

# Calibration of hydrological models allowing for timing offsets

J. Lerat<sup>a</sup> and R. S. Anderssen<sup>b</sup>

<sup>a</sup>*Bureau of Meteorology, Environmental Research Division, Canberra*

<sup>b</sup>*CSIRO Digital Productivity*

*Email: [J.Lerat@bom.gov.au](mailto:J.Lerat@bom.gov.au)*

**Abstract:** A central activity in hydrological modelling and simulation is the construction of predictors of extreme events, such as floods and droughts, on the basis of available historical observational data. For a given extreme event, the construction involves not only the choice of an appropriate parametric model for the predictor but also the identification of the key feature of the extreme event to be predicted, since it determines the focus for the associated parameter identification. For example, in flood forecasting, though the peak size of the flood represents a crucial aspect, the key feature is the accurate prediction of the timing of the arrival of the flood peak and its shape, after a rainfall event. The arrival time's importance relates not only to its practical significance from a community perspective but also to its theoretical significance in that the accurate prediction by the model of the size and shape of the flood peak depends crucially on the accurate prediction of the arrival time. On the other hand, in drought prediction, the key features are the duration and intensity of the drought.

However, the protocols that are commonly used to perform the associated parameter identification only focus on minimizing, with respect to some appropriately chosen error measure such as least squares, the *amplitude offset* discrepancy between the modelled and observed data at a representative set of time points. Such approaches can lead to significant errors in the identified parameters when the prediction, relative to the observed, contain (small) *timing offset* errors, reflected in the fact that the predicted flood peak arrival and shape does not accurately match the observed peak arrival and shape. In such situations, the parameter identification protocol has failed to successfully match the predicted flood peak arrival and shape with the observed.

The paper presents a new approach to overcome this problem by estimating jointly the amplitude and timing offsets. For a given flood event, it is based on comparing the difference between the cumulative sums of the predicted and observed flood, in order to assess the extent to which timing and/or amplitude offsets have occurred and how to correct for them. The method is applied to a synthetic case study where known amplitude and timing errors are introduced in the rainfall time series, and a simple hydrological model is fitted to perturbed data.

**Keywords:** *Timing errors, amplitude errors, rainfall-runoff modelling, non-linear storage model*

## 1 INTRODUCTION

The essence of the hydrological prediction problem is illustrated in Figure 1, where, for four representative floods on the Corang river, the observed and predicted flood events have been plotted in the top panels. Each prediction was generated by the following storage model

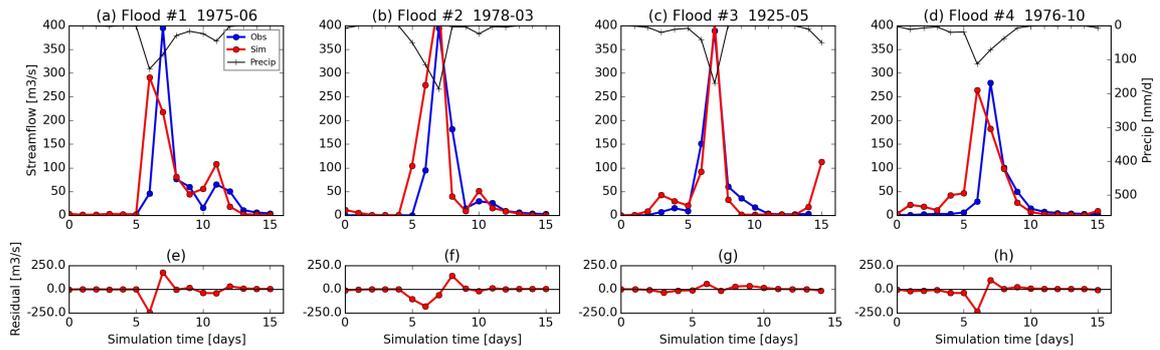
$$\frac{dV}{dt} = \theta_3 r(t) - g(t), \quad V = V(t) \quad (1)$$

$$g(t) = Q \left( \frac{V}{\theta_1} \right)^{\theta_2} \quad (2)$$

with  $g$ ,  $V$  and  $r$  respectively denoting, at time  $t$ , the simulated flow, the accumulated catchment storage volume and the rate of precipitation.  $\theta_1$  (storage volume,  $\text{m}^3$ ),  $\theta_2$  (storage exponent, non dimensional) and  $\theta_3$  (loss factor, non dimensional) are the model parameters and  $Q$  a fixed constant representing a reference flow. When  $\theta_2 = 1$  and  $\theta_2 = 2$ , the model can be solved analytically (Bentura and Michel, 1997). A comparison of the observed and predicted curves in Figures 1 illustrates the essence of the problem to be solved. For a given flood event, the match between the observed and predicted curves can be assessed in two ways:

- (i) *The Amplitude Offsets:* The heights and shapes of the predicted and observed should be similar, in that the indicative unimodal or multimodal structure in the observed is preserved in the predicted and that they generate similar volumetric flows.
- (ii) *The Timing Offsets:* The time of arrival of the predicted and observed flood peak and their overall shapes should be closely matched.

For example, for Flood #3, there appears to be a reasonable match, whereas, for the other three, the match is less satisfactory. In particular, the timing of the peak in floods #1 and #4 is predicted a day before it occurs in the observed data. These timing offsets results in large model residuals as can be seen in Figures 1.e and 1.h. However, the shape of the modelled hydrograph appears reasonably close to the observed one for both floods. As a result, it might be counter productive to minimise the amplitude residuals, when the modelled is already representing the flood dynamic correctly.



**Figure 1 :** Four representative flood events in the Corang River as measured at Hockeys (NSW), where  $\theta_1 = 200,000 \text{ m}^3$ ,  $\theta_2 = 2$ ,  $\theta_3 = 1.3$ . Precipitation  $r$  is represented in inverse axis.

Few authors have addressed the issue of estimating timing offsets. Liu et al. (2011) proposed a method based on the phase of complex wavelets. This method was successfully applied to hourly runoff simulations, but requires heavy computation. A more simple approach was developed by Ewen (2011) where amplitude and timing offsets are estimated jointly using pattern matching techniques. The approach proposed in this paper has a similar objective than the study by Ewen (2011), but introduces an additional constraint on mass balance as indicated in the next section.

## 2 TIMING AND AMPLITUDE OFFSET ESTIMATION

When timing offsets are ignored, the relationship between an observed ( $f$ ) and modelled ( $g$ ) times series is classically written

$$f(t) = g(t) + \epsilon(t), \quad (3)$$

where  $\epsilon(t)$  is the amplitude offset at time  $t$  (as presented in figures 1.e to 1.h), The model response  $g$  may be controlled by a set of parameters  $\theta$  that can be optimised by minimising a functional  $S$  of  $\epsilon$  over a simulation period  $T$ . For example, this functional can be the sum of squared amplitude offsets where

$$S(T, \epsilon) = \int_0^T \epsilon(t)^2 dt \quad (4)$$

In this paper, we extend Equation 3 by introducing a timing offset function  $\phi$  following:

$$f(t) = \dot{\phi}(t) g(\phi(t)) + \epsilon(t) \quad (5)$$

Where  $t \in [0, T]$ ,  $f$  is the observed signal,  $g$  is the modelled signal,  $\phi$  and  $\epsilon$  are the timing and amplitude offsets, respectively. The function  $\phi$  has the following properties:

$$\phi \text{ is continuous and } \forall t, \dot{\phi}(t) > 0 \quad (6)$$

$$\phi(0) = 0 \text{ and } \phi(T) = T \quad (7)$$

We can see that Equation 5 reverts to Equation 3 if  $\phi(t) = t$ , i.e. in the case where timing offset is negligible. Second, the conditions stated in Equations 6 and 7 impose that the timing offset function remains a plausible distortion of the time continuum.  $\phi$  has to be continuous because  $\phi(t)$  cannot jumps from one value to another. It has to be a strictly increasing function because time can progress forward only. The boundary conditions expressed in Equation 7 ensures that the distortion of time covers the whole simulation period. An immediate consequence of Equation 6 and 7 is that there exists a unique inverse function  $\phi^{-1}$  such that

$$\forall t \in [0, T] \quad \phi^{-1}(\phi(t)) = t \quad (8)$$

The role of the derivative of  $\phi$  in the right-hand side of Equation 5 is explained as follows. Assuming that Equation 5 is integrated over the simulation period  $T$  and using Equation 7, we get

$$\int_0^T f(t) dt = \int_0^T g(t) dt + \int_0^T \epsilon(t) dt \quad (9)$$

This last equation shows that the integral of the amplitude offset  $\epsilon$  over the simulation period is not dependent on the timing offset  $\phi$  because the integral of  $f$  and  $g$  are fixed constant. This property is important in hydrological modelling in order to ensure that the timing offset function is not introducing a change in the overall mass balance during the simulation period.

As indicated before, a special case of Equation 5 occurs when there is no timing offset, i.e. leading to the trivial timing offset function

$$\forall t \in [0, T] \quad \phi_1(t) = t \quad (10)$$

Another special case can be highlighted when there is no amplitude offset, i.e. when

$$\forall t \in [0, T] \quad \epsilon(t) = 0 \quad (11)$$

In this case, the timing offset function  $\phi$  can be found by integrating equation 5 from 0 to  $t$ , which leads to

$$\int_0^t f(t) dt = \int_0^t \dot{\phi}(t) g(t) dt \quad (12)$$

$$\Leftrightarrow F(t) = G(\phi(t)) \quad (13)$$

Where  $F$  and  $G$  are the integrals of  $f$  and  $g$ , respectively. Two problems can arise at this stage, depending on the characteristics of  $G$ . First, Equation 13 can be solved if the function  $G$  is invertible. In this case, the timing offset function becomes

$$\phi_2(t) = G^{-1}(F(t)) \quad (14)$$

However,  $G$  is invertible if and only if  $g$  is a strictly positive function, which may not be valid with run-off simulations as can be seen in Figure 1 where many zero flow values are produced. Second, assuming that the inverse  $G^{-1}$  exists, Equation 14 may violate the condition stated in Equation 7 if  $F(T) \neq G(T)$ . In this case  $\phi(T)$  is not equal to  $T$ . Consequently, a new function  $H$  is introduced such that

$$H(t) = \frac{F(T) - aT}{G(T)} G(t) + a t \quad (15)$$

Where  $a$  is a small positive constant such as  $a \ll F(T)/T$ . The timing offset function is finally derived by replacing  $G$  by  $H$  in equation 14:

$$\phi_2(t) = H^{-1}(F(t)) \quad (16)$$

It can be easily verified that  $\phi_2$  satisfies the conditions stated in equations 6 and 7.

The two special cases described in equations 10 and 11 are asymptotic assumptions that may not hold for real simulations such as the ones presented in Figure 1. As a result, the timing offset function proposed in this paper is a weighted combination of both the special cases  $\phi_1$  and  $\phi_2$ :

$$\phi_3(t) = \alpha \phi_1(t) + (1 - \alpha) \phi_2(t) \quad (17)$$

where  $\alpha$  is a constant varying between 0 and 1 that controls the importance of the amplitude offset compared to timing offset. It can easily be verified that  $\phi_3$  satisfies the conditions given in equations 6 and 7.

### 3 MODEL CALIBRATION EXPERIMENT

In the context of calibrating the model described in equation 1, one has to deal with data that are prone to errors which may significantly affect the fitting process. In particular, rainfall data are known to contain significant errors (Kavetski et al., 2006) due to the high spatial variability of rainfall and the limited number of raingauges that are used to estimate it. Streamflow data are also affected by measurement errors, but at a lower order of magnitude due to the precision of water level measurement devices. Consequently, the calibration experiments that are used in this paper are focused on errors introduced on the rainfall data.

In order to simplify the fitting process, the model parameter  $\theta_3$  is set to 1 (see Equation 1), inducing no loss or gain. In addition, the parameter  $\theta_2$  is set to a fixed value depending on the experiment as per Table 1. As a result, the only fitted parameter is  $\theta_1$ .

The experiments are run in three steps:

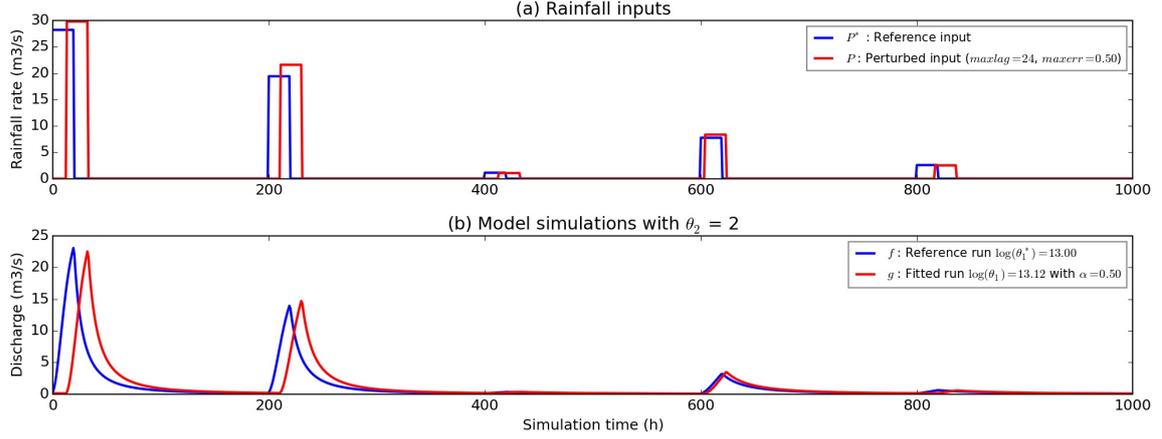
1. The model described in Equations 1 is first run using a value  $\theta^*$  for the parameter  $\theta$ , a reference flow  $Q = 3$ , and a series of step inputs:

$$r^*(t) = \sum_k^n A_k \left[ H(t - k L) - H(t - k L - D) \right] \quad (18)$$

Where  $H$  is the Heaveside step function,  $n$  is the number of event set to 50,  $L$  is the duration of an event, set to 200 hours,  $D$  is the duration of a step, set to 20 hours, and  $A_k$  is the amplitude of event  $k$  that is sampled from a log-normal distribution with a mean of 3 and a standard deviation of 1. This run produces what is considered to be the observed flow data  $f(t)$ .

2. Subsequently, the input data are corrupted by adding timing and amplitude errors:

$$r(t) = \sum_k^n A_k \left[ H(t - k L - \delta_k) - H(t - k L - D - \delta_k) \right] \nu_k, \quad (19)$$



**Figure 2.** : Example of calibration experiment results with  $\theta_1^* = 13.0$ ,  $\theta_2 = 2$ ,  $maxerr = 0.50$ ,  $maxlag = 24$  and  $\alpha = 0.5$ .

**Table 1.** : Range of configuration parameters used in the calibration experiments

Configuration element	Symbol	Values
Maximum lag error	$maxlag$	[0, 8, 16, ..., 48]
Maximum amplitude error	$maxerr$	[0, 0.1, 0.2, ..., 0.5]
Reference storage volume	$\theta_1^*$	[exp(11), exp(11.2), exp(11.4), ..., exp(13)]
Storage exponent	$\theta_2$	[1, 2]
Timing offset weight	$\alpha$	[0.1, 0.5, 1.0]

where  $(\delta_k)$  and  $(\nu_k)$  are series of amplitude and lag error coefficients, respectively. Both series are sampled from a uniform distribution with bounds  $[0, maxlag]$  for  $\delta_k$  and  $[1 - maxerr/2, 1 + maxerr/2]$  for  $\nu_k$ . The two constants  $maxlag$  and  $maxerr$  vary depending on the experiment as indicated in Table 1. An example of reference and corrupted inputs is presented in Figure 2.a.

- Finally, the model parameter  $\theta_1$  is fitted by minimising the residual sum of squared amplitude offsets described in equation 4, with the amplitude offsets computed as

$$\epsilon(t) = f(t) - \phi_3(t) g(\phi_3(t)) \quad (20)$$

During this process, the model is run with the corrupted inputs  $r$ . The calibration algorithm is the Nelder-Mead simplex algorithm. An example of reference and fitted streamflow simulations is presented in Figure 2.b. In this example, the model is non-linear (i.e.  $\theta_2 = 2$ ) and was fitted with  $\alpha = 0.5$ , which gives an equal weight to timing and amplitude offsets as per equation 17. We can see that the fitted parameter  $\theta_1 = \exp(13.12)$  is close to the reference value  $\theta^* = \exp(13.00)$  in spite of the large timing and amplitude errors introduced in the model inputs ( $maxlag = 24$  hours and  $maxerr = 0.5$ ).

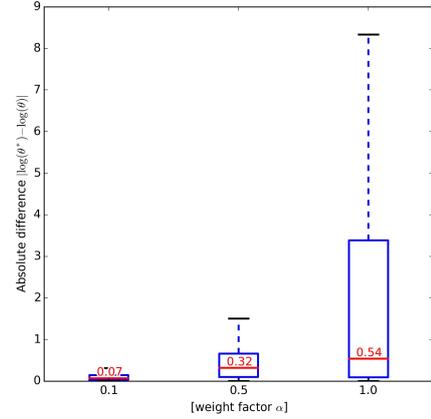
Overall, 5543 experiments were run by combining all configuration parameters indicated in table 1. Each experiment was run 20 times to account for the randomness of the input generation and perturbation process leading to a total of 110,086 experiment runs.

## 4 RESULTS

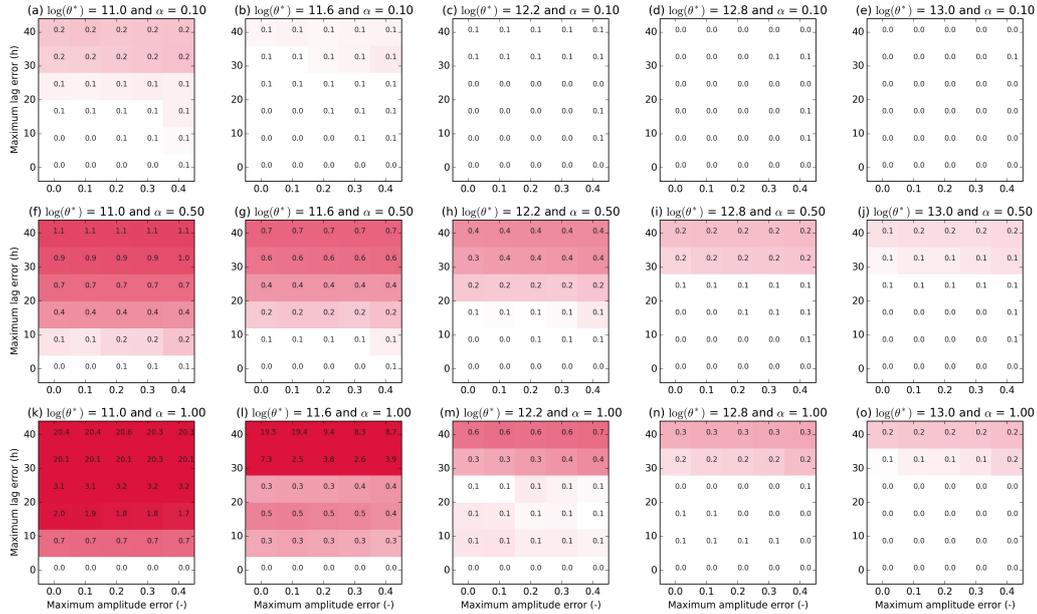
Figure 3 shows a distribution of the absolute difference between the log of the reference parameter  $\log(\theta_1^*)$  and the log of the fitted parameter  $\log(\theta_1)$  across the 110,086 experimental runs. The figure reveals that this difference increases for larger values of  $\alpha$  with the median difference varying from 0.07 for  $\alpha = 0.1$  to 0.54 for  $\alpha = 1.0$ . In other words, the parameter fitting process leads to larger mismatch on the parameter value when the objective function does not consider the timing offset (i.e. when  $\alpha = 1$ ). In this case, Figure 3 shows that the mismatch on the fitted parameter can be large with a significant fraction of the experiments leading to

an difference that is greater than 1 in log-space. This difference is equivalent to a deviation of more than 171% suggesting that lag errors in input data can significantly affect a model calibration process.

Figure 4 details the results presented in Figure 3 by showing the absolute difference obtained for experiments with  $\theta_2 = 1$  (linear model, see Equation 1). Each plot in this figure shows the results for experiments with identical reference parameter  $\theta_1^*$  and weighting factor  $\alpha$ . Lag (*maxlag*) and amplitude (*maxerr*) errors are indicated in X-axis and Y-axis, respectively. Large absolute differences are colored in red to highlight the configurations leading to a significant mismatch between  $\theta_1^*$  and  $\theta_1$ . Figure 4 reveals that the largest mismatch occurs for low values of  $\theta_1^*$  ( $\log(\theta_1^*) = 11.0$ ), when using an objective function that does not account for timing offsets ( $\alpha=1$ ) and moderate to large lag errors ( $maxlag \geq 8$ ), as can be seen in Figure 4.k. This result suggests that timing offsets can be particularly detrimental to the calibration process in catchments with a rapid response (i.e. small storage volume  $\theta_1$ ). In this case, even minor lag errors can lead to a large mismatch in parameter estimation as can be seen in the second rows of Figure 4.k (i.e.  $maxlag = 8$ ). In this plot, we note that the amplitude error (*maxerr*) does not affect the result. This indicates that the objective function is more sensitive to lag errors compared to amplitude errors. This result is similar for other configurations with all plots in Figure 4 showing a strong horizontal pattern, i.e. a limited effect of amplitude errors on the difference between  $\log(\theta_1^*)$  and  $\log(\theta_1)$ .



**Figure 3.** : Distribution of the absolute difference between  $\log(\theta_1^*)$  and  $\log(\theta_1)$  for three values of the weighting factor  $\alpha$  across all 110,086 experimental runs.



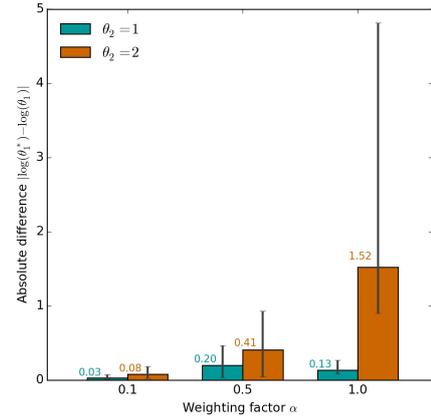
**Figure 4.** : Median value of the absolute difference  $|\log(\theta_1^*) - \log(\theta_1)|$  for given values of  $\theta_1^*$ ,  $\alpha$ , *maxlag* (X axis) and *maxerr* (Y axis). In all plots, a linear model is used, i.e.  $\theta_2 = 1$ .

Figure 4 confirms the finding highlighted in Figure 3 with lower values of the weighting factor  $\alpha$  resulting in lower parameter mismatch as can be seen by comparing the plots from the top and bottom row in Figure 4. The figure demonstrates accounting for timing offsets in the objective function makes the calibration more resilient to lag errors in input data. This finding is particularly important for catchments with a fast dynamic ( $\log(\theta_1^*) = 11.0$ ) and large lag errors ( $maxlag = 48$ ) as can be seen in Figure 4.a. The importance of timing offset assessment appears to be of lesser importance when the catchment dynamic is slow ( $\log(\theta_1^*) = 13.0$ ) as

can be seen by comparing Figure 4.e with 4.o where the parameter mismatch is of similar magnitude.

The results presented in Figure 4 are limited to the experiments where a linear model is used ( $\theta_2 = 1$ ). This constitutes a strong assumption that is not often supported by observed data in real catchments where strong non-linearity is usually the norm (Jakeman et al., 1993). In order to compare the results between linear and non-linear models, Figure 5 shows the median and interquartile range of the absolute difference in parameter value  $\log(\theta_1)$  for an average set of conditions where  $\log(\theta_1^*) = 12.2$ ,  $maxlag = 24$ ,  $maxerr = 0.2$ . The results are presented for three values of the weighting factor  $\alpha$  and two values of the storage exponent  $\theta_2$ .

This figure reveals that the non-linearity amplifies the mismatch in parameter  $\theta_1$  when timing offsets are not taken into account in the objective function. When  $\alpha = 1.0$ , we can see that the median difference goes from 0.13 when  $\theta_2 = 1$  to 1.52 when  $\theta_2 = 2$ . In addition the interquartile range displayed as a grey solid line in Figure 5, appears much larger for the non-linear model than for the linear one. As a result, the use of an objective function discarding timing offset is also affecting parameter uncertainty, which is likely to undermine the model performance when it is run with input data outside of the range observed during the calibration period (e.g. exceptional flood event).



**Figure 5.** : Median value of  $|\log(\theta_1^*) - \log(\theta_1)|$  for three values of  $\alpha$  and two values of  $\theta_2$ . Lag errors and amplitude errors are set to  $maxlag = 24$  and  $maxerr = 0.2$ .  $\theta_1^*$  is set to 12.2.

## 5 DISCUSSION AND CONCLUSION

This paper proposed a method to estimate simultaneously timing and amplitude offsets by comparing simulated and corresponding observed time series. The method was applied to a calibration problem where a model parameter is fitted to match observations using a least-squares criterion. Synthetic experiments were conducted by introducing random lag and amplitude errors in the model input and calibrating against a reference simulations considered as the observation.

The results show that the assessment of timing offset leads to calibration results that outperform the ones obtain with a classical approach, where only amplitude offsets are considered. The improvement were significant in the case of model dynamics with short response time and non-linear models.

This study was focused on optimisation criterion based on a least-squares approach. Further work needs to be undertaken to assess the impact of timing offset estimation when using other type of fitting criterion (e.g. using Box-Cox transformed variables as is often the case in hydrology).

## ACKNOWLEDGEMENT

The data from the Corang River was retrieved from the Hydrological Reference Stations (HRS) dataset, provided by the Bureau of Meteorology. Peter Thew and Ben Gouweleeuw are greatly acknowledged for their support and help.

## REFERENCES

- Bentura, P. L. and C. Michel (1997). Flood routing in a wide channel with a quadratic lag-and-route method. *Hydrol. Sciences Journal* 42(2), 169–189.
- Ewen, J. (2011). Hydrograph matching method for measuring model performance. *J. Hydrol.* 408(1), 178–187.
- Jakeman, A., G. Hornberger, et al. (1993). How much complexity is warranted in a rainfall-runoff model? *Water Resour. Res.* 29(8), 2637–2650.
- Kavetski, D., G. Kuczera, and S. W. Franks (2006). Bayesian analysis of input uncertainty in hydrological modeling: 2. application. *Water Resour. Res.* 42(3).
- Liu, Y., J. Brown, J. Demargne, and D.-J. Seo (2011). A wavelet-based approach to assessing timing errors in hydrologic predictions. *J. Hydrol.* 397(3), 210–224.