Exploration on hydrological model calibration by considering the hydro-meteorological variability

Binru Zhao, Jingqiao Mao, Qiang Dai, Dawei Han, Huichao Dai and Guiwen Rong

ABSTRACT

The hydrological response is changeable for catchments with hydro-meteorological variations, which is neglected by the traditional calibration approach through using time-invariant parameters. This study aims to reproduce the variation of hydrological responses by allowing parameters to vary over clusters with hydro-meteorological similarities. The Fuzzy C-means algorithm is used to partition one-month periods into temperature-based and rainfall-based clusters. One-month periods are also classified based on seasons and random numbers for comparison. This study is carried out in three catchments in the UK, using the IHACRES rainfall-runoff model. Results show when using time-varying parameters to account for the variation of hydrological processes, it is important to identify the key factors that cause the change of hydrological responses, and the selection of the time-varying parameters should correspond to the identified key factors. In the study sites, temperature plays a more important role in controlling the change of hydrological responses than rainfall. It is found that the number of clusters has an effect on model performance, model performances for calibration period become better with the increase of cluster number; however, the increase of model complexity leads to poor predictive capabilities due to overfitting. It is important to select the appropriate number of clusters to achieve a balance between model complexity and model performance.

Key words | Fuzzy C-means algorithm, hydrological model calibration, hydro-meteorological variability

Binru Zhao
Jingqiao Mao (corresponding author)
Huichao Dai
College of Water Conservancy and Hydropower
  Engineering,
Hohai University,
Nanjing 210098,
China
E-mail: *maojq@hhu.edu.cn*

Binru Zhao
Dawei Han
Water and Environmental Management Research
  Centre, Department of Civil Engineering,
University of Bristol,
Bristol,
UK

Qiang Dai
Key Laboratory of VGE of Ministry of Education,
Nanjing Normal University,
Nanjing 210023,
China

Guiwen Rong
College of Earth and Environment,
Anhui University of Science and Technology,
Huainan 232001,
China

INTRODUCTION

Understanding the hydrological response of catchments is crucial for various issues related to water resources management. Hydrological models with varying degrees of complexity have been developed to represent the rainfall-runoff transformation relationship. The accuracy of hydrological models is affected by multiple factors. The error associated with the observed data is one of these factors. Although improving the measuring technology

could reduce the error, for the existing data, noise reduction could be an effective pre-processing method and has been widely explored and applied (Chou 2014; Li *et al.* 2018). The model's representation of the hydrological process (or model structure) also affects the model performance. Melsen *et al.* (2016) studied the representation of spatial and temporal variability in large-domain hydrological models through investigating parameter transferability across different temporal and spatial resolutions. Magnusson *et al.* (2015) demonstrated the usefulness of a multimodel framework for identifying appropriate model structures according to data availability, properties of interest and computational

cost. Due to the lack of field measurements, hydrological model parameters are generally estimated through calibration, which is also a key procedure controlling the model capability. Some works investigated the role of the multi-objective calibration in improving model accuracy (Zhang *et al.* 2016, 2018b; Her & Seong 2018), and some works aimed at the improvement of the optimization algorithm for calibration (Wang *et al.* 2012). Taking into account the variation of the hydrological response during model calibration is another way to improve the model performance, which is also the task of this study.

It is widely accepted that the more stable the catchment conditions are, the better should the estimated parameters represent the hydrological response, and the more similar the calibration data is to the validation data, the better should be the performance of hydrological models. Based on these cognitions, many studies attempt to use varying model parameters to capture the variation of the hydrological response which is caused by climatic variation or land-cover changes in catchments. Efstratiadis *et al.* (2015) simulated the hydrological process of the Ferson Creek basin (USA) that has experienced growing urbanization over the past 30 years through employing a lumped conceptual model with one time-varying parameter and a semi-distributed scheme based on two hydrological response units with the time-varying surface area. Pathiraja *et al.* (2016a, 2016b) investigated the potential of data assimilation techniques to detect temporal patterns in hydrological model parameters from streamflow observations, and then examined the proposed method to paired catchment systems in Western Australia that have different extents of deforestation. The results demonstrate that the time-varying model structures are able to improve both predictions and modelling of changing catchments. Sadegh *et al.* (2019) proposed the Nonstationary Rainfall-Runoff Toolbox (NRRT) to permit time-varying realizations of hydrological models to predict a nonstationary hydrological response in watersheds, where the physical changes are manifested in time-varying parameters in a conceptual model. Their case study in the Wights catchment in Australia shows that the decrease of the maximum capacity of the production store ($S1_{max}$) of the GR4J model adequately represents the loss of near surface storage due to deforestation.

In addition to employing time-varying model parameters to capture the variation of the hydrological response that is caused by land-use changes, time-varying parameters are also used to account for the effect of climatic temporal variations on hydrological processes, including intra-annual variations (or seasonal variations) and inter-annual variations. For attempts focusing on seasonal variations of hydrological responses, explorations are made under the hypothesis that the hydrological response of different seasons can be reproduced by using different parameter sets. Paik *et al.* (2005) proposed a seasonal tank model to calibrate season-varying parameters for three four-month seasons. The application of the seasonal tank model to a watershed located in central South Korea indicates that the seasonal tank model has a smaller sum of square errors than those of the non-seasonal tank model for the calibration period. LéVesque *et al.* (2008) evaluated the hydrological behaviour of the SWAT model by distinguishing the hydrological dynamics related to winter and summer seasons for two watersheds in southeastern Canada. The summer performance was considerably improved when only summer streamflow was provided for calibration. However, calibration based solely on winter observations resulted in minor improvements in performance. Luo *et al.* (2012) explored the possible effects of the hydrologic non-stationarity by testing ten parameterization schemes at 12 catchments in eastern Australia. Results show that among all parameterization schemes, calibrating the model using the data from each individual month benefits the seasonal streamflow forecasting. Zhang *et al.* (2018a) configured a season-based probability-distributed model (PDM-CEMADEN) to simulate different hydrological responses during wet and dry seasons. The season-based models constructed for five basins in southeastern Brazil are adequate to reproduce the intra-annual and inter-annual variability of the streamflow.

As for the impact of inter-annual climatic variations on hydrological processes, Klemeš (1986) initially considered the necessity of verifying the hydrological model under different climate conditions. He proposed a differential split-sampling test to identify two periods with different climatic characteristics, the hydrological model was then calibrated and validated by the contrast periods. This method was applied to 273 catchments in Austria by Merz *et al.* (2011). They found that parameters representing snow and soil moisture processes have high correlations to

changing climatic conditions in more recent years, such as higher evapotranspiration and drier soil conditions. Under these changing climatic conditions, the simulation errors clearly increase as the time lag between the simulation and calibration periods increases. Brigode et al. (2013) classified the available records to four three-year sub-periods on the basis of the Aridity Index (here defined as the ratio of mean Penman potential evapotranspiration to mean precipitation): a wet sub-period, two dry ones and an intermediate one. The driest sub-period was used as the validation period and the three others were used as calibration periods separately. The results show that the model performance is the worst when the wet sub-period is used for calibration. Kim et al. (2016) investigated the calibration scheme where one parameter of the IHACRES model is selected to vary against time and climate conditions while other parameters remain fixed. They found that the model that takes into account the nonstationary effects works well for both calibration and validation periods.

\When using time-varying parameters to account for the variation of hydrological responses caused by climatic variations, clustering methods are commonly used to identify periods with similar climatic characteristics, during which parameter sets are assumed the same. Choi & Beven (2007) classified sub-periods with the length of 30 days into 15 clusters using the Fuzzy C-means algorithm, where the climatic conditions are described with six variables. They then calibrated and validated the TOPMODEL for each cluster in the Generalized Likelihood Uncertainty Estimation (GLUE) framework. Although satisfactory model performance could be achieved at the global level, there was no parameter set that performs well for all 15 clusters. de Vos et al. (2010) employed the k-means clustering algorithm to classify the historical data into 12 clusters with similar characteristics in terms of precipitation and soil moisture, and model parameters are allowed to vary over clusters. They improved the model structure by analyzing the variation pattern of parameters. A clustering method based on self-organizing maps (SOM) was also used to partition climatic conditions by Toth (2009). They found that an adequate distinction of the climatic conditions may considerably improve the rainfall-runoff modelling performance.

Although it is widely recognized that allowing parameters to vary according to the variations of climatic conditions could improve the hydrological modelling performance, some issues remain unexplored. For example, when identifying similar climatic conditions, which climatic factors are more related to the variation of the hydrological response? Besides, despite the fact that time-varying parameters could better reproduce the real hydrological response, it will increase the model complexity and has an effect on the predictive capabilities of the model. This study aims to investigate these problems. In order to identify periods with climatic similarities, the Fuzzy C-means (FCM) algorithm is used to partition one-month periods to different clusters. Here the FCM algorithm is executed based on the temperature information and rainfall information of one-month periods separately. One-month periods are also classified on the basis of seasons and random numbers for comparison. Parameters are allowed to vary over clusters during the calibration procedure. Model performances are then evaluated using the criteria R^2, R_{ln}^2 and relative bias to represent high flows, low flows and water balance. The trade-off between the model complexity and model performance is studied by evaluating the model performance under different numbers of the cluster. Three catchments in southwest UK are selected to carry out this study, with the use of a lumped conceptual rainfall-runoff model IHACRES.

STUDY SITES

Three catchments located in southwest UK are explored in this study: the Exe River at Thorverton (45001), Brue River at Lovington (52010) and Avon River at Great Somerford (53008). The main land use of these three catchments is grassland and horticulture, presenting little change in recent decades. Figure 1 shows the location of the selected catchments and the corresponding stream gauging stations. Information on these three catchments and available data are listed in Table 1. The average daily rainfall data are obtained from the NERC Environmental Information Data Centre (Tanguy et al. 2016). The catchment average temperatures are calculated with the use of the UKCP09 gridded observation data sets, and the National River Flow Archive (NRFA) provides the daily time series of observed streamflow data. All data have been checked for possible outliers and missing data, etc. The data during
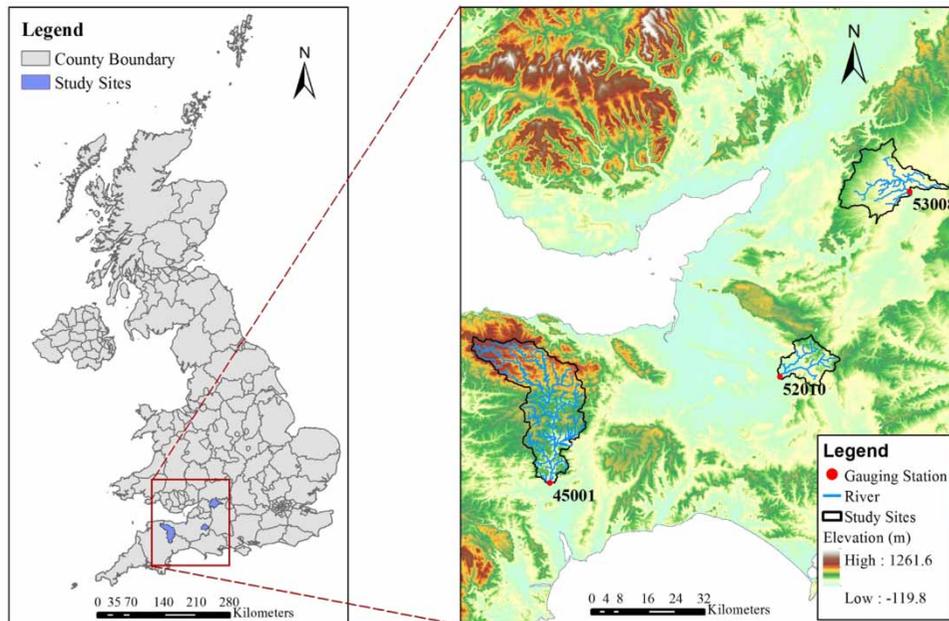
Figure 1 │ Location of study sites.

**Table 1** │ Information and characteristics of the study sites

Basin	Gauging station	Area (km²)	Available records		Annual rainfall (mm)	Annual streamflow (mm)
Exe River	Thorverton (45001)	600.9	Temperature Rainfall Streamflow	1960–2016 1961–2015 1956–2016	1295	850
Brue River	Lovington (52010)	135.2	Temperature Rainfall Streamflow	1960–2016 1961–2015 1964–2016	865	440
Avon River	Great Somerford (53008)	303	Temperature Rainfall Streamflow	1960–2016 1961–2015 1964–2016	841	345

the period 2003–2015 is selected for analysis because this period has minimal missing records for all types of data for all catchments.

Figure 2 shows the average temperature characteristics at the study sites. The UK has four seasons: spring (March–May), summer (June–August), autumn (September–November) and winter (December–February). From the distribution of the average monthly temperature in Figure 2(a), it is seen that the temperature difference between summer and winter months is not significant, ranging from 4.4 °C in February to 16.3 °C in July. The studied catchments are hardly affected by snow. Figure 2(b) shows the temporal variations of monthly temperature, a clear seasonal pattern is seen. The average annual temperature indicates the

temperature rise during the study period is not distinct compared with temperature intra-annual variations, therefore the climate change could be negligible in this study.

The distribution of the average monthly rainfall and measured streamflow for three catchments is shown in Figure 3. The three catchments show a similar pattern for both rainfall and streamflow. They have heavy rainfall all year round, and there is a seasonal pattern with wet autumns and winters and relatively dry springs and summers. It is clear that the monthly rainfall varies greatly over years, especially for the summer season. The average monthly streamflow shows a similar pattern to rainfall, where the autumn and winter have high streamflow, while the spring and summer have low streamflow. Despite this
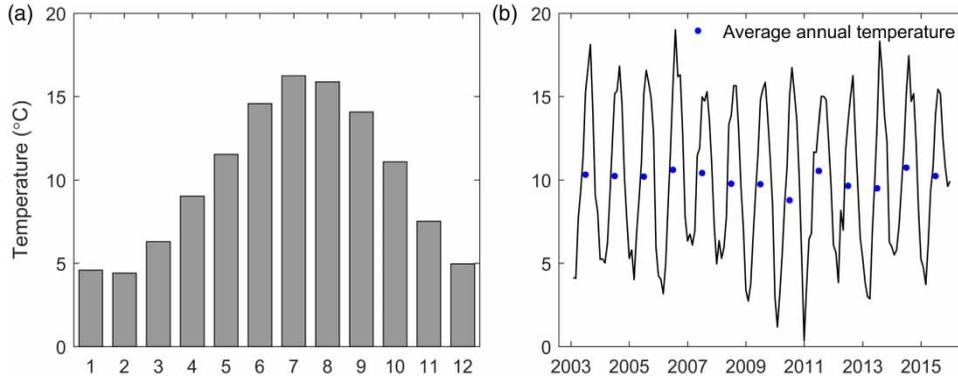
Figure 2 | Temperature characteristics of the studied sites over the period 2003–2015: (a) the average of monthly temperature, (b) the change of monthly temperature.
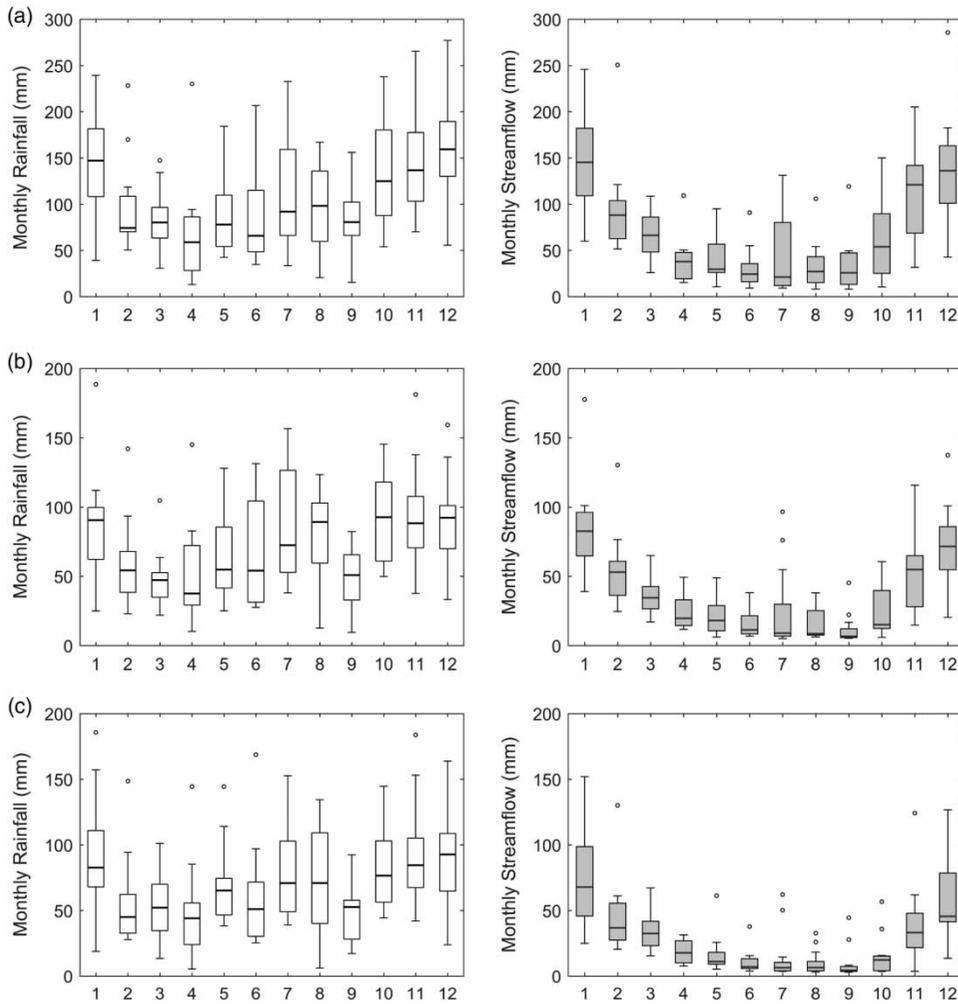
Figure 3 | The distribution of the average monthly rainfall (left) and streamflow (right) for three catchments over the period 2003–2015: (a) Thorverton catchment, (b) Lovington catchment, (c) Great Somerford catchment.

similar pattern, it is interesting to find that the decrease of streamflow in summer is more distinct compared with that of rainfall, and the summer streamflow has less interannual variations. This could be explained by the high temperature of summer, which plays a key role in controlling the runoff through affecting the evapotranspiration process. Given the

location of these three catchments, it is found that the monthly rainfall and monthly streamflow decrease from the west (Thorverton catchment) to the east (Great Somerford catchment).

METHODS

Rainfall-runoff model

The IHACRES model (Jakeman & Hornberger 1993) is a lumped conceptual rainfall-runoff model which has been widely applied to a range of catchments for hydrological analysis and climate impact studies due to its simple structure and less requirement for input data (Jakeman *et al.* 1993; Letcher *et al.* 2001; Kim & Lee 2014; Kim *et al.* 2016).

The IHACRES model consists of two modules in series: the non-linear loss module and the linear routing module, as shown in Figure 4. The non-linear loss module calculates effective rainfall by calculating the catchment wetness index on the basis of rainfall and temperature. The percentage of rainfall that becomes effective rainfall varies linearly from 0 to 100% as the catchment wetness index varies between zero to unity. The linear routing module then converts the effective rainfall to streamflow based on the unit hydrograph theory, where the catchment is conceptualized as a configuration of linear storages acting in series and/or parallel. Model parameters are listed in Table 2.

Calibration schemes

The parameters of the hydrological model are allowed to vary over clusters during the calibration procedure. These

Table 2 | List of parameters in the IHACRES model

Module	Parameter	Description
Non-linear	C	Mass balance
	τ_w	Reference drying rate
	f	Temperature modulation of drying rate
Linear	v_q	Relative volume of quick flow to total flow, $v_q = \beta_q/(1 - \alpha_q)$
	v_s	Relative volume of slow flow to total flow, $v_s = 1 - v_q$
	τ_q	Quick flow recession time constant, $\tau_q = -\Delta/ln(\alpha_q)$
	τ_s	Slow flow recession time constant, $\tau_s = -\Delta/ln(\alpha_s)$

clusters consist of one-month periods in the calibration period, and the one-month period is divided based on the calendar month. There are four types of cluster, which are identified based on temperature similarity, rainfall similarity, calendar seasons and random numbers, respectively. Temperature and rainfall are crucial variables affecting the hydrological response. Clusters considering their similarities are classified using the FCM algorithm. The reason for exploring seasons is that the study areas show seasonal variations of hydro-meteorological conditions, therefore clusters based on seasons could take into account the similarity of both temperature and rainfall, although these similarities are not as distinct as those identified using the FCM algorithm. Clusters based on random numbers are just used for comparison. In addition, for the purpose of better evaluating the parameter-varying calibration scheme, we also studied the traditional calibration approach which uses time-invariant parameters. As a result, there are five calibration schemes, referred to as Tradition scheme, FCM_T scheme, FCM_R scheme, Season scheme, and Random scheme.
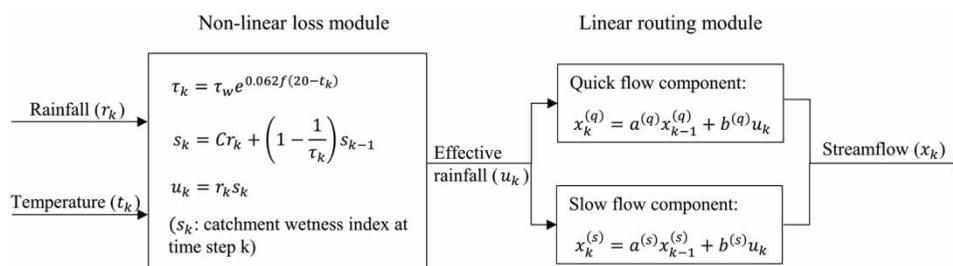
Figure 4 | Structure of the IHACRES model.

Fuzzy C-means algorithm

The FCM algorithm is an unsupervised clustering algorithm, initially proposed by Bezdek (1981). During clustering, objects with similar characteristics are classified into one cluster, and objects in different clusters are dissimilar in terms of the same characteristics (Sbai 2001; Pakhira *et al.* 2004). In this study, the FCM algorithm was used to partition multiple one-month periods based on hydro-meteorological conditions in terms of temperature and rainfall. The temperature information of the one-month period was described with four variables: average monthly temperature, maximum monthly temperature, minimum monthly temperature, and monthly temperature variance. The rainfall information of the one-month period was described with the following variables: monthly rainfall, maximum monthly rainfall, the rate of rainy days and monthly rainfall variance.

When classifying the multiple one-month periods $X = \{x_1, x_2, \ldots, x_n\}$ into k clusters represented as fuzzy sets $(F_j, j = 1, \ldots, k)$, the algorithm is carried out by minimizing the following objective function:

$$J_m = \sum_{i=1}^{n} \sum_{j=1}^{k} (\mu_{ij})^m \|x_i - c_j\|^2 \tag{1}$$

where μ_{ij} is the membership degree of x_i to the fuzzy cluster set F_j, $\sum_j \mu_{ij} = 1$. $m \in [1, \infty)$ is a weight exponent controlling the degree of fuzzification, c_j is the cluster centroids of the fuzzy cluster set F_j, and $\|x_i - c_j\|$ is a Euclidean norm between x_i and c_j. In this study, x_i is the ith variable of each hydro-meteorological factor.

Fuzzy partitioning is performed through an iterative optimization of the above objective function, with the membership degree μ_{ij} and the cluster centroids c_j updated until J_m cannot be further improved.
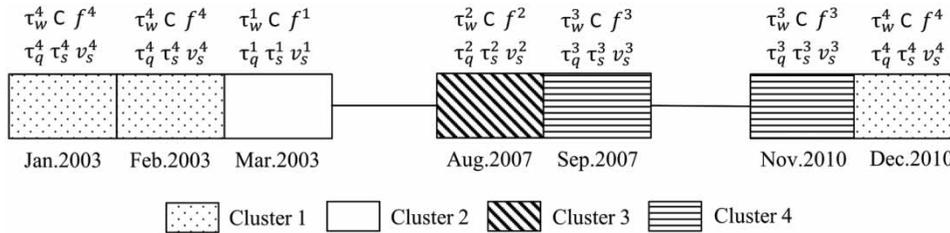
The number of clusters needs to be defined before the FCM algorithm is conducted. When exploring the performance of different calibration schemes, to avoid the effect of the difference in model complexity which is associated with the number of clusters, we define the number of clusters as four, which equals the number of seasons. When investigating the effect of model complexity, model performance under different numbers of clusters are tested.

Model calibration

When calibrating the parameter-varying hydrological model, there are two calibration methods: the parallel calibration scheme (PCS) and the serial calibration scheme (SCS) (Kim & Han 2017). For the PCS approach, parameter sets for different clusters are calibrated parallelly. Each time the model is run, only the data belonged to one cluster is used in the objective function although the model is run for the whole calibration period. In this way, the parameter set for this cluster could be calibrated. When there are n clusters, the model needs to be run for n times to calibrate n sets of parameters. With all parameter sets, the simulated streamflow could be obtained by extracting and combining the streamflow of each cluster that is simulated with its corresponding parameter set. As for the SCS approach, all parameter sets are calibrated simultaneously. Parameters vary according to the cluster the data belong to. When the calibration procedure switches from one one-month period to the following one, the subsequent period's state variables and streamflow are updated with the prior period's ones. The PCS approach is easy to implement and widely used; however, the state variables and simulated streamflow are discontinuous, which does not make sense. The SCS approach increases the complexity of the model while overcoming the discontinuous problem. In this study, the SCS approach is employed to calibrate the varying parameters.

The IHACRES model has six parameters as listed in Table 2: three parameters in the non-linear loss module and three parameters in the linear routing module $(v_s = 1 - v_q)$. During calibration, all parameters except for C vary over clusters, and a value of C is selected such that the volumes of effective rainfall and observed streamflow are equal over the calibration period. Therefore, when there are n clusters, the number of parameters is $5n + 1$. The calibration procedure is illustrated in Figure 5, which takes the season-based clusters as an example. As is seen, each cluster is assigned one set of parameters, when the data switch from one cluster to another, the parameters vary accordingly.

During the calibration procedure, Nash–Sutcliffe Efficiency (NSE) (Nash & Sutcliffe 1970) is used to minimize the difference between the observed and simulated

$$\tau_w^4 \, C \, f^4 \qquad \tau_w^4 \, C \, f^4 \qquad \tau_w^1 \, C \, f^1 \qquad\qquad \tau_w^2 \, C \, f^2 \qquad \tau_w^3 \, C \, f^3 \qquad\qquad \tau_w^3 \, C \, f^3 \qquad \tau_w^4 \, C \, f^4$$

$$\tau_q^4 \, \tau_s^4 \, v_s^4 \qquad \tau_q^4 \, \tau_s^4 \, v_s^4 \qquad \tau_q^1 \, \tau_s^1 \, v_s^1 \qquad\qquad \tau_q^2 \, \tau_s^2 \, v_s^2 \qquad \tau_q^3 \, \tau_s^3 \, v_s^3 \qquad\qquad \tau_q^3 \, \tau_s^3 \, v_s^3 \qquad \tau_q^4 \, \tau_s^4 \, v_s^4$$

Jan.2003   Feb.2003   Mar.2003   Aug.2007   Sep.2007   Nov.2010   Dec.2010

Cluster 1   Cluster 2   Cluster 3   Cluster 4

Figure 5 | Illustration of the calibration procedure which allows parameters to vary over clusters (clusters here are classified based on four seasons in the UK, and there are 21 parameters).

streamflow, defined as:

$$NSE = 1 - \frac{\sum_{i=1}^{n} (Q_i^o - Q_i^m)^2}{\sum_{i=1}^{n} (Q_i^o - \overline{Q^o})^2} \qquad (2)$$

where Q_i^o and Q_i^m are the *i*th observed and simulated streamflow, respectively. $\overline{Q_o}$ is the arithmetic mean of the observed streamflow and n is the total number of days in the calibration period. The NSE can vary from $-\infty$ to 1. An efficiency of 1 corresponds to a perfect match of the simulated streamflow to the observed streamflow.

The shuffled complex evolution (SCE-UA) method (Duan *et al.* 1992) is used to maximize the above objective function. We first used the period 2003–2010 as the calibration period and the period 2011–2015 as the validation period. In order to improve the reliability of results, the period 2008–2015 was then used to calibrate the model, and the period 2003–2007 was for validation. The first year of the calibration period was for the warm-up of the model. Model performance was assessed based on the average values of two calibration periods and two validation periods.

Model evaluation

When validating the hydrological model for FCM_T scheme, FCM_R scheme and Season scheme, each one-month period in the validation period is assigned one parameter set according to its similarity to the existing clusters. As for the Random calibration scheme, the one-month periods in the validation period are assigned parameter sets randomly. Once the parameter sets are assigned to the one-month period, the model is run with varying parameters.

Model performance is assessed for all the calibration periods and validation periods based on three evaluation criteria: R^2, R_{ln}^2 and relative bias:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (Q_i^o - Q_i^m)^2}{\sum_{i=1}^{n} (Q_i^o - \overline{Q^o})^2} \qquad (3)$$

$$R_{ln}^2 = 1 - \frac{\sum_{i=1}^{n} [\ln(Q_i^o + Q_{90}^o) - \ln(Q_i^m + Q_{90}^o)]^2}{\sum_{i=1}^{n} [\ln(Q_i^o + Q_{90}^o) - \overline{\ln(Q^o + Q_{90}^o)}]^2} \qquad (4)$$

$$\text{bias}(\%) = \frac{\sum_{i=1}^{n} (Q_i^o - Q_i^m)}{\sum_{i=1}^{n} Q_i^o} \times 100 \qquad (5)$$

where Q_i^o and Q_i^m are the *i*th observed and simulated streamflow, respectively. $\overline{Q_o}$ is the arithmetic mean of the observed streamflow and n is the total number of days in the calibration period. Q_{90}^o represents the 90th percentile of the observed non-zero streamflow. R^2 is commonly used to assess the overall fit of a hydrograph which is sensitive to high flow events (Croke 2009). R_{ln}^2, which is the logarithmic form of R^2, is often used to reduce the sensitivity of extreme values and results in the increasing sensitivity of low flows (Krause *et al.* 2005; Kim & Lee 2014). The value of R^2 and R_{ln}^2 can vary from $-\infty$ to 1, with the value of 1 corresponding to an optimal model. According to Moriasi *et al.* (2007), the model performance is considered very good when both R^2 and R_{ln}^2 are greater than 0.75, good when the values are in the range of 0.65–0.75, and satisfactory if they are 0.50–0.65. The relative bias is used to assess the water balance error for a certain period. The perfect result is achieved when the bias equals zero. The larger the absolute value, the worse the result, and the positive and negative values correspond to the underestimation and overestimation of streamflow, respectively.

RESULTS

Classification of one-month periods

Figure 6 shows the distribution of clusters identified based on temperature, rainfall, season and random numbers, respectively. Clusters for temperature are numbered according to the value of average monthly temperature, and clusters for rainfall are numbered based on the value of monthly rainfall (from a low level to a high level). As for seasons, the numbers 1–4 refer to winter, autumn, spring and summer, respectively. As is seen, for clusters identified using the FCM algorithm, the same month of different years does not always belong to the same cluster, indicating that there are inter-annual variations in terms of temperature and rainfall. From the distribution of temperature-based clusters it is found that there is a similar pattern to the seasonal-based clusters, where the warm clusters are distributed in the middle of the year and the temperature of months at the start and end of the year is lower. As for the rainfall-based cluster, it shows a large difference in annual rainfall over the study period. Taking the Thorverton catchment (Figure 6(a)) as an example, most months in 2003 belong to Clusters 1 and 2, while the months in 2012 mostly belong to Clusters 3 and 4, which indicates the year 2003 has less rainfall than 2012. In summary, the clustering
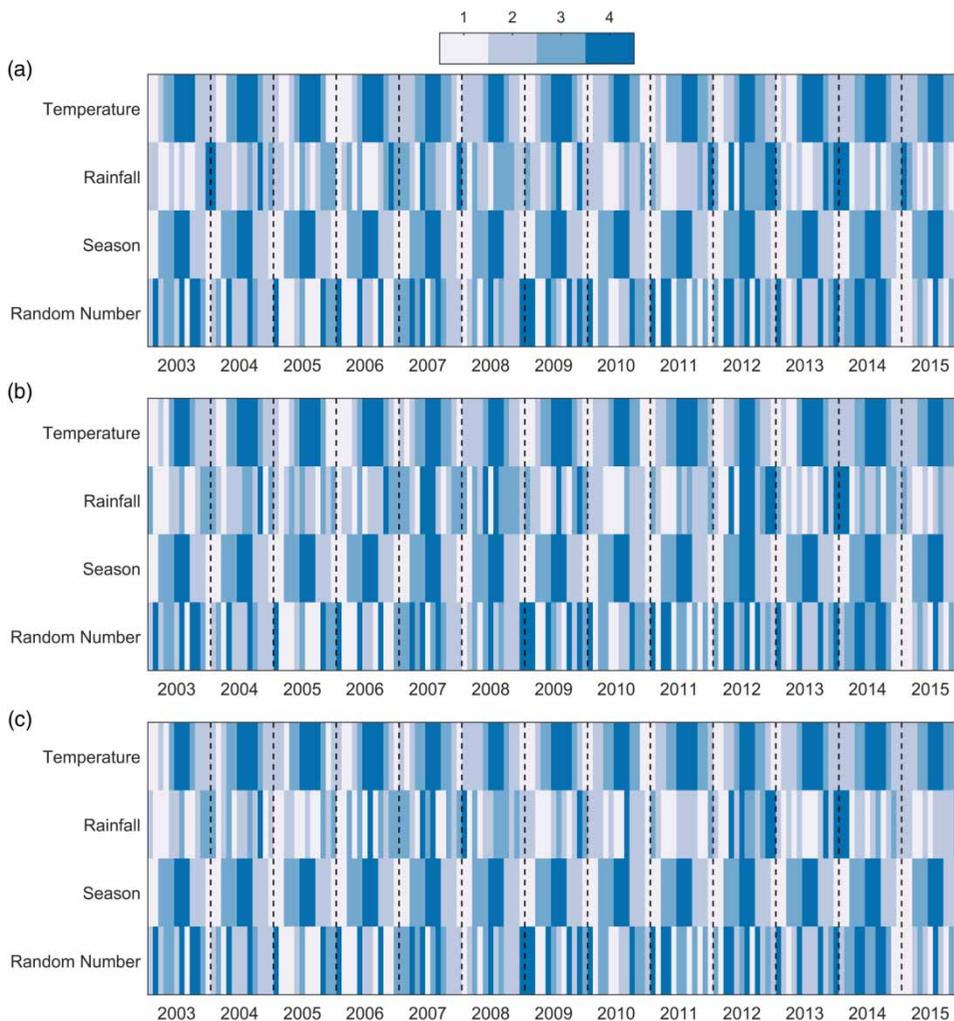
Figure 6 │ Distribution of clusters identified based on temperature, rainfall, season and random numbers for the study period 2003–2015: (a) Thorverton catchment, (b) Lovington catchment, (c) Great Somerford catchment.

algorithm shows superiority in considering the inter-annual and intra-annual variation of temperature and rainfall, compared with the season-based approach. By comparing the cluster distribution of these three catchments, it is found that clusters based on temperature and rainfall have a similar distribution, indicating that the variations in terms of temperature and rainfall is similar for these three catchments.

The distribution of average monthly temperature and monthly rainfall of the objects in different clusters is explored. It is found that the three catchments have a similar distribution, therefore, we take the Thorverton catchment as an example, whose corresponding distribution is shown in Figure 7. As is seen, the difference of average monthly temperature among clusters is most significant for temperature-based clusters, followed by season-based clusters. Clusters identified based on rainfall and random numbers have no distinct difference in terms of the average monthly temperature. As for the monthly rainfall, the distinct difference can be found among rainfall-based clusters, while the difference among other clusters is not significant. In addition to the difference among clusters, it is also found that the similarity of the objects in the same cluster varies greatly. For instance, the objects in each temperature-based cluster are highly similar in terms of the average monthly temperature, and the objects in each rainfall-based cluster also have a high level of similarity in terms of the monthly rainfall. As the temperature-based cluster and rainfall-based cluster are

identified using the FCM algorithm, it is inferred that the FCM algorithm has a better performance in grouping objects with similar characteristics and separating objects that are dissimilar in terms of the same characteristics.

Model performance in calibration

Figure 8 shows the model performance of five calibration schemes (Tradition scheme, FCM_T scheme, FCM_R scheme, Season scheme and Random scheme) over the calibration period. During calibration, the parameter C is determined to ensure that the volumes of the effective rainfall and observed streamflow are equal over the calibration period, so the relative bias of the calibration period is equal to zero. Therefore, only the criteria R^2 and R_{ln}^2 are used to evaluate the model performance in the calibration period. The value of these criteria is the average value of two calibration periods. The model performance in terms of R^2 is greater than 0.8 for almost all calibration schemes at the three catchments. However, the value of R_{ln}^2 is relatively low. This indicates that the model performs better in simulating high flows than low flows. This could be explained by the fact that the model is calibrated only based on NSE, so the calibration procedure only focuses on matching one aspect of the hydrological process reflected in the observations and ignores other hydrological processes, as NSE is sensitive to high flows, the calibrated model has a better performance in simulating high flows.
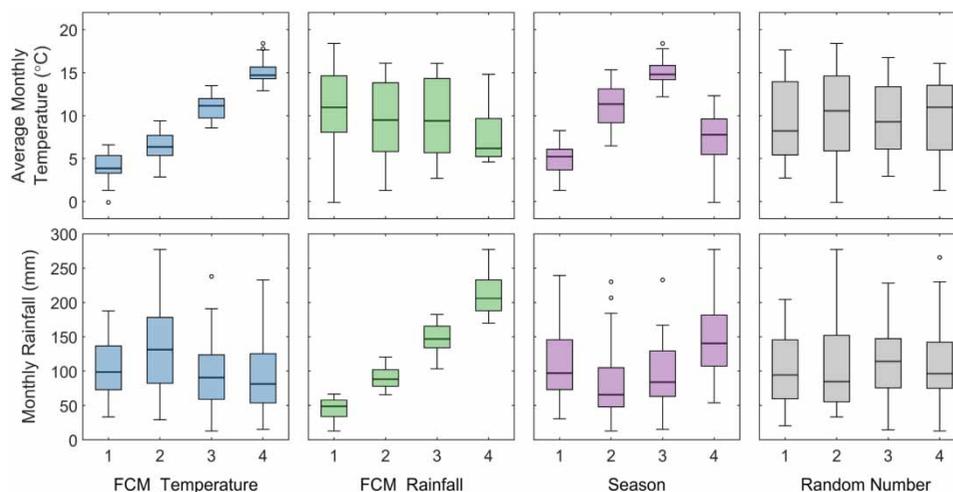
Figure 7 | Distribution of average monthly temperature and monthly rainfall of the objects in different clusters for the Thorverton catchment.
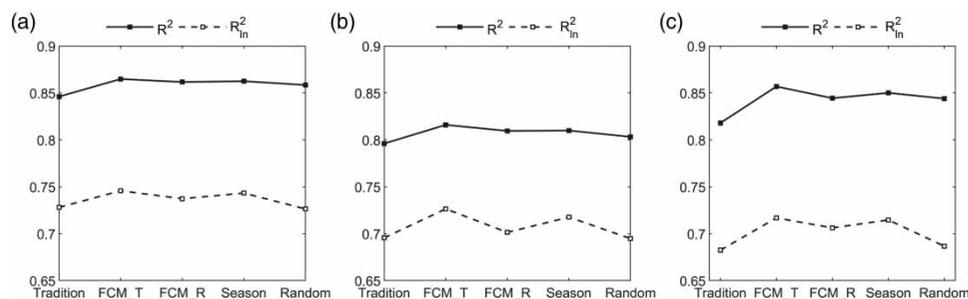
Figure 8 | Model performance (R^2 and R^2_{ln}) of different calibration schemes over the calibration period: (a) Thorverton catchment, (b) Lovington catchment, (c) Great Somerford catchment.

Through comparing the model performance of different calibration schemes the same result could be found for the three catchments. Calibration schemes that allow parameters to vary perform better than the traditional calibration approach in terms of R^2 and R^2_{ln}, indicating that allowing parameters to vary could better reproduce the hydrological process by considering the change of the hydrological response. Among calibration schemes that allow parameters to vary, the FCM_T scheme has the best performance, followed by the Season scheme, though the extent of the improvement caused by these two calibration schemes varies among three catchments. Given that the temperature-based clusters and season-based clusters have similar patterns in recognizing the temperature variation, it could be inferred that allowing parameters to vary according to temperature similarities could achieve better model performance in the calibration period for the studied catchments.

The FCM_T scheme and Season scheme were further explored by analyzing their seasonal performance with the use of R^2, as shown in Figure 9. The seasonal R^2 is calculated by using the seasonal data of the observed and simulated streamflow. From Figure 9, the model performance shows a distinct seasonal variation in terms of R^2.

Winter and spring have better performance, while the accuracy of summer streamflow simulations is poor. The possible reason is that the summer streamflow of the study sites is much lower compared with other seasons, and the model performs better in simulating high flows than low flows due to the high sensitivity of the objective function NSE to high flows. Despite the poor performance of summer, it is found there is a significant improvement produced by the FCM_T scheme and Season scheme. For other seasons, their R^2 values are also improved with the use of the FCM_T scheme and Season Scheme at different extents.

Model performance in validation

The calibrated parameter sets for the five calibration schemes over two calibration periods (2003–2010 and 2008–2015) were validated using the data in the periods 2011–2015 and 2003–2007 respectively. The model performance of the five calibration schemes for validation periods is compared in Figure 10 with the use of the average value of two validation periods. The improvement produced by parameter-varying calibration schemes (except for the Random scheme) is more significant compared with that
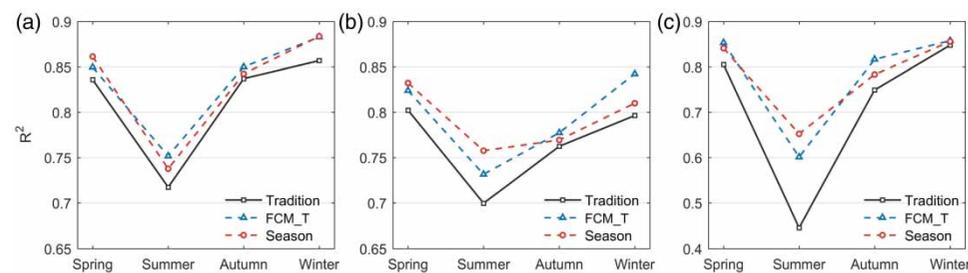
Figure 9 | Seasonal model performance (R^2) of calibration schemes (Tradition, FCM_T, Season) over the calibration period: (a) Thorverton catchment, (b) Lovington catchment, (c) Great Somerford catchment.
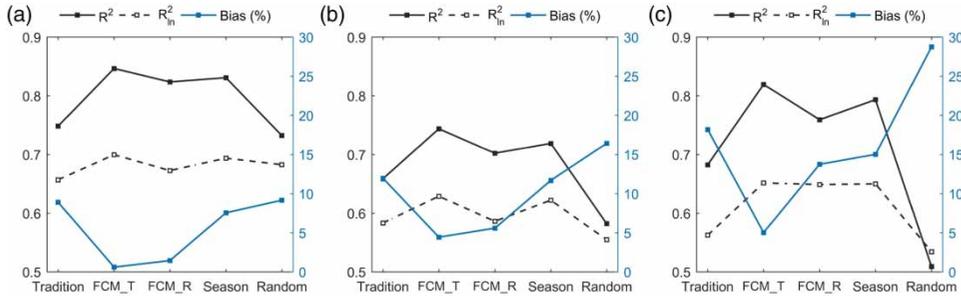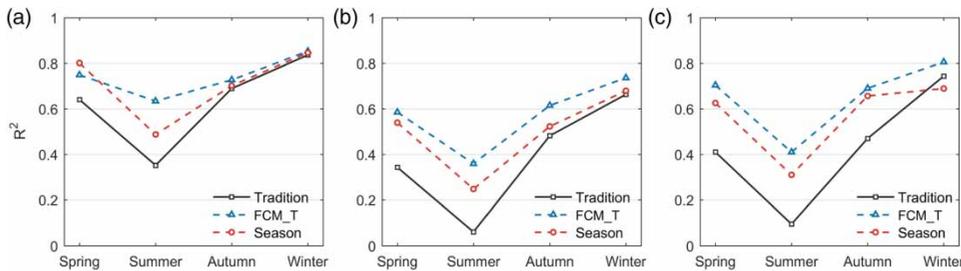
Figure 10 | Model performance (R^2, R^2_{ln} and Bias) of different calibration schemes over the validation period: (a) Thorverton catchment, (b) Lovington catchment, (c) Great Somerford catchment.

of the calibration period. The FCM_T scheme and the Season scheme have relatively higher values of R^2 and R^2_{ln} compared with other calibration schemes for all catchments, which is similar to the results of the calibration period. In the validation period, not all calibration schemes that allow parameters to vary could lead to better performances than the traditional scheme. For example, the model performance of the Random scheme in R^2 is poorer than the traditional scheme for both the Lovington and Great Somerford catchments. This indicates that although allowing parameters to vary could improve the hydrological model performance, it is of great importance to define the appropriate cluster which could represent the variation of the hydrological response, otherwise, the increased model complexity may have adverse impacts on the model's predictive capabilities. The value of R^2_{ln} is still lower than that of R^2 for all calibration schemes, indicating the calibrated model has better capabilities in simulating high flows than low flows, this is caused by the choice of the objective function for the calibration procedure. For the relative bias, it is seen that the FCM_T scheme has the smallest bias for three catchments. Although the bias of the Season scheme is not as good as the FCM_T scheme, it is still better than the

traditional approach. Despite the three catchments presenting similar results in terms of the improvement caused by the parameter-varying calibration schemes, it is clear that their model performance differs. In general, the Thorverton catchment has a better performance than the other two catchments, with relatively higher R^2 and R^2_{ln} and lower bias, which also applies to the calibration period. The difference in model performance could be caused by the difference of the catchment properties, for example, the Lovington catchment and the Great Somerford catchment have less rainfall than the Thorverton catchment, the vegetation conditions of the three catchments may show differences, etc. Although it is important to find the reasons that lead to the model performance difference, it is beyond the scope of this study and will be investigated in future work.

The FCM_T scheme has a better performance in terms of R^2, R^2_{ln} and relative bias in validation, and the Season scheme has higher values of R^2 and R^2_{ln}. The seasonal model performance (R^2) of these two schemes for validation periods are also compared with that of the Tradition scheme, as shown in Figure 11. The pattern of the improvement caused by the FCM_T scheme and the Season scheme in the validation period is similar to that in the validation

Figure 11 | Seasonal model performance (R^2) of calibration schemes (Tradition, FCM_T, Season) over the validation period: (a) Thorverton catchment, (b) Lovington catchment, (c) Great Somerford catchment.

period. The value of R^2 are improved in almost all seasons at different extents with the use of the FCM_T scheme and Season scheme. The improvement in summer is the most significant, though the value of R^2 for summer is still low.

The variation of model parameters

As the FCM_T calibration scheme has the superior model performance for both calibration and validation periods, the model parameters of this calibration scheme are used to analyse the variation of parameters against clusters. It is found that the variation pattern of parameters is the same for two calibration periods. Figure 12 shows the distribution of model parameters calibrated during the period 2003–2010 for three catchments. The average monthly temperature is lowest for one-month periods in Cluster 1 and highest for one-month periods in Cluster 4. It is seen that the parameter τ_w and f in the non-linear loss module show distinct variation patterns against clusters for all catchments, while there is no obvious variation pattern of the parameter and v_s in the linear routing module. The reference drying rate τ_w shows a decreasing trend with the increase of temperature. This is plausible since when the reference drying rate is small, according to the equations of the non-linear module in Figure 4, the soil tends to be

drier (smaller catchment wetness index), which is achieved when the temperature is high. The temperature modulation of drying rate f controls the sensitivity of drying rate τ_k to changes in temperature, showing an increase trend with the increase of temperature. For the cluster with lower temperatures, the difference between the real temperature and the reference temperature 20 °C is relatively larger and the reference drying rate τ_w is higher. Due to these two factors, the drying rate τ_k may have a larger variation range even the temperature varies within a small range. In this case, a smaller value of the parameter f could decrease the variation range of the drying rate τ_k in this cluster. In contrast, for the cluster with higher temperatures, the difference between the real temperature and the reference temperature 20 °C is relatively smaller and the reference drying rate τ_w is lower, which makes the variation of the drying rate τ_k insignificant even when the temperature shows an obvious difference. Here a larger value of the parameter f could solve this problem. Therefore, the variation pattern of the parameter f is also plausible.

It is interesting to find that the water loss process controlled by the parameter τ_w and f has a high correlation with the temperature, this is also the reason why these two parameters show distinct variation patterns against the change of temperature. The linear routing module aims at
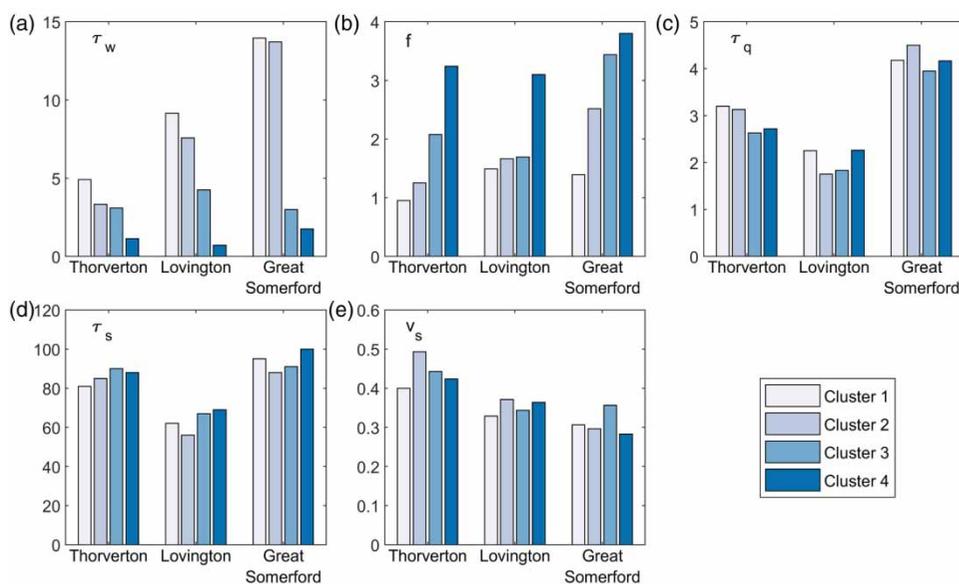

Figure 12 | Distribution of model parameters calibrated during the period 2003–2010 for the FCM_T calibration scheme (the average monthly temperature is lowest for one-month periods in Cluster 1 and highest for one-month periods in Cluster 4).

converting the effective rainfall to the streamflow, so the parameters τ_q, τ_s and v_s in this module are more related to the catchment characteristics. They are also associated with the rainfall characteristics, for instance, the higher intensity rainfall facilitates the quicker surface water flow. However, from Figure 7, the difference of the monthly rainfall is not significant among temperature-based clusters, so even though these parameters are associated with the rainfall characteristics, they do not show significant variations among temperature-based clusters. Based on the above results, it is inferred that after identifying the key factors that cause the change of the hydrological response in the catchment, the selection of the time-varying parameters should correspond to the identified key factors. In this study, the key factor controlling the change of the hydrological response is temperature, because the FCM_T calibration scheme presents the superior model performance for both calibration and validation periods than other calibration schemes. The parameters that are related to the effect of temperature in the IHACRES model are the parameter τ_w and f, so these two parameters vary significantly among clusters.

The effect of the cluster number

The FCM_T calibration scheme is used to explore the effect of the cluster number on model performances, owing to its superior performance in both calibration and validation periods than other schemes. The effect of the cluster number is based on the trade-off between the bias and variance, as shown in Figure 13 (Han 2011). If the number of clusters is too small, the classification may not be flexible
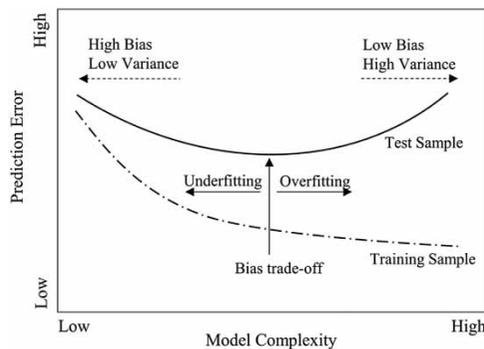

Figure 13 | Trade-off between the bias and the variance to explain the model overfitting and underfitting (Han 2011).

enough to recognize specific similarities, and the corresponding calibration scheme may have limitations to capture the variation of the hydrological response. In this case, underfitting will be caused, with high bias and low variance. On the other hand, if the number of clusters is too large, even the noise will be recognized, which will lead to overfitting, with low bias and high variance.

In order to avoid too many parameters, the FCM_T calibration scheme with the cluster number ranging from one to six is explored. Their model performance in terms of R^2 is compared in Figure 14. The typical trend of bias trade-off (Figure 13) can be found in Figure 14 for all catchments. The model performance for the calibration period improves with the increase of the cluster number. However, the increase of model complexity leads to overfitting and poor predictive capabilities of the model. The cluster number has an effect on the model performance; therefore, it is important to choose an appropriate cluster number to avoid both underfitting and overfitting. From Figure 14, the appropriate number of clusters is four for the study sites, where the model performance for both calibration period and validation period is the optimal.

DISCUSSION AND CONCLUSIONS

This study attempts to improve the hydrological model performance by using time-varying parameters to represent the variation of the hydrological response. However, allowing parameters to vary according to similarities of catchment conditions will increase the model complexity, which may lead to overfitting and affect the predictive capabilities of the model. Two issues of concern are the identification of clusters with similarities and the effect of the increased model complexity. In this study, four types of clusters are explored. Clusters based on the similarity of temperature and rainfall are identified using the Fuzzy C-means (FCM) algorithm. Clusters are also classified based on seasons and random numbers. The component of these clusters are the data of one-month periods which are divided according to the calendar month. From the distribution of different clusters, the FCM algorithm performs better in grouping objects with similar characteristics and separating objects that are dissimilar in terms of the same characteristics.
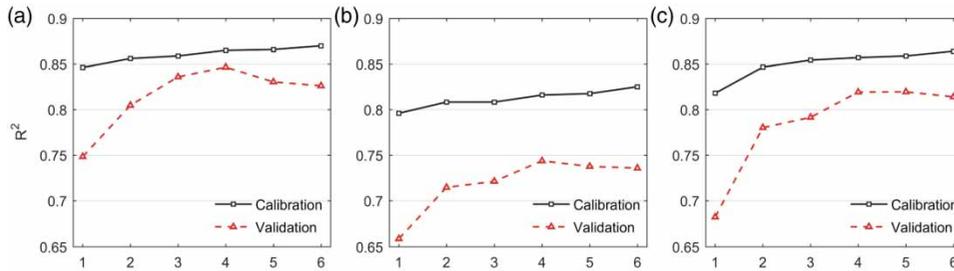
Figure 14 | Comparison of model performance (R^2) for FCM_T calibration schemes with cluster number from one to six: (a) Thorverton catchment, (b) Lovington catchment, (c) Great Somerford catchment.

The clusters identified using the clustering algorithm could better account for the inter-annual and intra-annual variation of temperature and rainfall, compared with the season-based approach. It is noted that when identifying the clusters we divided the data into multiple one-month periods according to the calendar month. For the choice of one-month periods, there is a problem of balance between the computational efficiency and model performance. For example, if daily data are used for clustering, it is possible that four consecutive days belong to four different clusters respectively. In this case, parameters vary more frequently and decrease the computational efficiency. In contrast, if the period length is too long, there may be different hydro-meteorological conditions during this period, and the difference cannot be identified when they belong to the same period. The reason for selecting one month as the period length is that the one-month period is often regarded as the minimum unit to describe the hydro-meteorological conditions in the previous studies (LéVesque *et al.* 2008; Luo *et al.* 2012). In order to directly compare with the season-based clusters, the one-month periods are divided according to the calendar month.

With the different types of clusters, parameters are calibrated by varying over clusters. For the purpose of comparison, the traditional calibration approach which uses time-invariant parameters is also explored. During the calibration procedure, the difference between the observed and simulated flow is minimized by maximizing the NSE. The performance of the calibrated model for the calibration period and validation period is evaluated with the use of criteria R^2, R^2_{ln} and relative bias to represent high flows, low flows and water balance respectively. The studied three catchments in southwest UK show similar results. The FCM_T calibration scheme provides a more accurate simulation in high and low flows and water balance for

both calibration period and validation period, followed by the Season calibration scheme. Given both temperature-based clusters and season-based clusters have capabilities of recognizing the temperature variation, it could be inferred that the temperature plays a crucial role in affecting the hydrological response in our study sites, and model performances could be improved by allowing parameter sets to vary according to temperature similarities. Through analysing the variation pattern of parameters in the FCM_T calibration scheme, it is found that the parameter τ_w and f in the non-linear loss module show distinct variation patterns against temperature-based clusters for all catchments, while there is no obvious variation pattern of the parameter τ_q, τ_s and v_s in the linear routing module. Given that the parameter τ_w and f are more related to the water loss process, which is highly associated with temperature, it could be concluded that after identifying the key factors that cause the change of the hydrological response in the catchment, the selection of the time-varying parameters should correspond to the identified key factors. This conclusion provides inspiration for applying time-varying parameters to more complicated models. Complicated hydrological models like the physically based models always involve multiple parameters and, for these models, it is not feasible to allow each parameter to vary among clusters because too many parameters may lead to overfitting and poor computational efficiency. In this case, the parameters that are more related to the changes of the catchment are selected to vary while the other parameters remain unchanged, which could avoid too large a number of parameters while taking into account the changes of the hydrological response.

We also explored the seasonal model performance of the FCM_T scheme and the Season scheme, the improvement is found at different extent for different seasons,

compared with the Tradition scheme. Although the model performance of the summer season is improved most significantly, the value of R^2 for the summer season is still low. This is because the calibrated model has poor performance in simulating low flows since the objective function NSE is more sensitive to high flows, and for the study sites, the streamflow in summer is much lower than other seasons. The use of the single objective of NSE also causes the fact that for all calibration schemes, the model performance in terms of R^2 is better than that of R^2_{ln}, indicating the model performs better in simulating high flows than low flows. Using the single objective optimization to calibrate the model is one limitation of our study, multiple objective optimization will be investigated in future work.

When allowing parameters to vary among clusters with similarities, the model complexity is highly correlated with the number of clusters, which raises the question of the trade-off between the model complexity and model performance. Through changing the number of clusters in the FCM_T calibration scheme, the effect of the cluster number on model performance is investigated. It is found that the model performance for the calibration period becomes better with the increase of the cluster number; however, the increase of model complexity leads to poor predictive capabilities of the model due to overfitting.

Overall, the main findings of this paper are as follows: among two hydro-meteorological factors: rainfall and temperature, temperature plays a more crucial role in controlling the change of the hydrological response in the study sites, so allowing parameters to vary among temperature-based clusters could improve the model performance. When using the time-varying parameters to account for the variation of the hydrological response, it is important to identify the key factors that cause the change of the hydrological response, and the selection of the time-varying parameters should correspond to the identified key factors. The clustering algorithm is an effective method to identify data with similarities of characteristics of interest. The number of clusters has an effect on model performance, therefore it is of great importance to select the appropriate cluster number to achieve a balance between the model complexity and model performance. In this study, the optimal performance for both the calibration period and the validation period is achieved when the cluster number is equal to four.

This study only used one hydrological model at three catchments, which really limits the generalization of conclusions. However, the methodology proposed in this study is generic and applicable to other catchments and hydrological models. We hope this paper will stimulate more studies to explore a variety of sites with different hydrological models using the proposed methodology to gain more knowledge about the variation of the hydrological response.

ACKNOWLEDGEMENTS

REFERENCES

Bezdek, J. C. 1981 *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Brigode, P., Oudin, L. & Perrin, C. 2013 Hydrological model parameter instability: a source of additional uncertainty in estimating the hydrological impacts of climate change? *J. Hydrol.* **476**, 410–425.

Choi, H. T. & Beven, K. 2007 Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework. *J. Hydrol.* **332** (3–4), 316–336.

Chou, C.-M. 2014 Random modeling of daily rainfall and runoff using a seasonal model and wavelet denoising. *Math. Prob. Eng.* **2014**, 1–11.

Croke, B. 2009 Representing uncertainty in objective functions: extension to include the influence of serial correlation. In: *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation*, pp. 3372–3378.

de Vos, N. J., Rientjes, T. H. M. & Gupta, H. V. 2010 Diagnostic evaluation of conceptual rainfall-runoff models using temporal clustering. *Hydrol. Process.* **24** (20), 2840–2850.

Duan, Q., Sorooshian, S. & Gupta, V. 1992 Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.* **28** (4), 1015–1031.

Efstratiadis, A., Nalbantis, I. & Koutsoyiannis, D. 2015 Hydrological modelling of temporally-varying catchments: facets of change and the value of information. *Hydrol. Sci. J.* **60** (7–8), 1438–1461.

Han, D. 2011 *Flood Risk Assessment and Management*. Bentham Science Publishers, Bussum, The Netherlands.

Her, Y. & Seong, C. 2018 Responses of hydrological model equifinality, uncertainty, and performance to multi-objective parameter calibration. *J. Hydroinform.* **20** (4), 864–885.

Jakeman, A. & Hornberger, G. 1993 How much complexity is warranted in a rainfall-runoff model? *Water Resour. Res.* **29** (8), 2637–2649.

Jakeman, A., Littlewood, I. & Whitehead, P. 1993 An assessment of the dynamic response characteristics of streamflow in the Balquhidder catchments. *J. Hydrol.* **145** (3–4), 337–355.

Kim, K. B. & Han, D. 2017 Exploration of sub-annual calibration schemes of hydrological models. *Hydrol. Res.* **48** (4), 1014–1031.

Kim, K. B., Kwon, H.-H. & Han, D. 2016 Hydrological modelling under climate change considering nonstationarity and seasonal effects. *Hydrol. Res.* **47** (2), 260–273.

Kim, H. S. & Lee, S. 2014 Assessment of a seasonal calibration technique using multiple objectives in rainfall-runoff analysis. *Hydrol. Process.* **28** (4), 2159–2173.

KlemeŠ, V. 1986 Operational testing of hydrological simulation models. *Hydrol. Sci. J.* **31** (1), 13–24.

Krause, P., Boyle, D. & Bäse, F. 2005 Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.* **5**, 89–97.

Letcher, R., Schreider, S. Y., Jakeman, A., Neal, B. & Nathan, R. J. 2001 Methods for the analysis of trends in streamflow response due to changes in catchment condition. *Environmetrics* **12** (7), 613–630.

LéVesque, É., Anctil, F., Van Griensven, A. N. N. & Beauchamp, N. 2008 Evaluation of streamflow simulation by SWAT model for two small watersheds under snowmelt and rainfall. *Hydrol. Sci. J.* **53** (5), 961–976.

Li, X., Sha, J., Li, Y.-M. & Wang, Z.-L. 2018 Comparison of hybrid models for daily streamflow prediction in a forested basin. *J. Hydroinform.* **20** (1), 191–205.

Luo, J., Wang, E., Shen, S., Zheng, H. & Zhang, Y. 2012 Effects of conditional parameterization on performance of rainfall-runoff model regarding hydrologic non-stationarity. *Hydrol. Process.* **26** (26), 3953–3961.

Magnusson, J., Wever, N., Essery, R., Helbig, N., Winstral, A. & Jonas, T. 2015 Evaluating snow models with varying process representations for hydrological applications. *Water Resour. Res.* **51** (4), 2707–2723.

Melsen, L., Teuling, A., Torfs, P., Zappa, M., Mizukami, N., Clark, M. & Uijlenhoet, R. 2016 Representation of spatial and temporal variability in large-domain hydrological models:

case study for a mesoscale pre-Alpine basin. *Hydrol. Earth Syst. Sci.* **20** (6), 2207–2226.

Merz, R., Parajka, J. & Blöschl, G. 2011 Time stability of catchment model parameters: implications for climate impact analyses. *Water Resour. Res.* **47** (2), W02531.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D. & Veith, T. L. 2007 Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* **50** (3), 885–900.

Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models part I – a discussion of principles. *J. Hydrol.* **10** (3), 282–290.

Paik, K., Kim, J. H., Kim, H. S. & Lee, D. R. 2005 A conceptual rainfall-runoff model considering seasonal variation. *Hydrol. Process.* **19** (19), 3837–3850.

Pakhira, M. K., Bandyopadhyay, S. & Maulik, U. 2004 Validity index for crisp and fuzzy clusters. *Pattern Recognit.* **37** (3), 487–501.

Pathiraja, S., Marshall, L., Sharma, A. & Moradkhani, H. 2016a Detecting non-stationary hydrologic model parameters in a paired catchment system using data assimilation. *Adv. Water Resour.* **94**, 103–119.

Pathiraja, S., Marshall, L., Sharma, A. & Moradkhani, H. 2016b Hydrologic modeling in dynamic catchments: a data assimilation approach. *Water Resour. Res.* **52** (5), 3350–3372.

Sadegh, M., AghaKouchak, A., Flores, A., Mallakpour, I. & Nikoo, M. R. 2019 A multi-model nonstationary rainfall-runoff modeling framework: analysis and toolbox. *Water Resour. Manage.* **33** (9), 3011–3024.

Sbai, E. 2001 Cluster analysis by adaptive rank-order filters. *Pattern Recognit.* **34** (10), 2015–2027.

Tanguy, M., Dixon, H., Prosdocimi, I., Morris, D. G. & Keller, V. D. J. 2016 Gridded estimates of daily and monthly areal rainfall for the United Kingdom (1890–2015) [CEH-GEAR]. NERC Environmental Information Data Centre.

Toth, E. 2009 Classification of hydro-meteorological conditions and multiple artificial neural networks for streamflow forecasting. *Hydrol. Earth Syst. Sci.* **13** (9), 1555–1566.

Wang, W.-C., Cheng, C.-T., Chau, K.-W. & Xu, D.-M. 2012 Calibration of Xinanjiang model parameters using hybrid genetic algorithm based fuzzy optimal model. *J. Hydroinform* **14** (3), 784–799.

Zhang, R., Moreira, M. & Corte-Real, J. 2016 Multi-objective calibration of the physically based, spatially distributed SHETRAN hydrological model. *J. Hydroinform.* **18** (3), 428–445.

Zhang, R., Cuartas, L. A., de Castro Carvalho, L. V., Reis Deusdará Leal, K., Mendiondo, E. M., Abe, N., Birkinshaw, S., Samprogna Mohor, G., Seluchi, M. E. & Nobre, C. A. 2018a Season-based rainfall-runoff modelling using the probability-distributed model (PDM) for large basins in southeastern Brazil. *Hydrol. Process.* **32** (14), 2217–2230.

Zhang, R., Liu, J., Gao, H. & Mao, G. 2018b Can multi-objective calibration of streamflow guarantee better hydrological model accuracy? *J. Hydroinform.* **20** (3), 687–698.