Kim, K. B., & Han, D. (2016). Exploration of sub-annual calibration schemes of hydrological models. Hydrology Research, [nh2016296]. DOI: 10.2166/nh.2016.296

Exploration of sub-annual calibration schemes of hydrological models

Kue Bum Kim[1*] and Dawei Han[1]

* Corresponding author: kbkim93@gmail.com

[1] Water and Environmental Management Research Centre, Department of Civil Engineering, University of Bristol, Bristol, UK

Abstract

This study has compared hydrological model performances under different sub-annual period calibration schemes using two conceptual models, IHACRES and HYMOD. In several publications regarding sub-annual period calibration, the authors showed that such an approach generally performed better than the conventional whole period method. Hence, there are advantages in dividing (or clustering) the data into sub-annual periods for calibration. However, little attention has been paid to the issue of how to calibrate the non-continuous sub-annual period. It is therefore important to explore reliable calibration schemes for such a situation. Unlike the conventional whole period calibration which assumes time-invariant parameters for the entire calibration period, the model parameters vary in sub-annual calibration. We have explored two sub-annual calibration schemes, serial calibration scheme (SCS) and parallel calibration scheme (PCS). We assume that the relationships between the rainfall and runoff could be different for each sub-annual period and consider intra-annual variations of the system. The models are then evaluated for a different validation period to avoid over-fitting (or, over parameterisation) and the optimal sub-annual calibration period is explored. Overall, we have found that PCS performed slightly better than SCS and the optimal calibration periods are seasonal and bimonthly for IHACRES and biannual for HYMOD at the study catchment. Since there are pros and cons in both SCS and PCS, we recommend choosing the method depending on the purpose of the model usage. Although the catchment is specific in the study, the methodology proposed is general and applicable to other catchments.

Keywords: calibration, sub-annual, hydrological model, model performance, optimal calibration period

INTRODUCTION

Hydrological modelling is an essential tool for understanding the hydrological behaviour of a catchment (Madsen, 2000; Wagener *et al.*, 2003), and it is a complicated task (De Vos *et al.*, 2010). The most common method for identifying the optimised model parameters is through the calibration with the use of historical observation data. Objective functions, such as Nash-Sutcliffe Efficiency (Nash and Sutcliffe, 1970) are used to minimise the difference between the observed and simulated flows. This calibrating scheme is widely applied (Sorooshian, 1991; Gan and Biftu, 1996; Gupta *et al.*, 1998; Gupta *et al.*, 2009). The validation is a standard practice in hydrological modelling (Andréassian *et al.*, 2009) to test the model with the data outside of calibration period to evaluate the model performance.

A recognised issue in hydrological modelling is uncertainties which are attributed to the model structural errors and parameterisation errors. The uncertainty due to model structure errors is generally quantified by using several different models, and numerous methods are proposed regarding quantification of the uncertainty of parameterisation problem. The time varying parameters which may arise from catchment change (such as land use/cover change), climate variability and climate change (such as change of evapotranspiration dynamics of vegetation due to higher or lower temperatures) may be another source of uncertainty. Such changes may be significant within a year (e.g., contrasting vegetation cover between winter and summer in many parts of the world). Recently, there have been some studies about the stability of the model performances and the effect of parameter values (Xu, 1999; Li *et al.*, 2014; Patel and Rahman, 2014; Yan and Zhang, 2014). The reasons of time varying model parameters can be explained by several reasons (Merz *et al.*, 2011). First, the hydrological model has structure errors and the calibrated parameters may change for different time periods in order to compensate these problems with the model structures (Wagener *et al.*, 2003). Secondly, catchment characteristic change (Brown *et al.*, 2005) such as land use and vegetation variations (Merz and Blöschl, 2009) can also lead to the change of calibrated parameters. However, the correlation between parameters is complicated (Wagener, 2007) which make it hard to understand the reason of the parameter changes in time (Wagener *et al.*, 2010).

Therefore, most parameterisation scheme in hydrological model is based on the assumption that the parameters do not change for the entire calibration period. There has been some research about the adequate data length for calibration (Xu and Vandewiele, 1994; Zeng et al., 2016). In addition, some researchers have attempted to develop more accurate models by adapting sub-annual calibration, which is based on seasonal or monthly time periods (i.e. non-continuous periods) in order to better simulate temporal variations of the varying catchment condition within the year. In such a scheme, intra-annual variation of the data is taken into account. For example, Luo et al. (2012) examined ten different parameterisation schemes at the catchments in Australia. Their results have shown that calibrating the model for each individual month separately makes the model performance better than the other schemes, which is particularly evident for the dry months when the flow is low and difficult to forecast. Levesque et al. (2008) conducted seasonal calibration, in which summer and winter seasons were calibrated separately. When the model was calibrated based on the summer (dry period) data, the model performance improved considerably. However, there was no advantage when only the winter (wet period) data were used for calibration compared with the conventional calibration method which used the entire data over the whole period. Paik et al (2005), Kim and Lee (2014), Zhang et al (2015) and Kim et al (2016) also used the seasonal calibration method.

Similar to the sub-annual discrete calibration, Hartmann and Bardossy (2005) investigated the transferability of hydrological models by dividing the observation period into different climatic conditions (i.e. warm, cold, wet and dry). They conducted the calibration only for the chosen years which were discontinuous, and the model was running continuously for the entire observation period. De Vos et al (2010) proposed clustering time series according to hydrological similarities and allowing the parameters to vary over the clusters during calibration. Seiller et al. (2012) selected five non-continuous hydrologic years for four contrasting climate conditions: dry/warm, dry/cold, humid/warm and humid/cold. Calibration was done on each period and validation was conducted on contrasting climate conditions. The model was kept running continuously on the entire time series while the optimisation was done only for the chosen years. The above mentioned studies demonstrate that model parameter values can vary over time in accordance with seasonal variations and indicate that there are some advantages in using this scheme, although further investigation and improvements are needed.

In this study, regarding sub-annual period calibration, we focus on two issues that should be resolved, which have not been considered yet in the literature. First, guidance on calibration scheme for non-continuous time periods should be provided. Most studies have conducted calibrations based on only the chosen sub-annual period (i.e. only the selected sub-annual period data are used for parameter optimisation in the objective function) while the entire time series data were used to run the model continuously. However, the downside of such an approach is that there are discontinuities in the time series of soil moisture when the separately calibrated flows from individual sub-annual periods are combined to examine the performance of the entire time period. Another calibration scheme is to optimise the entire varying parameters simultaneously. De Vos et al (2010) divided the data into 12 clusters and the model parameters were set differently for each of the 12 clusters which resulted in the number of degrees of freedom equal to 12 clusters times the number of parameters. Then these parameters are optimised simultaneously. These different methodologies point towards the need for a guidance on the non-continuous sub-period calibration scheme. Second, most non-continuous sub-period calibration studies are not interested in the model performance of the total time series but that of the particular time periods (e.g. dry season or wet season). Therefore the questions are, "How can we assess the model efficiency of the entire time series using non-continuous sub-annual period calibration? Is it reasonable to just combine the flows calibrated separately?", "In this case, what about the issue of discontinuity in soil moisture simulation?" Third, different sub-annual calibration periods may present different model efficiencies due to underfitting or overfitting issues and there may be an optimal period which can be evaluated by using cross validation. So far, there is no consensus method in the literature regarding how to calibrate non-continuous sub-annual time period. Given such a background, this paper explores the following questions:

(1) What schemes are applicable for non-continuous sub-annual calibration?

(2) Which schemes are both logical and practical?

(3) What are the pros and cons of these schemes?

(4) What is the optimal time period for sub-annual calibration?

In this study, two sub-annual calibration schemes are employed to explore these questions, i.e., serial calibration and parallel calibration schemes. The sub-annual calibration has been performed on five different time scales: annual, biannual, seasonal, bimonthly and monthly. We assume that the system is changing due to various reasons such as vegetation change (e.g., seasonal change of deciduous vegetation and crop rotation), soil structure change (e.g., seasonal soil compaction by farm animals or farming machines), etc. Therefore the system response, i.e. the relationships between the rainfall and runoff may be different for different time scales and there may be an optimal time period for sub-annual calibration. In other words, different sets of varying parameters are optimised to consider intra-annual variations. As the number of sub-annual calibration period increases (i.e. from annual to monthly) the model is more likely to fit to the observations since it has more flexibility to cope with the change of the system. On the other hand it is more likely to fit to the noise which is the well-known trade-off between the bias and variance in mathematical modelling. Therefore the models are evaluated for different validation periods to overcome the overfitting (or. over-parameterisation) issue and to explore the best sub-annual model.

Both calibration schemes are applied to one catchment located in the southwest of England. Since the main aim of this study is to introduce the concept and the logic of non-continuous sub-period calibration scheme, we believe only one catchment is sufficient to prove the concept and it is hoped that a wide application of the proposed methodology in other catchments under different conditions will help the hydrological community to find useful patterns on this important issue.

This paper is organised as follows. The study area and hydrological models used are presented in Section 2. The methodology used to examine the calibration schemes are described in Section 3. The results are presented in Section 4 followed by the discussion and conclusion in Section 5 and Section 6.

CASE STUDY AREA AND THE HYDROLOGICAL MODELS

Study area and data

The Thorverton catchment is used in this study, which has an area of 606km^2, and is a sub-catchment of the Exe catchments. The Exe catchment is located in the southwest of England with an area of 1,530 km^2 and an average annual rainfall of 1,088 mm. Figure 1 shows the overview of the Thorverton catchment area. Daily time series of the observed precipitation, potential evapotranspiration and flow data (1961-1990) over the Thorverton catchment is obtained from the UK Met Office and daily temperature data has been downloaded from the UKCP09 gridded observation data sets.

Figure 1. Location of the Thorverton catchment (the thick highlighted line) in the UK.

Hydrological model

IHACRES

The first model used in this paper is a conceptual rainfall-runoff model IHACRES (Jakeman and Hornberger, 1993) which has eight parameters. This model has been widely applied to a variety of catchments for hydrological analysis and climate impact studies (Jakeman *et al.*, 1993; Littlewood, 1999; Letcher *et al.*, 2001; Kim and Lee, 2014). The model is composed of a non-linear module and a linear module as shown in Figure 2 and model parameters are listed in Table 1. A non-linear module converts rainfall to effective rainfall which is calculated from the following equations.

$$U_k = [C(\emptyset_k - l)]^p r_k \qquad (1)$$

where, r_k is the observed rainfall, C is the mass balance, l is the soil moisture index threshold and p is the power on soil moisture respectively. The soil moisture ($\emptyset_k$) is calculated from:

$$\emptyset_k = r_k + (1 - \frac{1}{\tau_k})\emptyset_{k-1} \qquad (2)$$

where, τ_k is the drying rate given by:

$$\tau_k = \tau_w \exp[0.062 f(t_r - t_k)] \qquad (3)$$

where, τ_w is the drying rate at reference temperature, f is the temperature modulation, t_r is the reference temperature, and t_k is the observed temperature. A linear module assumes that there is a linear relationship between the effective rainfall and flow. Two components in this module, quick flow and slow flow, can be connected in parallel or in series. In this study two parallel storages in the linear module is used because it reflects the catchment conditions and the streamflow (x_k) at time step k is defined by the following equations:

$$x_k = x_k^{(q)} + x_k^{(s)} \tag{4}$$

$$x_k^{(q)} = \beta_q U_k - \alpha_q x_{k-1}^{(q)} \tag{5}$$

$$x_k^{(s)} = \beta_s U_k - \alpha_s x_{k-1}^{(s)} \tag{6}$$

where, $x_k^{(q)}$ and $x_k^{(s)}$ are quick flow and slow flow respectively, and α and β are recession rate and peak response respectively. The relative volumes of quick flow and slow flow can be calculated from:

$$V_q = 1 - V_s = \frac{\beta_q}{1+\alpha_q} = 1 - \frac{\beta_s}{1+\alpha_s} \tag{7}$$

Therefore, one parameter is determined if the other three parameters are known among the four linear module parameters (α_q, α_s, β_q, β_s).

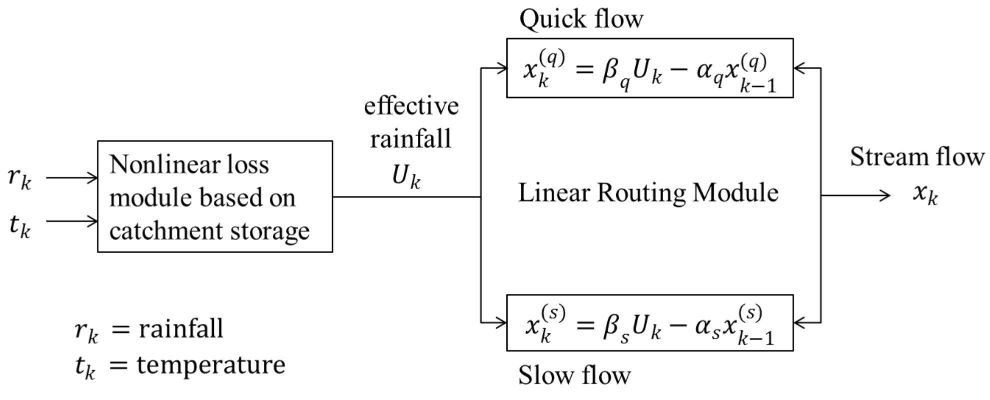

Figure 2. Structure of the IHACRES model.

Table 1. Parameters in the IHACRES model

Module	Parameter	Description	Range
Non-linear	c	Mass balance	0-0.04
	τ_w	Reference drying rate	0-50
	f	Temperature modulation of drying rate	0-4
	l	Soil moisture index threshold	0-50
	p	Power on soil moisture	0-3
Linear	α_q, α_s,	Quick and slow flow recession rate	-1-0
	β_q, β_s	Fractions of effective rainfall for peak response	-1-0

HYMOD

Another conceptual rainfall-runoff model used in this study is HYMOD which has five parameters. The model consists of a simple rainfall excess model based on the probability distributed principle (Moore, 1985) and was applied by several recent studies (Boyle, 2001; Wagener *et al.*, 2001; Vrugt *et al.*, 2003; De Vos *et al.*, 2010). The model parameters are described in Table 2 and the model structure is illustrated in Figure 3. The cumulative distribution function of the water storage capacity *C* is in the following form.

$$F(C) = 1 - (1 - \frac{C(t)}{C_{max}})^{b_{exp}}, \qquad 0 \leq C(t) \leq C_{max} \tag{8}$$

where, C_{max} is the maximum soil moisture storage capacity in the catchment and b_{exp} controls the degree of spatial variability of the soil moisture capacity. The excess rainfall is treated as the runoff which is divided into quick flow and slow flow based on the partitioning factor *a*. The runoffs are routed through three identical quick flow tanks and a parallel slow flow tank. The flow rates are determined by the recession coefficient for quick flow tank (R_q) and slow flow tank (R_s).

Table 2. HYMOD model parameters.

Parameter	Unit	Range	Description
C_{max}	mm	1-500	Maximum soil moisture storage capacity
b_{exp}	-	0.01-1.99	Spatial variability of soil moisture capacity
α	-	0.01-0.99	Quick/slow flow distribution factor
R_s	day	0.01-0.99	recession coefficient for slow flow tank
R_q	day	0.01-0.99	recession coefficient for quick flow tank

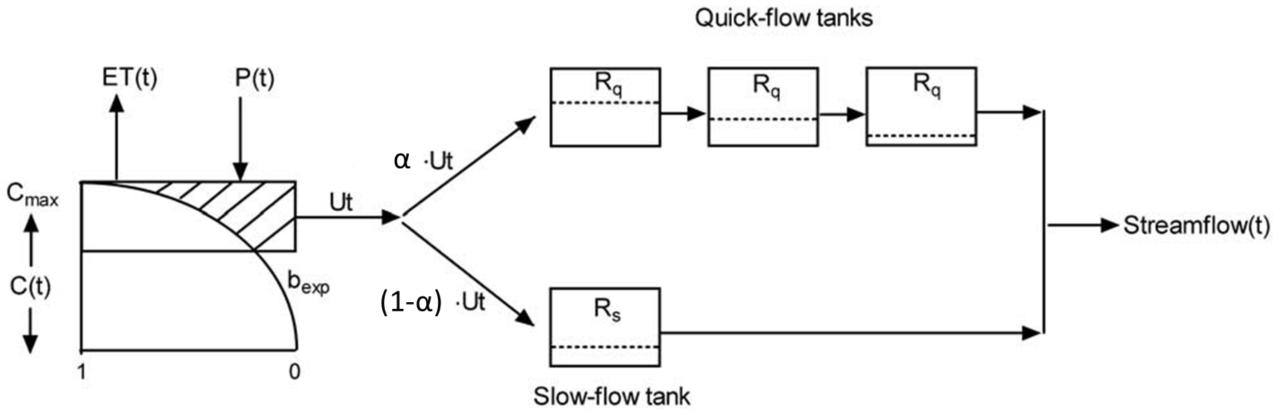

Figure 3. Structure of the HYMOD model (adopted from Vrugt et al.[2002]).

METHODOLOGY

Optimisation method

For the optimisation algorithm, we used dynamically dimensioned search (DDS) (Tolson and Shoemaker, 2007) which is a simple, single objective, heuristic global search algorithm. DDS searches the parameter space globally and incrementally localises as the number of iterations reaches the maximum allowable number of simulations. The procedure from global to local scales is done by probabilistically decreasing the number of model parameters in the neighbourhood. New search avoids poor local optima and parameter values are updated by perturbing the current solution values in the randomly selected dimensions with perturbation magnitudes randomly sampled from a normal distribution with a zero mean. More detail can be found in Tolson and Shoemaker (2007).

Model parameterisation schemes

The calibration has been conducted with the use of root mean square error (RMSE) to minimise the difference between the observed and simulated flow and the optimised parameters have been tested in the validation period.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(Q_{obs,i} - Q_{sim,i}\right)^2}{N}} \quad (9)$$

where, Q_{sim} and Q_{obs} are the simulated and observed runoff, respectively. i is the i^{th} day, and N is the number of days in the calibration period. In this study, only RMSE is used as the objective function. Exploration of the effect of different objective functions and their combinations in calibrating rainfall-runoff models can be another research topic to extend the current study (e.g. Jie et al, 2016).

We conducted three different calibration schemes, i.e., the conventional whole period calibration scheme and dynamic sub-annual calibration scheme. The common assumption of the conventional calibration scheme is to use time invariant parameters. The model parameters do not change during the entire calibration period. On the other hand, the dynamic calibration schemes are based on the idea that the optimised parameter values can vary according to different climatic and catchment conditions. In this study we have assumed sub-annually (annual, biannual, seasonal, bimonthly and monthly) varying parameters. With regard to a monthly model, for instance, the model parameters are the same for each month for all years but differ between the months. This scheme can be conducted in two different ways: serial calibration scheme (SCS) and parallel calibration scheme (PCS). Among 5 different time scales for sub-annual calibration, illustrations of the monthly calibration schemes are presented in Figure 4. SCS optimises the whole parameter sets simultaneously, i.e., 12 monthly models are calibrated at the same time. For example, 96 parameters (8 parameters times 12 sets) in IHACRES and 60 parameters (5 parameters times 12 sets) in HYMOD are optimised simultaneously. On the other hand, PCS optimises 12 different monthly models separately. Only the data from individual months are used in the objective function while the model is running for the whole period. The calibrated flow data from each separate monthly model are then combined together to compare with the observed flow.

Calibration has been done for three time periods 1960s, 1970s and 1980s. Normally a warm-up period (e.g. one or two years) is used during calibration to reduce the influence of the initial values of state variables. In this study, instead of using a warm-up period, we set the initial value of soil moisture as a parameter to be optimised to avoid the warm-up problem (this is fine for calibration, but not suitable for validation in which a warm-up period of one year is still needed because the initial value of soil moisture is not a model parameter).

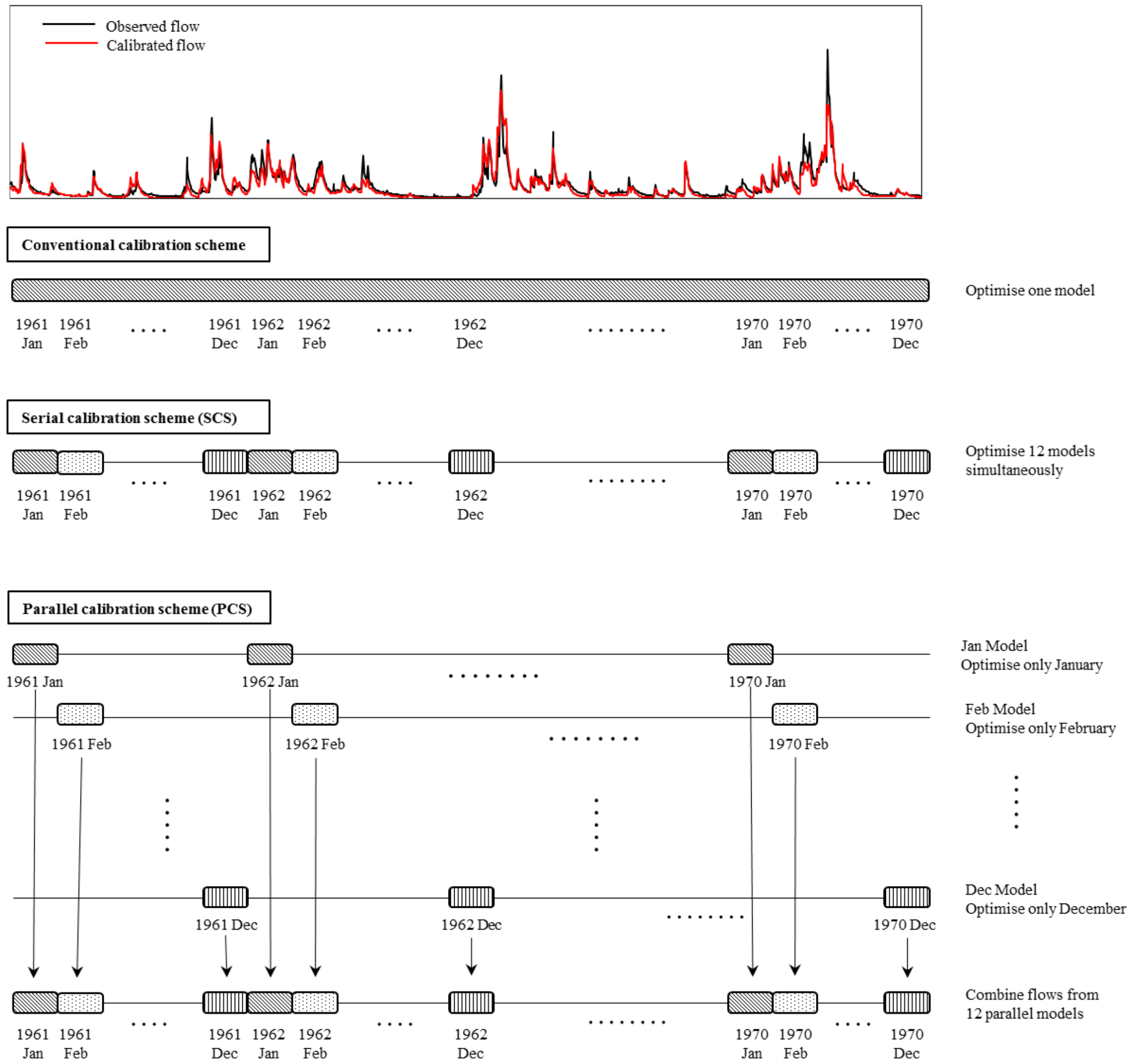

Figure 4. Illustration of the conventional calibration, SCS and PCS for the calibration period from 1961 to 1970.

Evaluation of the parameterisation schemes and the optimal calibration period

To explore the optimal time period (i.e. optimal number of groups) of sub-annual calibration which is based on a balance between the bias and the variance (Figure 5), the model performance is evaluated in the validation period with the use of root mean square error (RMSE). If the sub-annual calibration period is too long it may not be flexible to capture the change of the system and may lose the temporal information due to underfitting (high bias and low variance). On the other hand if the period is too short even the noise of the system will be matched due to overfitting (low bias and high variance). Hence, it is possible that there could be an optimal sub-annual calibration period. So far there are no reported studies on this issue.

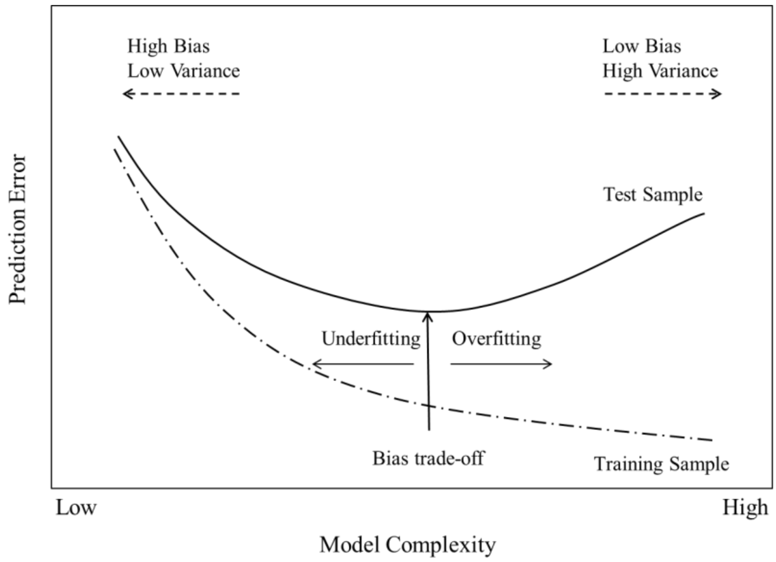

Figure 5. Trade-off between the bias and the variance to explain the model over-fitting and under-fitting (Han, 2011).

The three-fold cross validation is applied to evaluate the performance of the model and to find the optimal sub-annual calibration period (Figure 6). The RMSE values are averaged for all the six validation periods.

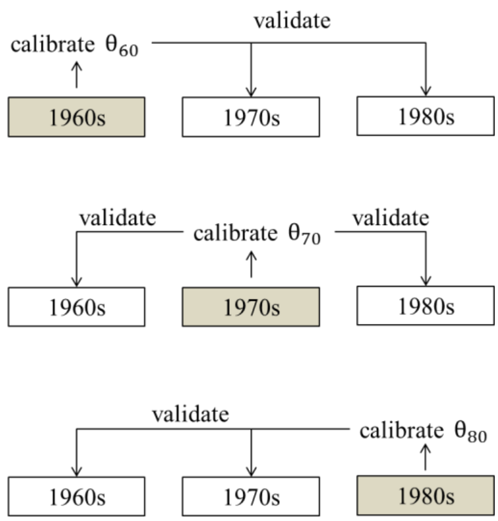

Figure 6. Three fold cross-validation scheme (θ is optimised parameter set).

Sub-annually optimised parameters from SCS and PCS can be applied in two ways for validation: serial run and parallel run. For a monthly model, for instance, the serial run is to apply the monthly varying parameters in series at one model run, while the parallel run is to apply 12 sets of parameters individually (i.e. run from January parameters to December parameters separately). Therefore, the parameters from SCS can be applied in series and the parameters from PCS can be applied in parallel. The parameters optimised from SCS and PCS are expressed as follows.

$$\theta_y^{c,m}, \qquad y = \begin{cases} 60, & \text{for 1960s} \\ 70, & \text{for 1970s} \\ 80, & \text{for 1980s} \end{cases}, \qquad c = \begin{cases} S, & \text{for SCS} \\ P, & \text{for PCS} \end{cases}, \qquad m = \text{number of groups} \tag{10}$$

where, *y* is the calibration period, *c* is the type of the calibration scheme and *m* is the number of groups (e.g. 12 for monthly calibration).

For example, 1960s monthly parameter sets from SCS ($\theta_{60}^{s,m}$) are applied to the validation period in which the monthly parameters are switched every month while the model runs continuously ((1) in Figure 7). One hydrograph is produced in this way. On the other hand, applying monthly parameter sets from PCS ($\theta_{60}^{p,m}$) in parallel runs result in 12 hydrographs. Each monthly parameter set is applied separately while the model is run for the whole validation period ((2) in Figure 7). Note that in this figure, the hydrographs are represented only for the corresponding month for illustration purpose, although the hydrographs are continuous in reality. The flow data from each month period is picked up and combined to build the entire flow data for comparison with the observed flow.

As previously noted the initial value of the soil moisture is set as if it is a ‘model parameter’ and is optimised during calibration. However, in reality, the soil moisture is not a hydrological model parameter but a state variable, hence, the optimised initial soil moisture value estimated from the calibration period should not be used in the validation period. If we know the correlation between the soil moisture and the flow, the initial soil moisture value in the validation period could be estimated from the initial flow value. However, since this is not the main point of the study, we just assume that the initial soil moisture value evolves to an appropriate value after one year instead of building the soil moisture and flow relationship. Therefore, the first one year of the validation period has not been taken into account in evaluating the model efficiency.

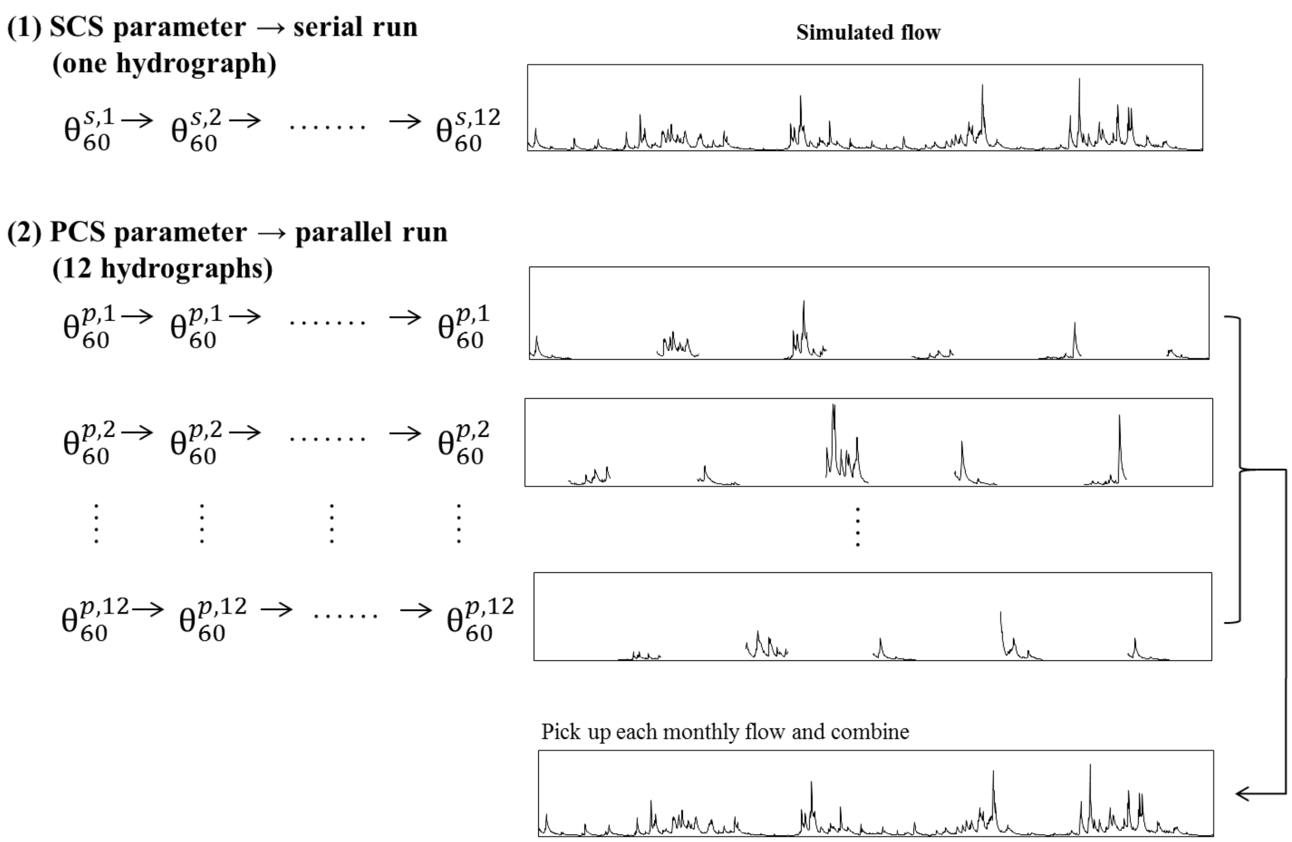

Figure 7. Illustration of the evaluation of 1960s monthly calibrated parameters. The flows are simulated by applying the parameters to 1970s (or 1980s) rainfall.

RESULTS OF MODEL EVALUATION

Figure 8 presents the models performance for SCS and PCS. The calibration and validation results show the typical trend of bias trade-off (Figure 5). The more calibration groups we divide the data into the more flexible the model becomes and the smaller the error becomes for the calibration scheme. However, the improvement in the calibration period comes with the cost of a worsened validation results due to overfitting (or over-parameterisation). For the study catchment, the optimal sub-annual calibration periods, where the RMSE of the validation result is the smallest, are bimonthly and seasonal for SCS and PCS respectively for the IHACRES. For HYMOD, biannual is the optimal period for both SCS and PCS. The implication of this result is that it is reasonable to assume that the catchment response is gradually changing within the year and to capture the change of the system intra-annual variation should be considered for calibrating a hydrological model.

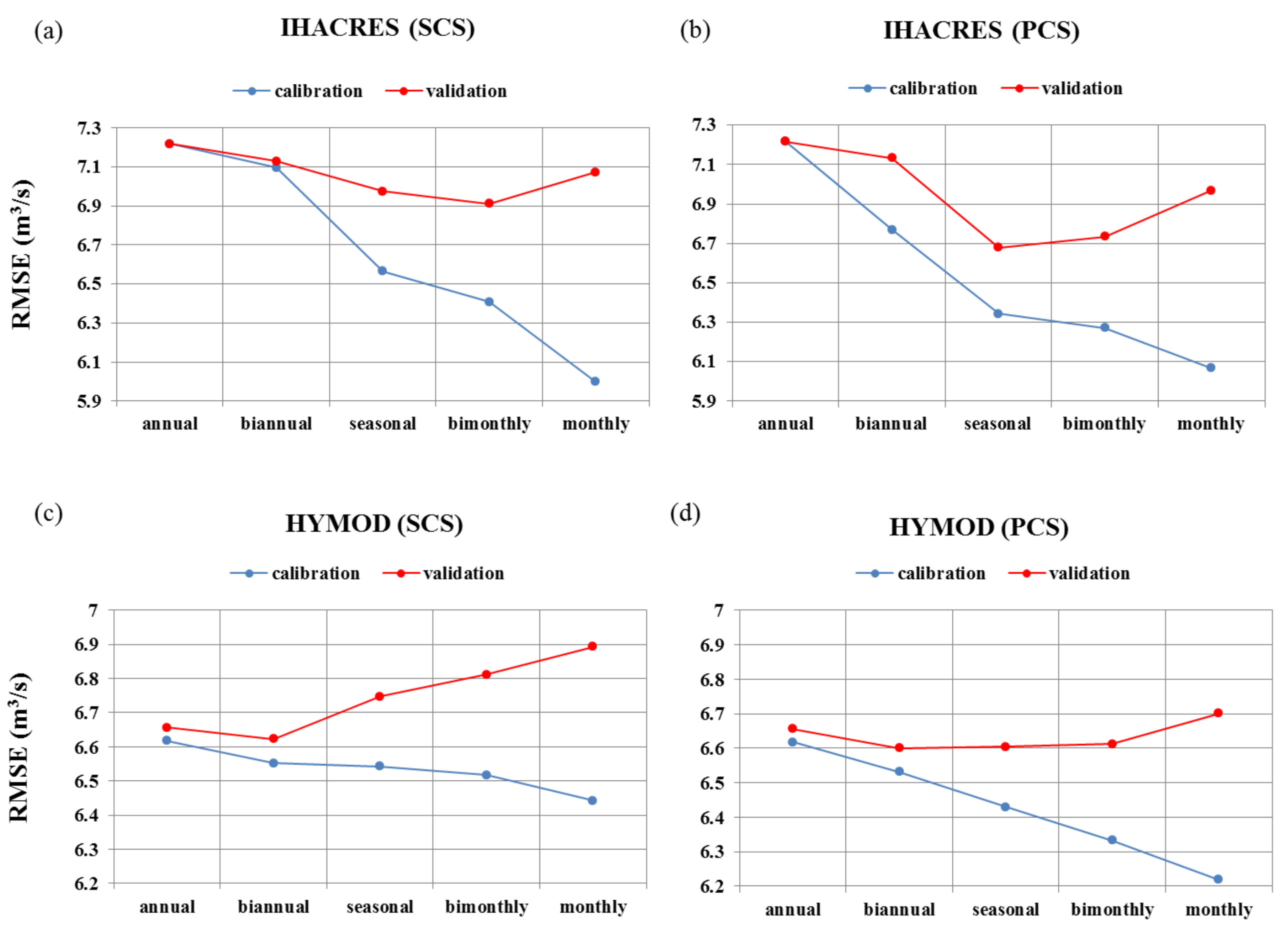

Figure 8. Calibration and validation results for SCS and PCS for IHACRES and HYMOD.

Figure 9 compares the SCS and PCS performance of the validation results. It is reasonable to suppose that the SCS will be better than the PCS since the SCS has a sound logic by continuously running the model like the real system from the first to the last in series, while the PCS considers only the chosen time period for estimating the model efficiency although each sub-annual model runs continuously. However, both IHACRES and HYMOD show that the PCS is better than SCS. This might be because the parameters in PCS are adjusted to provide the best fit to one specific period without being affected by other periods, where the optimal parameter values are different, while the SCS has an issue of optimisation in high dimensional spaces and the interdependency of numerous parameters.

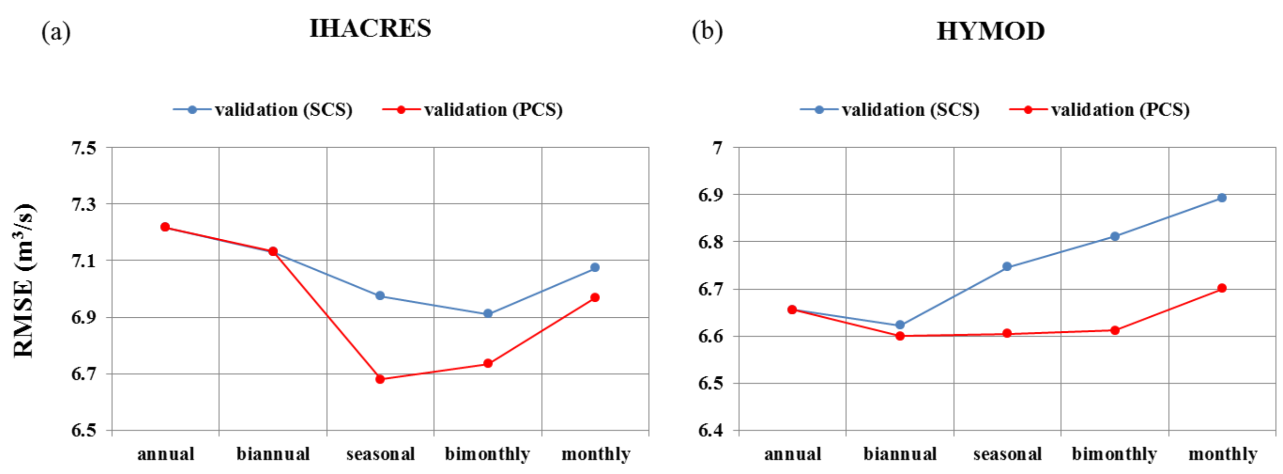

Figure 9. Comparison of SCS and PCS performance for IHACRES and HYMOD.

DISCUSSION

Sub-annual calibration and over-parameterisation

There is a possible concern that the sub-annual calibration scheme has an over-parameterisation (over-fitting) problem. For example, the monthly calibration scheme in IHACRES has 96 parameters (8 parameters times 12 sets), which could be recognised as 12 monthly models of 8 parameters equivalent to one model of 96 parameters. However, this argument is not exactly right due to misinterpretation of multiplying the model parameter numbers by the number of models. Figure 10 explains why the two concepts are not equivalent. Suppose we have a linear model which is calibrated on monthly basis. Each monthly model has one parameter (a_1, a_2 … a_{12}). However, when these 12 separately calibrated models are combined (Figure 10(a)) to evaluate the annual performance, it is apparent that this combined model is different to a single model with 12 parameters (Figure 10(b)) since the model structure between the two is different. Therefore the two models, a combined model and one model with 12 parameters are not equivalent and should not be confused among them.

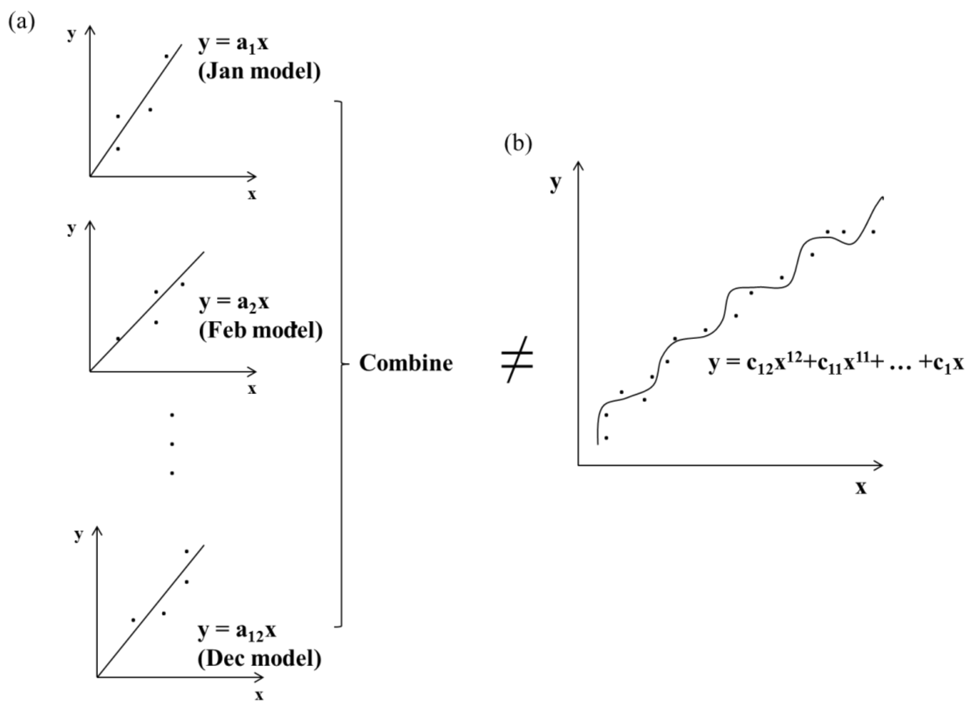

Figure 10. Schematic of 12 monthly models of one parameter and one model of 12 parameters.

We have done an experiment with a synthetic flow to prove that these two models are not equivalent. It is assumed that the system does not change in order to simplify the problem and SCS is applied for five different time scales: annual, biannual, seasonal, bimonthly and monthly. To make a stationary system, we assumed the 1960s parameters calibrated in the conventional method as a 'true parameter set' which are used to generate the 'true flow'. A random error (noise) is added to the 'true flow' to make it as a real flow data (Eq(11)).

$$Q_r = Q_t + Q_t \times \alpha_r \times \mu \tag{11}$$

where, Q_r is the random error added flow, Q_t is the 'true flow', α_r is a random error intensity coefficient, which is 0.2 in this study and μ is a random sample from a Gaussian distribution. This flow series generated from a stationary system is then used to calibrate the hydrological model, IHACRES.

The hypothesis is that a simple model (*m* parameters) which is separately calibrated on a sub-annual basis (*n* groups) is equivalent to a complicated single model with $m \times n$ parameters (Eq(12)).

$$\mathrm{y} = C_n x^n + C_{n-1} x^{n-1} + \cdots + C_1 x \qquad n = \begin{cases} 1 & \text{for annual calibration} \\ 2 & \text{for biannual calibration} \\ 4 & \text{for seasonal calibration} \\ 6 & \text{for bimonthly calibration} \\ 12 & \text{for monthly calibration} \end{cases} \tag{12}$$

The hypothesis will be rejected if the errors are similar even though the number of sub-annual calibration groups (n) increase. This is because given that the hypothesis is true, the model becomes more complicated (i.e. one model with $m \times n$ parameters) when the number of groups n increases, which results in less error due to overfitting (Figure 11). The calibration has been performed on annual, biannual, seasonal, bimonthly and monthly time scale based on SCS. The model performances are presented in Table 3. The RMSE values are similar among different time scales which mean that the sub-annual calibration scheme does not improve the model performance. Therefore the hypothesis should be rejected. The reason of no improvements in the complicated model is that the model structures are still the same although the sub-annual calibration has been performed in different time scales. Hence, the sub-annually (n groups) calibrated model (m parameters) is not equivalent to a single model with $m \times n$ parameters.

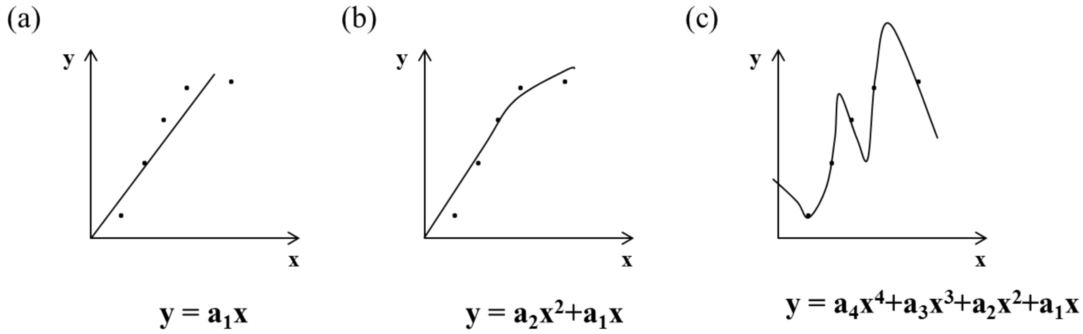

Figure 11. Schematic of overfitting issue when the number of groups increases. (a) a model with one parameter; (b) a model with two parameters; (c) a model with 4 parameters.

Table 3. RMSE of different calibration period.

Calibration period	annual	biannual	seasonal	bimonthly	Monthly
RMSE (m^3/s)	4.24	4.23	4.31	4.40	4.33

Another possible concern is that a time-variant parameter is not a parameter any more but a state variable. This shouldn't be a problem because such a time-variant parameter scheme has already been widely adopted in the "adaptive control" (Tao, 2003) which is used to control the system with varying parameters.

Calibration of nonstationary system and the optimal calibration period

As mentioned in the previous section, if the system is stationary there is no advantage in applying the sub-annual calibration scheme. However, in real life the system is nonstationary and it is changing with time. The response of the nonstationary system may be different dependent on the catchment change (e.g. seasonal vegetation change, and seasonal soil structure change, etc.) which means that different hydrological model parameters may be needed. In this changing system, the more calibration groups divided, the less the error will be since the model has more flexibility to fit to the observation. Since there is a trade-off between bias and variance, there is an optimal calibration period between annual and monthly models as shown in Figure 8. It is understandable that IHACRES (8 parameters) has shorter optimised calibration period than HYMOD (5 parameters) since IHACRES has more parameters than HYMOD, which can cope with the change of the system better.

Comparison of soil moisture

As stated in Section 1, the sub-annual parameterisation scheme mostly applied in hydrological model calibration is PCS. The model is kept running continuously on the entire time series while the optimisation is done only for the chosen sub-annual period (e.g. monthly or seasonal). However, this method has a drawback. Figure 12 presents a part of the time series of soil moisture from the monthly model for SCS and PCS. It is apparent that in the PCS the time series shows discontinuity at each monthly boundary, which is illogical and a downside of this method although the flow simulation performance is excellent as presented in the previous sections. On the other hand, in the SCS the time series is continuous, which is a realistic simulation. This is because in the serial calibration, the model runs continuously, while in the parallel calibration, although each monthly model runs continuously only the corresponding month data are picked up and combined together.

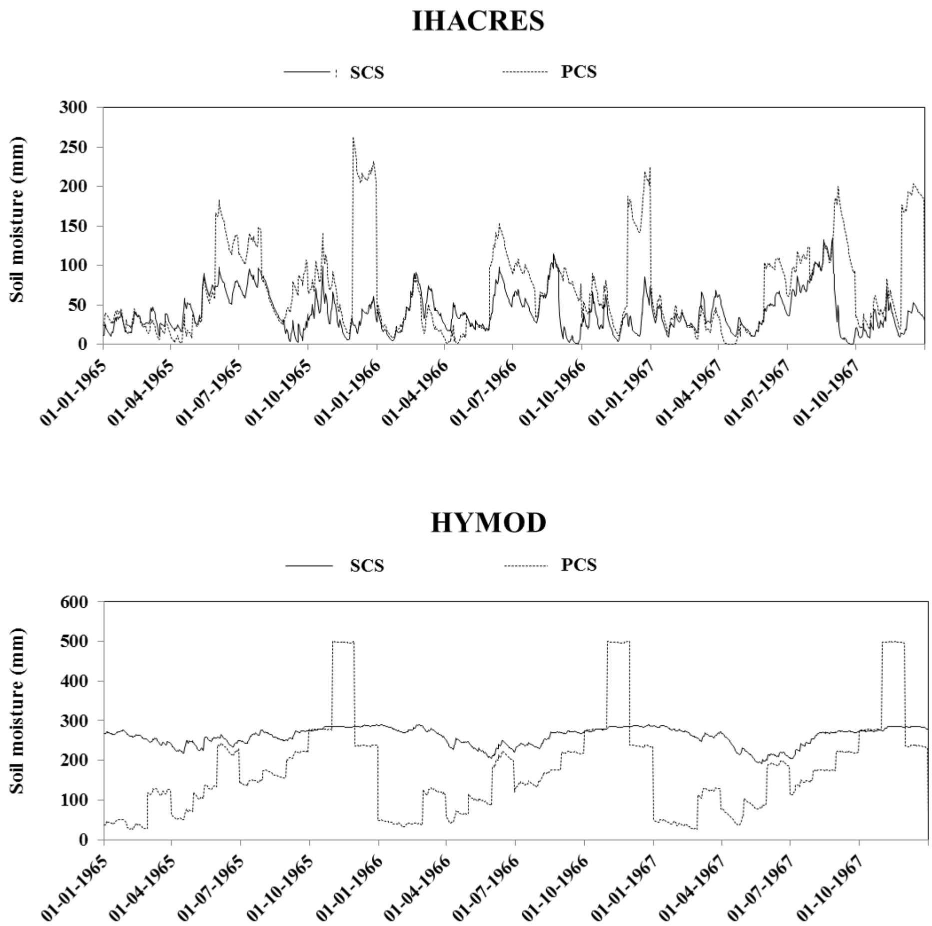

Figure 12. Time series of soil moisture for different calibration schemes.

To see whether the time varying parameter is reasonable between different time periods of calibration, we have investigated the state variable, soil moisture for IHACRES and HYMOD. In Figure 13, we can see that among sub-annual calibration schemes, the patterns of soil moisture of monthly and seasonal calibrations have been ruined compared with the conventional annual calibration method (low in summer, high in winter), while the biannual calibration result shows a similar pattern. From flow point of view, the optimal calibration period for SCS is the bimonthly model (Fig 9 (a)) while, from soil moisture point of view, at least the biannual model should be used although the RMSE between the flows is not the least in this case. If the calibration period is divided more in detail than the biannual period, the pattern of soil moisture would not be realistic.

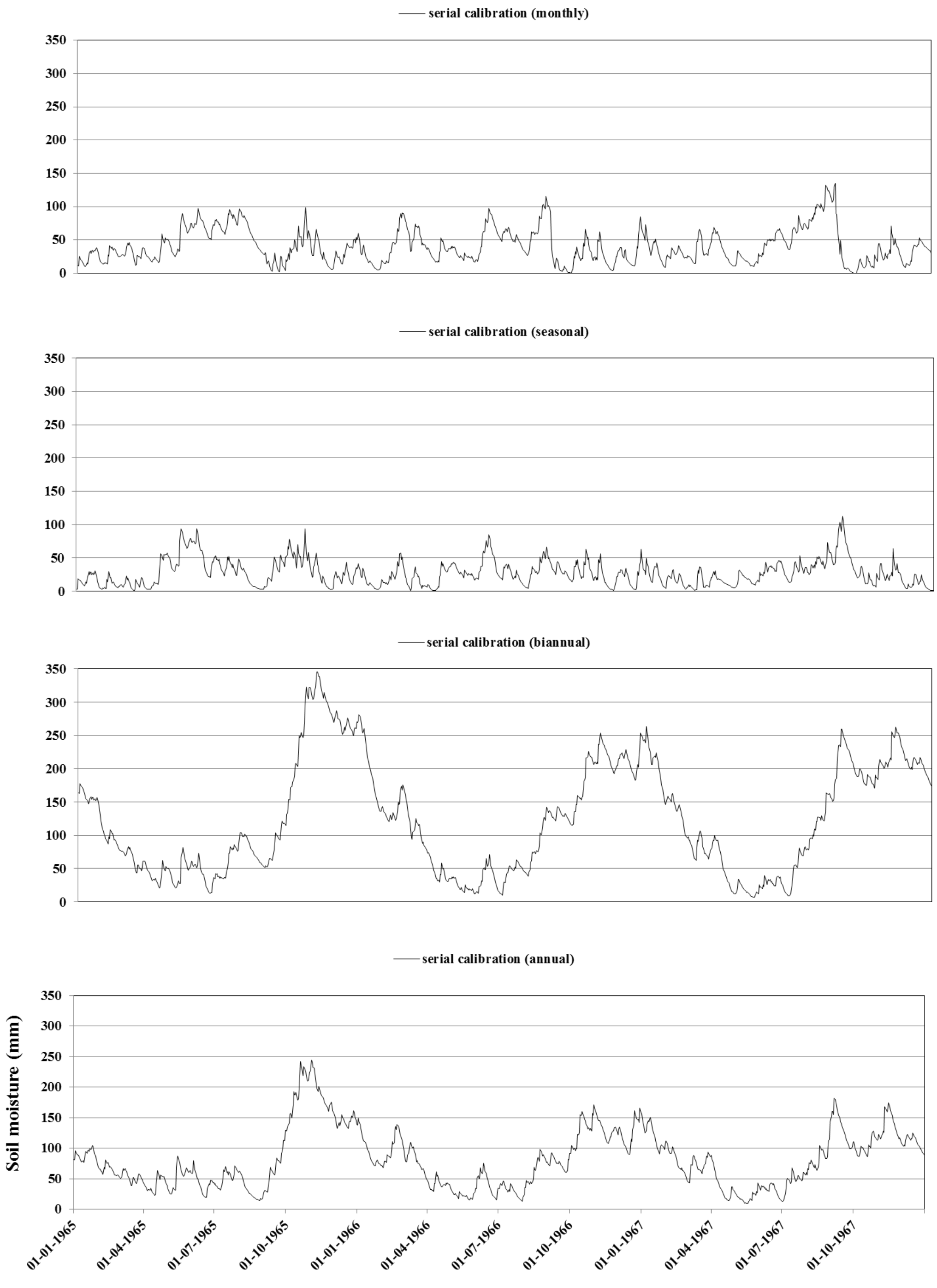

Figure 13. Time series of soil moisture from 1965-1967 in different calibration periods for IHACRES.

In contrast to the IHACRES, the HYMOD generates more reasonable soil moisture patterns (Figure 14). Although the sub-annual calibration period is divided in more detail, the pattern of soil moisture is maintained. Therefore, the biannual calibration period which is the optimal for SCS (Fig 9 (b)) is valid for HYMOD.

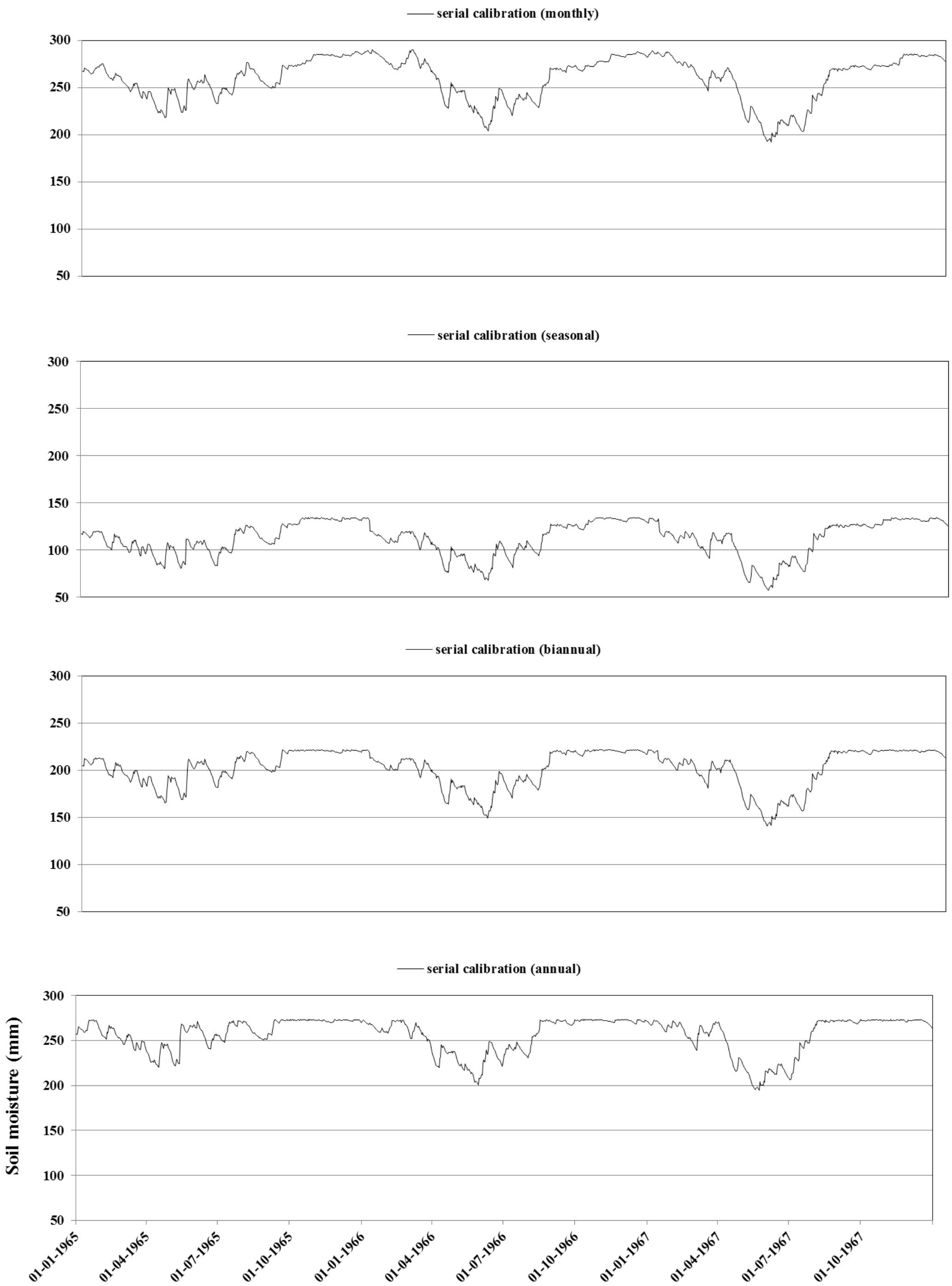

Figure 14. Time series of soil moisture from 1965-1967 for different calibration period for HYMOD.

From this result, it is found that realistic flow and soil moisture simulations may not be both achieved at the optimal state. Sometimes we may get the right answers (predicting the best flow for practical purposes) for the wrong reasons (unrealistic soil moisture). That is, the flow may be optimum at the expense of soil moisture due to the drawback of the hydrological model itself. There have been relevant researches regarding this issue by Zhuo and Han (2016a and 2016b).

Which calibration scheme is more reliable?

We have explored three options of the calibration schemes for hydrological modellers to choose from: 1) the conventional method considering inter-annual variations of the data, 2) SCS and 3) PCS considering intra-annual variations of the data. Logically, the SCS should be better than the PCS since the model should run in series by logic but the results show that the PCS beats the SCS. A possible interpretation of this result may be in part due to the issue of optimisation in a high dimensional space and the interdependency of numerous parameters in SCS.

Then the question is which method is more reliable? Our recommendation depends on the purpose of the calibration. If one is interested in the volume of the flow water, we recommend the PCS since it represents the best flow simulation result and is easy to calibrate due to the small number of parameters although there are discontinuities in the time series of soil moisture. On the other hand, when one is interested in soil moisture as well as flow, we recommend the SCS which is more logical and the performance is not much worse in flow but shows continuity in soil moisture simulation. However, the downside of this method is that it is not efficient and takes longer time to calibrate than PCS since the number of degrees of freedom equals to the number of sub-annual period (e.g. 12 for monthly calibration) times the number of parameters.

CONCLUSIONS

This study has compared the hydrological model performances under different sub-annual period calibration schemes using two conceptual models, IHACRES and HYMOD. There are some studies regarding sub-annual period calibration methods, but little attention has been paid to the issue of how to calibrate (e.g. in series or in parallel) the non-continuous time series and what is an optimal time period for sub-annual calibration. It is therefore important to investigate reliable calibration schemes for non-continuous sub-annual period calibration. In this study, we have proposed two alternative approaches to calibrate hydrological models by sub-annual calibration schemes in which unique model parameter sets are estimated for each sub-period (annual, biannual, seasonal, bimonthly and monthly). In one approach unique parameter sets for each sub-period is calibrated simultaneously and parameter values are thus changes for each sub-period - the serial calibration scheme (SCS). In the second approach the model is calibrated *n* (the number of groups, e.g. 4 for seasonal) times but only using data from one sub-period at the time. The *n* models are then combined to get the result of a complete time series - the parallel calibration scheme (PCS). Then three fold cross validation is used to find the optimal sub-annual calibration period and it is possible that this optimal calibration period is related to local catchment change and the purpose of the data usage. Overall, we have found that the model calibrated on the sub-annual period schemes generally perform better than the model calibrated in the conventional way, which implies that it is worth considering intra-annual variations in calibrating hydrological models in a changing environment. For the study catchment, from the flow point of view (predicting flow only for practical purposes), the optimal calibration periods for IHACRES are bimonthly and seasonal for SCS and PCS respectively. For HYMOD, biannual is the best sub-annual model for both SCS and PCS. However, from the soil moisture point of view, dividing sub-annual calibration period sometimes may not produce realistic a soil moisture pattern, which indicates the improvement of hydrological model structure is needed to achieve both the simulated flow and soil moisture at the optimal state. Therefore, not only the simulated flow but also the soil moisture needs to be considered to find the optimal calibration period. Among the dynamic calibration schemes, PCS performed slightly better than SCS. Since there are pros and cons in both SCS and PCS, we recommend choosing the method depending on the purpose of the sub-period calibration. Although the study catchment is specific in southwest England, the methodology proposed in this study is generic and applicable to other catchments. Since only one catchment is explored in this investigation, it is clear that such a study has not completely solved this problem and we hope this paper will stimulate the hydrological community to explore a variety of sites with different hydrological models so that valuable experiences and knowledge could be gained to improve our understanding on such a complex model calibration issue.

Acknowledgement

The first author is grateful for the financial support from the Government of Republic of Korea for carrying out his PhD study in the University of Bristol. The data used in this study are available upon request from the corresponding author via email (kk12496@bristol.ac.uk).

References

Andréassian V, Perrin C, Berthet L, Le Moine N, Lerat J, Loumagne C, Oudin L, Mathevet T, Ramos M, Valéry A. 2009. Crash tests for a standardized evaluation of hydrological models. Hydrology and Earth System Sciences Discussions: p. 1757-p. 1764.

Boyle DP. 2001. Multicriteria calibration of hydrologic models.

Brown AE, Zhang L, McMahon TA, Western AW, Vertessy RA. 2005. A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation. Journal of hydrology, **310**: 28-61.

De Vos N, Rientjes T, Gupta H. 2010. Diagnostic evaluation of conceptual rainfall–runoff models using temporal clustering. Hydrol Process, **24**: 2840-2850.

Gan TY, Biftu GF. 1996. Automatic calibration of conceptual rainfall-runoff models: Optimization algorithms, catchment conditions, and model structure. Water Resour Res, **32**: 3513-3524.

Gupta HV, Kling H, Yilmaz KK, Martinez GF. 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, **377**: 80-91.

Gupta HV, Sorooshian S, Yapo PO. 1998. Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. Water Resour Res, **34**: 751-763.

Han D. 2011. Flood risk assessment and management. Bentham Science Publishers.

Jakeman A, Hornberger G. 1993. How much complexity is warranted in a rainfall-runoff model? Water Resour Res, **29**: 2637-2649.

Jakeman A, Littlewood I, Whitehead P. 1993. An assessment of the dynamic response characteristics of streamflow in the Balquhidder catchments. Journal of Hydrology, **145**: 337-355.

Jie MX, Chen H, Xu CY, Zeng Q, Tao XE. 2016. A comparative study of different objective functions to improve the flood forecasting accuracy. Hydrology Research DOI:10.2166/nh.2015.078

Kim H, Lee S. 2014. Assessment of a seasonal calibration technique using multiple objectives in rainfall–runoff analysis. Hydrol Process, **28**: 2159-2173.

Kim KB, Kwon H-H, Han D. 2016. Hydrological modelling under climate change considering nonstationarity and seasonal effects. Hydrology Research, **47**: 260-273.

Letcher R, Schreider SY, Jakeman A, Neal B, Nathan R. 2001. Methods for the analysis of trends in streamflow response due to changes in catchment condition. Environmetrics, **12**: 613-630.

Li H, Beldring S, Xu C-Y. 2014. Stability of model performance and parameter values on two catchments facing changes in climatic conditions. Hydrological Sciences Journal.

Littlewood I. 1999. Improved unit hydrograph characterisation of the daily flow regime (including low flows) for the River Teifi, Wales: towards better rainfall-streamflow models for regionalisation. Hydrology and Earth System Sciences, **6**: 899-911.

Madsen H. 2000. Automatic calibration of a conceptual rainfall–runoff model using multiple objectives. Journal of hydrology, **235**: 276-288.

Merz R, Blöschl G. 2009. A regional analysis of event runoff coefficients with respect to climate and catchment characteristics in Austria. Water Resour Res, **45**.

Merz R, Parajka J, Blöschl G. 2011. Time stability of catchment model parameters: Implications for climate impact analyses. Water Resour Res, **47**.

Moore R. 1985. The probability-distributed principle and runoff production at point and basin scales. Hydrological Sciences Journal, **30**: 273-297.

Nash J, Sutcliffe JV. 1970. River flow forecasting through conceptual models part I—A discussion of principles. Journal of hydrology, **10**: 282-290.

Paik K, Kim JH, Kim HS, Lee DR. 2005. A conceptual rainfall-runoff model considering seasonal variation. Hydrol Process, **19**: 3837-3850.

Patel H, Rahman A. 2014. Probabilistic nature of storage delay parameter of the hydrologic model RORB: a case study for the Cooper's Creek catchment in Australia.

Sorooshian S. 1991. Parameter estimation, model identification, and model validation: conceptual-type models. In: Recent advances in the modeling of hydrologic systems, Springer, pp: 443-467.

Tao G. 2003. Adaptive control design and analysis. John Wiley & Sons.

Tolson BA, Shoemaker CA. 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resour Res, **43**.

Vrugt JA, Gupta HV, Bouten W, Sorooshian S. 2003. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resour Res, **39**.

Wagener T. 2007. Can we model the hydrological impacts of environmental change? Hydrol Process, **21**: 3233-3236.

Wagener T, Boyle DP, Lees MJ, Wheater HS, Gupta HV, Sorooshian S. 2001. A framework for development and application of hydrological models. Hydrology and Earth System Sciences Discussions, **5**: 13-26.

Wagener T, McIntyre N, Lees M, Wheater H, Gupta H. 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. Hydrol Process, **17**: 455-476.

Wagener T, Sivapalan M, Troch PA, McGlynn BL, Harman CJ, Gupta HV, Kumar P, Rao PSC, Basu NB, Wilson JS. 2010. The future of hydrology: An evolving science for a changing world. Water Resour Res, **46**.

Xu C-Y, Vandewiele G. 1994. Sensitivity of monthly rainfall-runoff models to input errors and data length. Hydrological sciences journal, **39**: 157-176.

Xu C-y. 1999. Operational testing of a water balance model for predicting climate change impacts. Agricultural and forest meteorology, **98**: 295-304.

Yan C, Zhang W. 2014. Effects of model segmentation approach on the performance and parameters of the Hydrological Simulation Program–Fortran (HSPF) models. Hydrology Research, **45**: 893-907.

Zhang DJ, Chen XW, Yao HX, Lin BQ. 2015. Improved calibration scheme of SWAT by separating wet and dry seasons. Ecological Modelling, **301**:54-61.

Zeng Q, Chen H, Xu C-Y, Jie M-X, Hou Y-K. 2016. Feasibility and uncertainty of using conceptual rainfall-runoff models in design flood estimation. Hydrology Research, in press. doi: 10.2166/nh.2015.069.