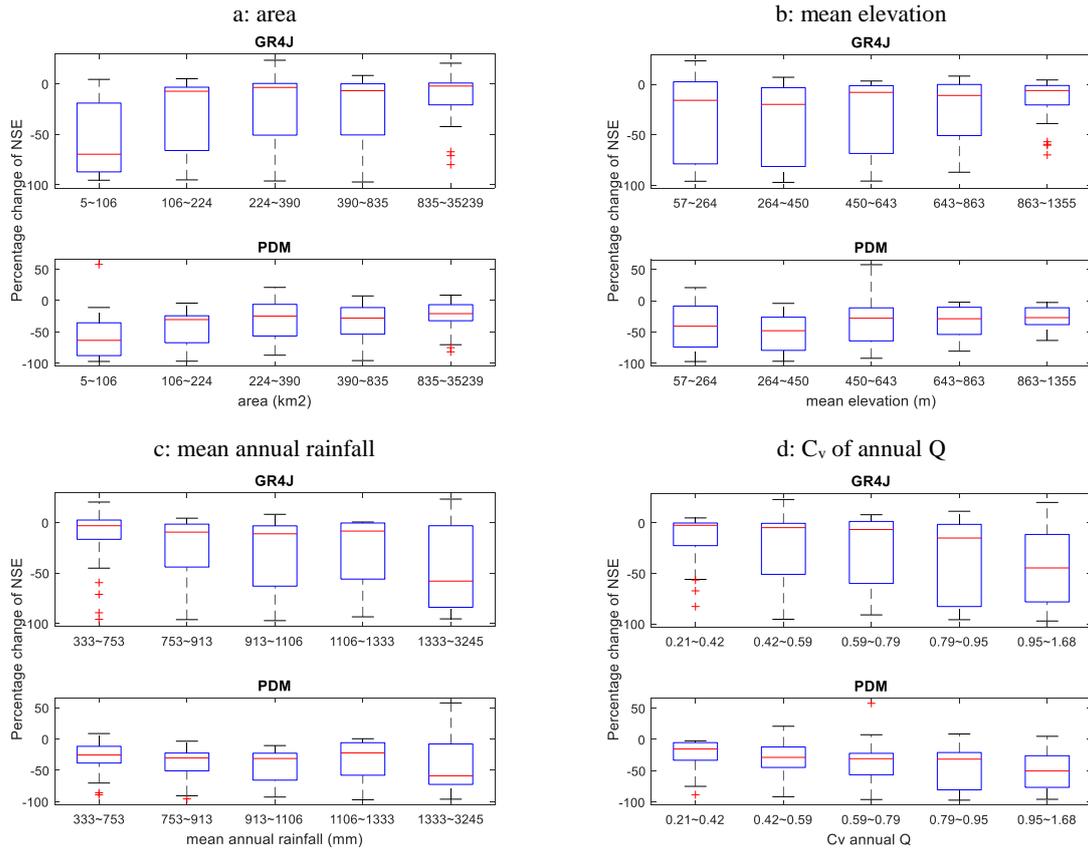
Figure 3. The distribution of NSE change for each catchment group for four typical variables
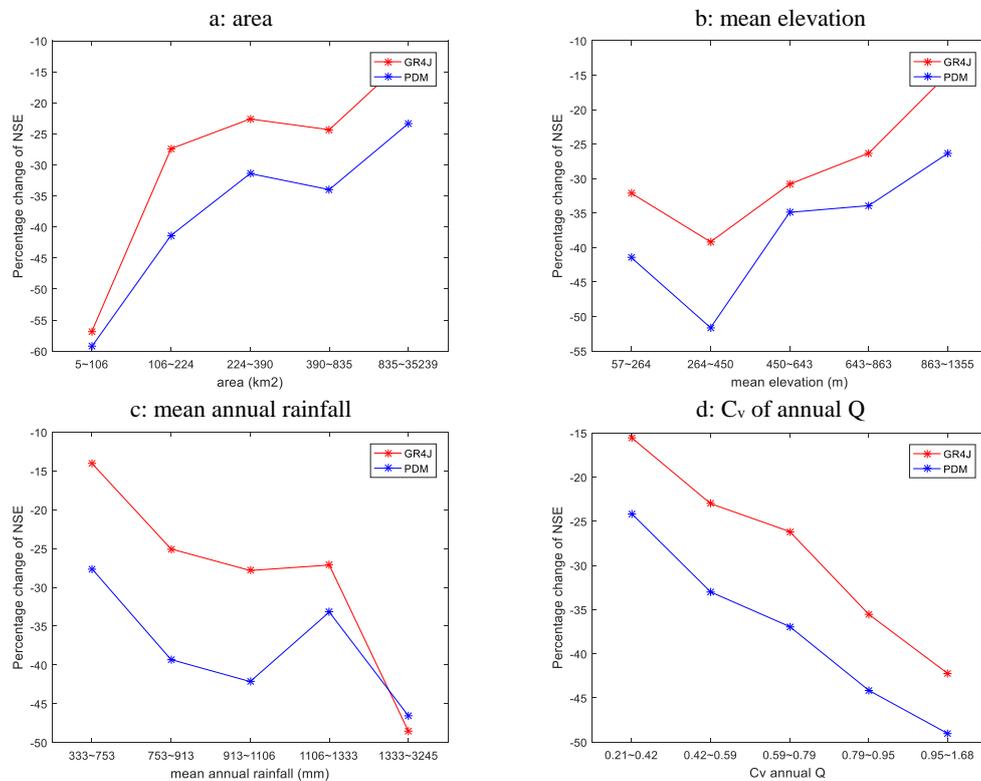
Figure 4. Mean values of each catchment group in Figure 3

Jian *et al*., Does a fifteen-hour shift make much difference? – Influence of time lag between rainfall and discharge data on model calibration

5. CONCLUSION

Accurate input forcing data and discharge data are essential in hydrological modelling, especially for the models which use daily data. However, due to the inconsistent calculation methods and different time intervals in the calculation of daily data, time lags between rainfall and discharge are common in data used by hydrologists.

One dataset from the Bureau of Meteorology (WDO) has a 15-hour time lag between input forcing data and discharge data. Calibrations based on matched and mismatched data were conducted and the results showed that the time lag issue could significantly decrease the calibration performances. The influences were related to the routing processes and tended to happen in the catchments with smaller areas, lower elevations, wetter climate conditions and higher interannual variability. PDM was more likely to be affected by the time lag issue, while GR4J was more sensitive to hydroclimatic variables such as mean annual rainfall and mean daily discharge. It is recommended that closer attention should be paid to generating discharge datasets with matching daily time-spans to climatic input data.

REFERENCES

Dijk, A. I., Beck, H. E., Crosbie, R. S., Jeu, R. A., Liu, Y. Y., Podger, G. M., . . . Viney, N. R. (2013). The Millennium Drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society. *Water Resources Research, 49*(2), 1040-1057.

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research, 34*(4), 751-763.

Moore, R. (2007). The PDM rainfall-runoff model. *Hydrology and Earth System Sciences Discussions, 11*(1), 483-499.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology, 10*(3), 282-290.

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

Saft, M., Peel, M. C., Western, A. W., & Zhang, L. (2016). Predicting shifts in rainfall-runoff partitioning during multiyear drought: Roles of dry period and catchment characteristics. *Water Resources Research*.

Vörösmarty, C., Askew, A., Grabs, W., Barry, R., Birkett, C., Döll, P., . . . Kitaev, L. (2001). Global water data: A newly endangered species. *Eos, Transactions American Geophysical Union, 82*(5), 54-58.