Hydrol. Earth Syst. Sci., 21, 251–265, 2017
www.hydrol-earth-syst-sci.net/21/251/2017/
doi:10.5194/hess-21-251-2017

Hydrology and Earth System Sciences
Open Access

Physically based distributed hydrological model calibration based on a short period of streamflow data: case studies in four Chinese basins

Wenchao Sun[1,2], Yuanyuan Wang[1,2], Guoqiang Wang[1,2], Xingqi Cui[1,2], Jingshan Yu[1], Depeng Zuo[1,2], and Zongxue Xu[1,2]

[1]College of Water Sciences, Beijing Normal University, Xinjiekouwai Street 19, Beijing 100875, China
[2]Joint Center for Global Change Studies (JCGCS), Beijing 100875, China

Correspondence to: Jingshan Yu (jingshan@bnu.edu.cn)

Received: 25 April 2016 – Published in Hydrol. Earth Syst. Sci. Discuss.: 17 May 2016
Revised: 6 December 2016 – Accepted: 7 December 2016 – Published: 11 January 2017

Abstract. Physically based distributed hydrological models are widely used for hydrological simulations in various environments. As with conceptual models, they are limited in data-sparse basins by the lack of streamflow data for calibration. Short periods of observational data (less than 1 year) may be obtained from fragmentary historical records of previously existing gauging stations or from temporary gauging during field surveys, which might be of value for model calibration. However, unlike lumped conceptual models, such an approach has not been explored sufficiently for physically based distributed models. This study explored how the use of limited continuous daily streamflow data might support the application of a physically based distributed model in data-sparse basins. The influence of the length of the observation period on the calibration of the widely applied soil and water assessment tool model was evaluated in four Chinese basins with differing climatic and geophysical characteristics. The evaluations were conducted by comparing calibrations based on short periods of data with calibrations based on data from a 3-year period, which were treated as benchmark calibrations of the four basins, respectively. To ensure the differences in the model simulations solely come from differences in the calibration data, the generalized likelihood uncertainty analysis scheme was employed for the automatic calibration and uncertainty analysis. In the four basins, contrary to the common understanding of the need for observations over a period of several years, data records with lengths of less than 1 year were shown to calibrate the model effectively, i.e., performances similar to the benchmark calibrations were achieved. The models of the wet Jinjiang and Donghe basins could be effectively calibrated using a shorter data record (1 month), compared with the dry Heihe and upstream Yalongjiang basins (6 months). Even though the four basins are very different, when using 1-year or 6-month (covering a whole dry season or rainy season) data, the results show that data from wet seasons and wet years are generally more reliable than data from dry seasons and dry years, especially for the two dry basins. The results demonstrated that this idea could be a promising approach to the problem of calibration of physically based distributed hydrological models in data-sparse basins, and findings from the discussion in this study are valuable for assessing the effectiveness of short-period data for model calibration in real-world applications.

1 Introduction

Globally, flood and droughts are the two most prevalent natural disasters, considered to have affected 140 million people annually, on average, between 2005 and 2014 (United Nations Offices for Disaster Risk Reduction, 2016). Mitigating the possible damages associated with these disasters relies on precise forecasting in terms of timing and scale (Callahan et al., 1999; McEnery et al., 2005). Hydrological models are tools commonly used for simulating the water cycle at basin scale and for predicting streamflow at the basin outlet, which represents the integrated output of all the hydrological processes within a basin. Many parameters of hydrological models are conceptual without explicit physical mean-

Published by Copernicus Publications on behalf of the European Geosciences Union.

ing, which makes it necessary to identify parameter values through model calibration based on streamflow data (Gupta et al., 2005). For physically based models, although values of parameters with explicit physical meaning can be measured, the scale of measurement and model simulation is different, which makes it difficult to apply measured values to hydrological models directly. Also, these measurements require intensive field surveys, which are not available in most studies. Therefore, parameters of such model are usually also obtained from model calibration based on streamflow data. However, because of resource constraints (e.g., financial and human resources), there has been a general decline in the networks designed to monitor streamflow (Wohl et al., 2012), especially in developing countries, which has become a major obstacle to the applications of hydrological models in basins where streamflow data are sparse (Hrachowitz et al., 2013).

The usual approach regarding data-sparse basins is regionalization, which estimates model parameters using information from similar gauged basins. One major concern with regionalization is prediction uncertainty, which is determined by the degree of similarity and by the method chosen to describe the similarity (Sivapalan, 2003). To reduce the uncertainty introduced by regionalization, many researchers have tried to improve parameter estimation by introducing limited information from ungauged basins. For example, Viviroli and Seibert (2015) combined short-term streamflow observations with parameter regionalization and showed that parameter identifications could be improved compared with using information only from donor basins. Many recent works have tried to use available in situ or remote sensing observations of hydrological processes other than streamflow for model calibration, e.g., soil moisture (e.g., Silvestro et al., 2015; Vrugt et al., 2002), evapotranspiration (Vervoort et al., 2014; Winsemius et al., 2008), and groundwater level (e.g., Khu et al., 2008), as a new direction to solve the calibration problem. These studies have shown promising performances for identifying parameters that describe the processes being measured. However, none of these observations have the similar capability to streamflow data for constraining hydrological model parameters. Another appealing approach is the use of river water surface area, width, or stage derived from remote sensing as a surrogate of streamflow for model calibration (e.g., Revilla-Romero et al., 2015; Sun et al., 2015; Getirana, 2010); however, such an approach depends on the availability of effective satellite observations. Furthermore, the reported higher simulation uncertainty in comparison with calibration based on streamflow data is another concern (Sun et al., 2010, 2012).

From the above, it is clear that streamflow observations play a critical role in identifying hydrological model parameters. For an ungauged basin, although a long time series of observations is unavailable, short-period records of streamflow or occasional observations from field surveys might be obtainable. If such data are to be used for calibration, it is important to know how many observations are needed to calibrate model parameter. It is usually suggested that streamflow records covering several years are necessary (Yapo et al., 1996); however, several researchers have attempted to challenge this common understanding using discontinuous or short-period records of less than 1 year in basins within different climatic regions (e.g., Perrin et al., 2007; Kim and Kaluarachchi, 2009; Seibert and Beven, 2009; Tada and Beven, 2012). For conceptual models, these studies indicated that with observations of the order of several scores, reasonable parameter estimates could be derived. And model performance similar to those obtained from calibrations using records covering several years could be obtained, highlighting the possibility that calibration with limited numbers of observations is a promising alternative to the classical regionalization approach. For hydrological simulations or predictions in changing environments, when the model is expected to evaluate influences of change in climate or the basin's physical characteristics to the water cycle, physically based distributed hydrological models are usually preferred, because of their better description of the spatial heterogeneity and details of the water cycle at the basin scale (Finger et al., 2012; Wu and Liu, 2012). However, the use of limited observations to address the calibration problem of such models in ungauged basins has rarely been discussed in the literature, probably because of the complexity of model structures and the corresponding considerable demands for computation time.

The objective of this study is to explore how short-period observations of daily continuous streamflow might support the calibration of a physically based distributed model in data-sparse basins. In the real world, such observations might be obtained from fragmentary historical records of previously existing gauging stations or from short-period field surveys. The commonly used soil and water assessment tool (SWAT) model was adopted for the investigation. Previous research has shown that the requirements of calibration data differ significantly among basins (e.g., Lidén and Harlin, 2000). Therefore, we selected four basins with different climatic conditions (two in humid regions and the other two in dry regions) to improve the generality of our findings. The evaluation relies on comparison with the conventional calibration using observations covering several years, which was adopted as the benchmark calibration. The evaluation requires an objective calibration and uncertainty analysis framework to ensure the differences among the calibration results derived solely from the differences in the observations. Considering this issue, the generalized likelihood uncertainty estimation (GLUE) (Beven and Binley, 1992; Freer and Beven, 1996) method was used for the model calibration, for which all the settings during the calibration were verifiable and satisfied the requirements of the evaluation. By reducing the number of observations used in the model calibration in a designed manner and by comparing each with the benchmark calibration, the influence of the length of observational records on the calibration could be analyzed and the feasibility of using limited data discussed.

2 Materials and method

2.1 Hydrological model

SWAT is a popular physically based distributed hydrological model developed by the United States Department of Agriculture. It operates on a daily-time step, and it is capable of simulating the water cycle and transportation of sediment and pollutants at the basin scale. The model is fully integrated with the geographic information system (GIS). Based on a river network derived from a digital elevation model, the study basin can be discretized into many subbasins. Moreover, based on GIS data of the soil type and land cover, each subbasin can be separated into several unique hydrological response units for describing the heterogeneity in runoff generation. The hydrological processes considered in the model include precipitation, interception, infiltration, evapotranspiration, snowmelt, surface runoff, percolation, baseflow, and flow movement in river channels. Because of the complex model structure, many parameters need to be identified via calibration. Further details about the SWAT model are available in Arnold et al. (1998) and Gassman et al. (2007).

2.2 Calibration and uncertainty analysis method

Considering the objective of this study, manual calibration is not feasible for the comparison of calibrations because it relies on subjective judgments about model performance (Madsen, 2003). Therefore, an automatic calibration procedure that optimizes an objective function by searching parameter spaces to find combinations reflecting the characteristics of the target basin was required (Muleta and Nicklow, 2005). Another concern is the phenomenon of equifinality (Beven, 2001) – that many very different parameter sets might exhibit similar performances. Thus, it is necessary to quantify the uncertainty introduced by equifinality for the evaluation. Here, the GLUE method was employed as the automatic calibration and uncertainty analysis scheme. It was integrated into the SWAT model in the calibration package SWAT-CUP (SWAT calibration uncertainty procedures) (Yang et al., 2008). To describe the equifinality in a quantitative manner, it regards all those parameter sets performing better than a predefined threshold as behavioral parameter sets, for which the corresponding simulations, with weights assigned based on performance, are then used to produce an ensemble simulation. Several subjective choices must be made when using the GLUE method, but they are made explicitly and they can be examined at any time (Beven and Binley, 1992). For different calibrations, if all subjective settings except the calibration data remain the same, the GLUE method can ensure that differences in the calibration results derive purely from the different observations used in the calibration, which is ideal for the comparison needed for the evaluation of this study. Here, the procedure for the implementation of the GLUE method was as follows:

1. Generate random parameter sets. Usually, the prior information about parameter distributions is unknown, and therefore assuming uniform distributions is reasonable (Beven and Freer, 2001). Then, the Latin hypercube sampling scheme is adopted to generate parameter sets randomly from parameter space.

2. Select behavioral parameter sets. A likelihood measure is defined to quantify the degree of goodness with which each parameter set can reproduce the observations. Then, based on a threshold set by the modeler, the good parameter sets (named behavioral parameter sets) are selected. Here, the Nash–Sutcliffe efficiency (NSE) was used as the likelihood measure:

$$\mathrm{NSE} = 1 - \frac{\sum(Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i})}{\sum(Q_{\mathrm{obs},i} - Q_{\mathrm{obs,avg}})}, \tag{1}$$

where $Q_{\mathrm{obs},i}$ ($\mathrm{m^3\,s^{-1}}$) and $Q_{\mathrm{sim},i}$ ($\mathrm{m^3\,s^{-1}}$) represent the observed and simulated streamflow, respectively, at time step i, and $Q_{\mathrm{obs,avg}}$ ($\mathrm{m^3\,s^{-1}}$) is the average value of the streamflow observations.

3. Calculate the behavioral parameter sets' posterior likelihood. Every identified behavioral set is included to make an ensemble simulation. The posterior likelihood of each set (i.e., the weight of the streamflow simulation of each behavioral parameter set in the ensemble simulation) is computed based on the Bayes equation:

$$L_p[\theta|Q_{\mathrm{obs}}] = CL[\theta|Q_{\mathrm{obs}}]L_o[\theta], \tag{2}$$

where $L_o[\theta]$ is the prior likelihood of parameter set θ, under the assumption of a uniform prior distribution (which is the same value for all sets), $L[\theta|Q_{\mathrm{obs}}]$ is the NSE that quantifies the performance of reproducing Q_{obs}, and C is a scaling factor that makes unity the sum of posterior likelihood for all behavioral parameter sets.

4. Make an ensemble prediction. At each time step t, the cumulative distribution of the simulation is calculated:

$$P_t(Q_t < q) = \sum_{i=1}^{m} L_p[\theta_i|Q_{t,i} < q], \tag{3}$$

where $P(Q_t < q)$ is the cumulative probability of the simulated streamflow Q_t less than an arbitrary value q, $L_p[\theta_i]$ is the posterior likelihood of set θ_i (for which the simulated streamflow is less than q) and m is the amount of the parameter sets that satisfy the condition of $Q_{t,i} < q$. The streamflow corresponding to the lower 2.5 % and upper 97.5 % quantiles of the posterior distribution at each time step consists of the lower and upper

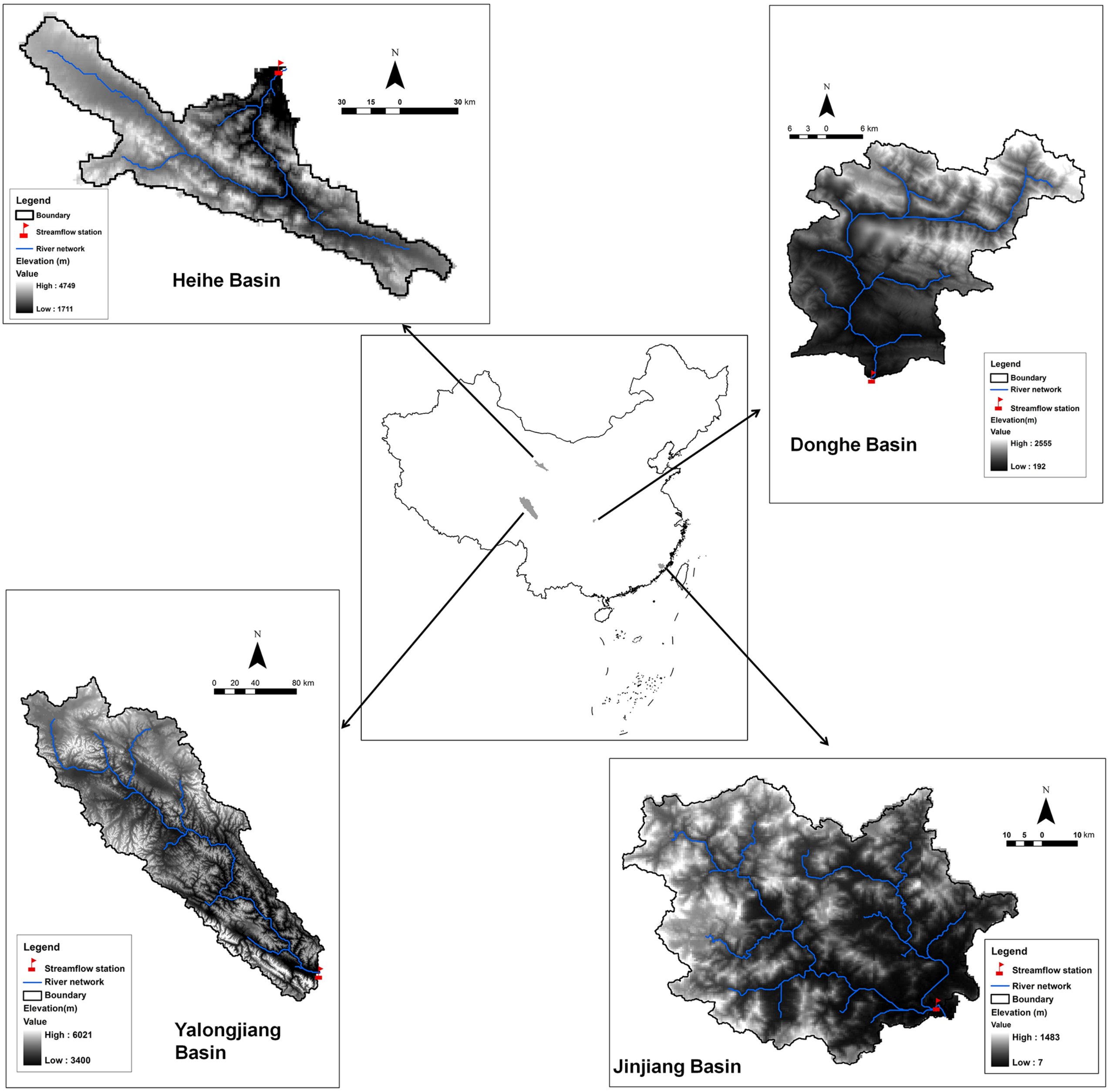

Figure 1. Topography, river networks, and the streamflow gauging stations of the four basins and their locations in China.

limits of the ensemble simulation, respectively. The predicted streamflow corresponding to the best performing parameter set (judged from likelihood) is treated as the best estimate of streamflow.

2.3 Study basins

We used four basins located in China (Fig. 1) to test the method. The basins are spread over the country to ensure that various hydrological, climatic, and geophysical conditions are included in our study. They are located in different climatic regions and characteristics of topography, annual precipitation, and temperature are quite different.

The Jinjiang Basin is located on the west coast of the Taiwan Straits in Fujian Province, China. The area of the basin is 5629 km^2. The river system has two major tributaries that flow from mountainous area of the north, join at Shuangxikou, and then flow to the low plain region in the southeast (elevation ranges from 50 to 1366 m). The dominant land covers are forest and crop land, and the main soil types are paddy soil, red soil, and yellow soil. The basin is in a subtropical marine monsoon climatic region, with

Table 1. Main characteristics of the four basins being studied.

Basin	Streamflow station	Area (km^2)	Climate	Annual rainfall (mm)	Annual average temperature (°)	Ranges of elevation (m)
Jinjiang	Shilong	5629	Subtropical marine monsoon climate	1651	20	50–1366
Donghe	Wenquan	1089	Subtropical monsoon climate	1247	18	192–2569
Heihe	Yingluoxia	8843	Continental monsoon climate	423	6	1711–4749
Yalongjiang	Ganzi	32 535	Continental plateau climate	520	8	3400–6021

Table 2. The calibration and validation periods for the benchmark calibrations of the four basins.

Basin	Calibration period	Validation period
Jinjiang	2005–2007	2008–2009
Donghe	2002–2004	2005–2006
Heihe	2003–2005	2006–2008
Yalongjiang	2005–2007	2008–2010

warm dry winters and hot rainy summers. Annual precipitation ranges from 1000 to 1800 mm, most of which falls in summer. The hydrological modeling was conducted for the upstream area of the Shilong gauging station.

The Donghe River is one of the major tributaries of the Pengxi River in the upstream region of the Three Gorges Reservoir. The length of the mainstream is about 106 km, and the drainage area is 1089 km^2. The main land covers are cropland, shrub, and pasture, and the main soil types are flat stone yellow sandy soil and lime yellow clay. The basin is in a warm wet subtropical monsoon climate region. Annual precipitation ranges from 1100 to 1500 mm. Most of precipitation falls in summer. The hydrological model was calibrated and validated by the streamflow data in Wenquan gauging station.

The Heihe Basin is in the arid northwest of China. It is the second-largest inland basin in China with an area of about 128 900 km^2. From the southern mountainous region to the northern high-plain area, the elevation decreases from about 5000 to 1000 m. The hydrological simulation was executed for the upstream mountainous region of Yingluoya gauging station, encompassing an area of around 8843 km^2. The elevation of the study area varies from around 4700 m in the headwater region to around 1700 m at Yingluoya station. The primary land cover types are forest, grassland, and Gobi, and alpine meadow soil and frost desert soil occupy more than 74 % of the basin area. The region has an inland continental climate with cold dry winters and hot summers, and average annual precipitation of around 400 mm.

The Yalongjiang River originates in the Tibetan plateau, which is the largest tributary to the Jinshajiang River in the upper Yangtze River. The hydrological modeling was conducted in the upstream region of the Ganzi streamflow station, the elevation of which ranges from 3400 to 6021. The area of hydrological simulation by SWAT is 32 535 km^2. Plateau meadow soil is the main soil type, and the shrub meadow is the main land-cover type. This basin has a continental plateau climate. Average annual precipitation in a recent period of 50 years is about 520 mm, 73 % percentage of which concentrated in June to September. A long, cold winter and a cool, wet summer exists in this basin, with strong radiation all year round.

In the four basins, most precipitation occurs in summer. Based on annual precipitation, the four basins were divided into two groups. The Jinjiang and Donghe basins are considered as representatives of wet basins, and Heihe and Yalongjiang basins represent dry basins. For each basin, the influence of the observational record length on the calibration will be explored. Then, discussion about the differences among the four basins will be performed to obtain more general insights. The characteristics of the four basins are shown in Table 1. The diversity among the basins is very helpful for making relatively general conclusions from the findings of this study.

2.4 Experiment design

Considering the availability of streamflow records, the benchmark calibration for the Jinjiang Basin was made based on full daily observations for 2005–2007 and it was validated using data for 2008–2009. For the other three basins, the benchmark calibrations were also conducted using 3-year continuous daily streamflow observations and the models were validated using 2- or 3-year streamflow data, based on data availability. The details about the calibration and validation periods for the benchmark calibrations conducted in the four basins are shown in Table 2. As an initial trial for showing the potential of the method for distributed models, we sought to explore whether there are records of certain short length or certain number of continuous daily observations achieving a similar performance to benchmark calibration, not to determine whether all records of that specific length can calibrate the model effectively. Simulations by distributed models are time-consuming and the calibration using the GLUE method requires models to be run a large number of times. Therefore, it was hard to follow the studies of conceptual models (e.g., Perrin et al., 2007; Seibert and Beven, 2009) that could conduct calibrations many times. Considering the two above-mentioned issues, to perform the

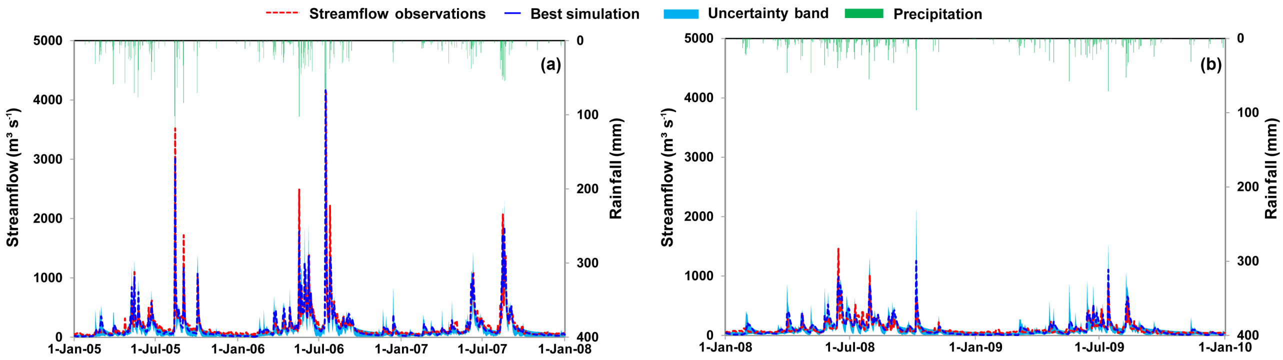

Figure 2. Simulated streamflow for the benchmark calibration of the Jinjiang Basin in both the calibration (2005–2007) and validation (2008–2009) periods.

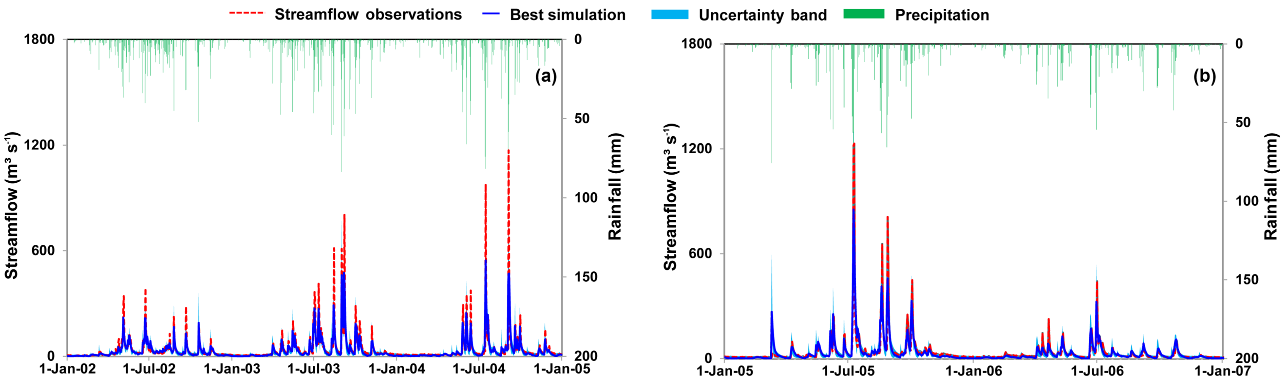

Figure 3. Simulated streamflow for the benchmark calibration of the Donghe Basin in both the calibration (2002–2004) and validation (2005–2006) periods.

calibration in manageable times, the experiment of calibration using short-period records, which are subsets of the calibration data of the benchmark calibration, was conducted in two stages. In the first stage, three calibrations using a 1-year data record that covered both the rainy and dry seasons, and five calibrations using a 6-month data record that covered either a rainy season or a dry season were undertaken. The short periods for which corresponding data were used for the calibrations in the first stage are listed in Table 3. If there are calibrations using 6-month data that could achieve performances similar to the benchmark calibration, stage two of the experiment was initialized, in which the subsets of 6-month data records were used for calibration to explore the performance of a calibration period shorter than six months. Kim and Kaluarachchi (2009) and Yapo et al. (1996) showed that data from high-flow periods are more informative than data from low-flow periods for model calibration. As our study explored the possibility of the highest performance of certain lengths of records for calibration, the 3-month, 1-month data and 1-week datasets with highest average streamflow in the 6-month records were employed as the representatives to calibrate the model and conduct the evaluation at these three temporal scales.

Perrin et al. (2007) showed that model performance in the calibration period could be very good when using very limited numbers of observations, because it is easy to fill only a small number of points in the hydrograph. Conversely, the performance in the validation period could be very poor, because there are no observations to constrain the model simulation. Therefore, the evaluation of limited numbers of streamflow data needs to consider the performance in both calibration and validation periods, mostly in the validation period. For each basin, in order to compare model performance of calibration using short-period data with the benchmark calibration in an objective manner, the validation periods of these calibrations were made the same as the benchmark calibrations. The evaluation of each calibration was performed in terms of the aspects of general performance and simulation uncertainty. The general performance was represented by the NSE of the best behavioral parameters set (i.e., the one with the highest likelihood value identified by the calibration data) for the calibration and validation periods. Two indexes were utilized to assess the simulation uncertainty:

Table 3. Short periods for which corresponding data were used for the calibrations at stage one of the evaluation.

Length of the period	Jinjiang Basin	Donghe Basin	Heihe Basin	Yalongjiang Basin
1 year	2005 2006 2007	2002 2003 2004	2003 2004 2005	2005 2006 2007
6 months	Apr to Sep 2005 Oct 2005 to Mar 2006 Apr to Sep 2006 Oct 2006 to Mar 2007 Apr to Sep 2007	Apr to Sep 2002 Oct 2002 to Mar 2003 Apr to Sep 2003 Oct 2003 to Mar 2004 Apr to Sep 2004	Apr to Sep 2003 Oct 2003 to Mar 2004 Apr to Sep 2004 Oct 2004 to Mar 2005 Apr to Sep 2005	Apr to Sep 2005 Oct 2005 to Mar 2006 Apr to Sep 2006 Oct 2006 to Mar 2007 Apr to Sep 2007

The P_factor is the percentage of observations embraced by the 95 % prediction intervals. The R_factor is a measure of the average width of 95 % simulation intervals

$$\text{R\_factor} = \frac{\sum_{i=1}^{m} (Q_{97.5\,\%,i} - Q_{2.5\,\%,i})}{m \times \sigma_{Q_{\text{obs}}}}, \tag{4}$$

where $Q_{97.5\,\%,i}$ and $Q_{2.5\,\%,i}$ represent the 97.5 and 2.5 % quantiles of the simulated streamflow at time step i, respectively, m is the total time step of the simulation, and $\sigma_{Q_{\text{obs}}}$ is the standard deviation of the streamflow observations. A low value of R_factor combined with a high value of the P_factor indicates low simulation uncertainty. For the evaluation, we put more weight on NSE and P_factor, as they are important and explicit for judging the model performance. After NSE and P_factor, then R_factor will be considered and less weight is applied to it.

3 Results and discussion

3.1 Performances of benchmark calibrations

Before we can apply the model for evaluating the method proposed in this study, the model robustness in the four basins must be examined, through assessing model performance corresponding to the benchmark calibrations. Ten commonly calibrated SWAT parameters from the literature were selected for the automatic calibration using GLUE, and their prior ranges were set based on the recommendation from SWAT-CUP. The parameters and their prior ranges (Table 4) were the same for all calibrations, to exclude the influence of parameter uncertainty and ease the calibration comparisons. For each calibration, 10 000 parameter sets were generated randomly using the Latin hypercubic sampling method to run the GLUE scheme. For the Heihe Basin, the threshold for likelihood was set to 0.5. For the Jinjiang Basin, too many parameter sets could result in reasonable simulations, and therefore the threshold for likelihood was set to 0.7. The threshold for Donghe Basin and Yalongjiang Basin was 0.5 and 0.4, respectively. For the benchmark calibrations of the four basins, the results of the calibration are summarized in Table 5, and the best simulations and uncertainty bands of the ensemble simulation are shown in Figs. 2–5. The NSEs of the best simulation in four cases were satisfactory and they could reproduce the observed hydrographs well. Furthermore, the uncertainty bands covered most of the observations. All these facts indicate that the model applications in the four basins were successful. The results of these calibrations were treated as the benchmarks for each basin. The only difference between the benchmark calibrations and the other calibrations was the calibration data, which were therefore the only cause of the differences in the calibration results.

3.2 Evaluation of the Jinjiang Basin case

The performances of the ensemble simulations corresponding to the 1-year and 6-month calibration datasets are shown in Fig. 6. For the 1-year period, all three calibrations performed similarly to the benchmark calibration. Figure 7a presents the cumulative distribution of the available annual streamflow for Shilong station from 1958 to 2009. For the three years that streamflow data was used for benchmark calibration, 2006 is a very wet year and 2005 and 2007 are normal-to-wet years. The two years of 2008 and 2009 are both dry years. For the benchmark calibration, the model performance in the validation period decreases, compared with the calibration period. The decrease in model performance is consistent with other studies (e.g., Todorovic and Plavsic, 2016), in which model efficiency also decreases if the calibration period is wetter than the validation period. When using only 1-year data for calibration, the performances in the validation period are similar to the benchmark calibration. On the 6-month timescale, the corresponding five calibrations exhibited considerable differences: no parameter sets were identified as behavioral parameter sets when using calibration data for the period October 2006 to March 2007, indicating that no parameter sets could capture the characteristics of the hydrological processes of that period. The other

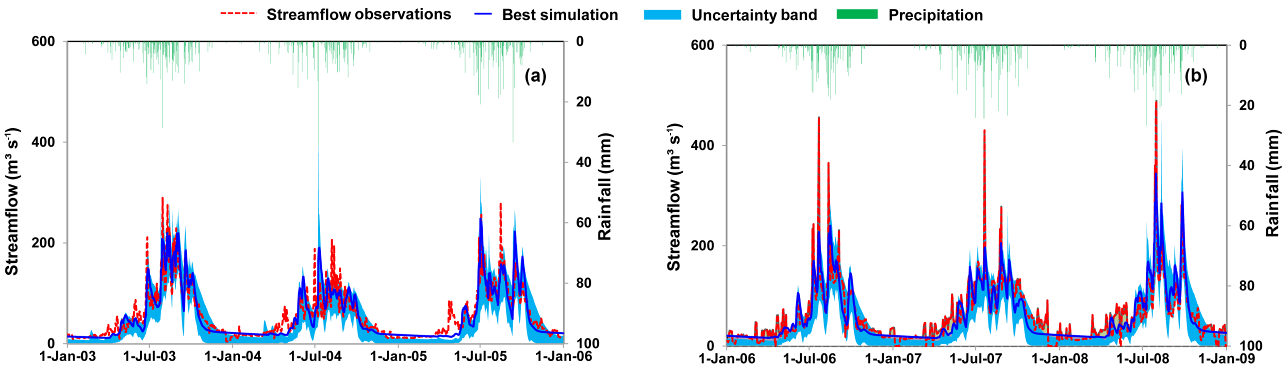

Figure 4. Simulated streamflow for the benchmark calibration of the Heihe Basin in both the calibration (2003–2005) and validation (2006–2008) periods.

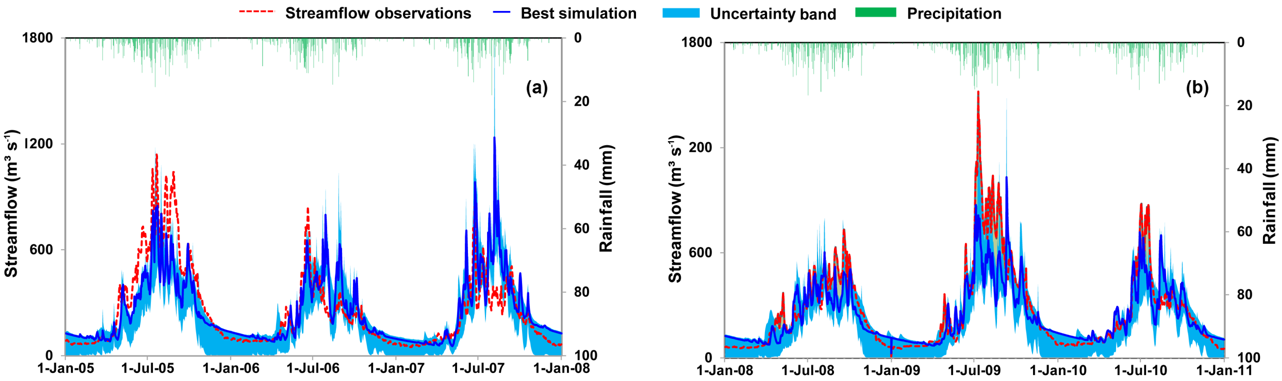

Figure 5. Simulated streamflow for the benchmark calibration of the Yalongjiang Basin in both the calibration (2005–2007) and validation (2008–2010) periods.

four records of 6 month could achieve a similar performance to the benchmark calibration. The second stage of the experiment was undertaken using 3-month (June–August), 1-month (July), and 1-week (14–20 July) datasets, with the highest streamflow during April to September 2006. Figure 6 shows that when calibrating the SWAT model using the 1-week dataset, the uncertainty increased and the NSE decreased distinctly in the validation period compared with the benchmark calibration. The calibration using the 1-month dataset still achieved a similar performance to the benchmark calibration, judging from the indexes. Figure 8 shows the simulated streamflow of best-performing parameter sets corresponding to the benchmark calibration, the calibration using the 1-month data and the 1-week data. The difference of simulations between the benchmark calibration and calibration using 1-month data is minor. But the difference between the benchmark calibration and calibration using 1-week data is obvious. The latter seems to fail to reproduce streamflow in low flow period, indicating the information content in the observations is not sufficient for model calibration. In summary, it is indicated that in the Jinjiang Basin, it is possible that 1 month's continuous daily observations can contain much of the information content in the 3-year continuous streamflow data for model calibration.

3.3 Evaluation of the Donghe Basin case

Figure 9 describes the general performance and simulation uncertainty of calibration using 1-year and 6-month data in the Donghe Basin. As long time series of annual streamflow is unavailable, we cannot judge the frequency of annual streamflow in the 5 years being simulated. Streamflow at the basin outlet is generated by precipitation within the basin. There is a close relationship between them. Based on this understanding, we use the annual precipitation frequency derived from a national climatic dataset with spatial resolution of 5 km, which was developed by the Land–Atmosphere Interaction Research Group at Beijing Normal University (available at: http://globalchange.bnu.edu.cn/research/forcing), as a surrogate of streamflow frequency to infer whether each year is a wet year or a dry year. For the 3 calibration years, as shown in Fig. 7b, 2002 is a dry year and 2003 and 2004 are wet years. Of the two validation years, 2005 and 2006 are a wet year and extremely dry year, respectively. No matter whether

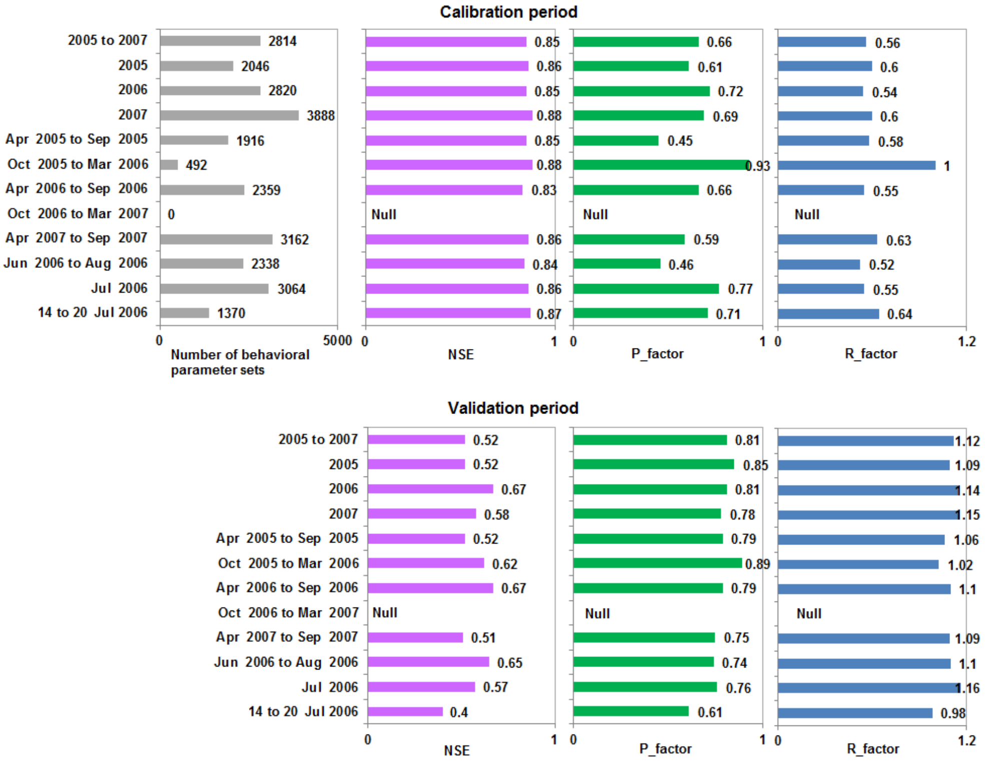

Figure 6. Model performance for the calibrations using short-period data in Jinjiang Basin.

Table 4. SWAT model parameters being calibrated.

Name	Description	Initial range
CN2	SCS runoff curve number	20–90
EPCO	Plant uptake compensation factor	0.01–1
GW_DELAY	Groundwater delay time (days)	30–450
SLSUBBSN	Average slope length (m)	10–150
ESCO	Soil evaporation compensation coefficient	0.8–1
ALPHA_BF	Baseflow recession coefficient	0–1
OV_N	Manning coefficient for overland flow	0–0.8
CH_K2	Hydraulic conductivity in main channel (mm h^{-1})	5–130
SOL_AWC	Available soil water capacity (mm H_2O mm $Soil^{-1}$)	0–1
SOL_K	Soil saturated hydraulic conductivity (mm h^{-1})	0–2000

the 1-year data from a dry year (2002 or 2004) or a wet year (2003) was used, the streamflow in the validation period are all reproduced well. Also, all of the calibrations using 6-month data, either from rainy seasons or dry seasons, achieve similar performances to the benchmark calibration. Like in the case of Jinjiang Basin, if some parameter sets are identified as behavioral ones using short-period data of 1 year and 6 months, their performances in the validation period can resemble the benchmark calibration. Stage two of the evaluation was carried out using the data of July to September 2003, September 2003, and 1 to 7 July 2003 as the representatives of 3-month, 1-month and 1-week data. There is no parameter set that could reach the threshold of likelihood when using the 1-week record. For the other two calibrations, they can all work as well as the benchmark calibration. Like in the Jinjiang Basin, 1-month data could also calibrate the model successfully in this basin.

3.4 Evaluation of the Heihe Basin case

The results of the calibration are shown in Fig. 10. The calibrations using the 1-year datasets of 2003 and 2005 achieved almost the same performance as the benchmark calibration. For the calibration using data from 2004, the number of identified behavioral parameter sets decreased significantly and the NSE of the best simulation in the validation period de-

Table 5. Model performance for the benchmark calibration in the four basins.

	Number of behavioral parameter sets	NSE		P_factor		R_factor	
		Calibration	Validation	Calibration	Validation	Calibration	Validation
Jinjiang Basin	2814	0.85	0.52	0.66	0.81	0.56	1.12
Donghe Basin	1644	0.70	0.75	0.82	0.81	0.42	0.40
Heihe Basin	1445	0.78	0.78	0.56	0.54	1.00	0.91
Yalongjiang Basin	1831	0.59	0.73	0.72	0.79	0.92	0.83

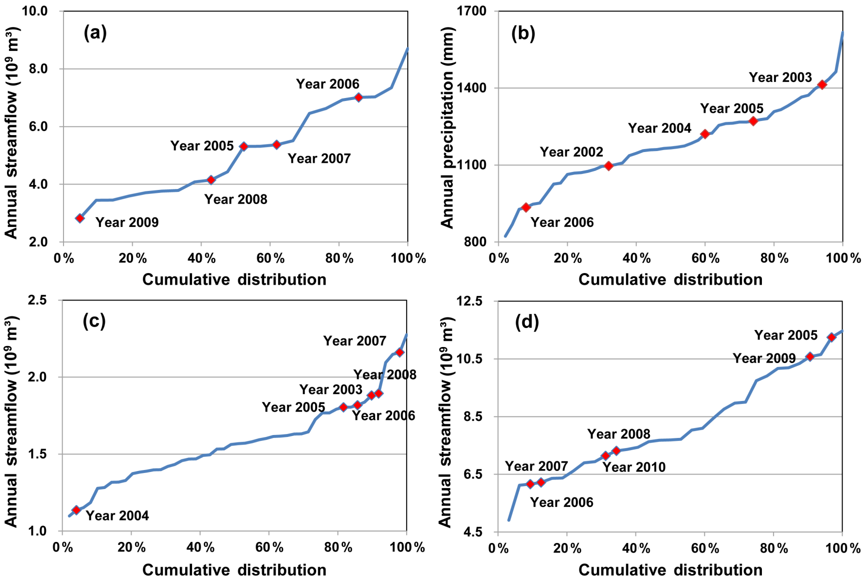

Figure 7. (a) Cumulative distribution of annual streamflow in the Jinjiang Basin (at the Shilong station) for the period of 1958 to 2009. **(b)** Cumulative distribution of annual precipitation in the Donghe Basin for the period of 1961 to 2010. **(c)** Cumulative distribution of annual streamflow in the Heihe Basin (at the Yingluoxia station) for the period of 1960 to 2008. **(d)** Cumulative distribution of annual streamflow in the Yalongjiang Basin (at the Ganzi station) for the period of 1980 to 2011.

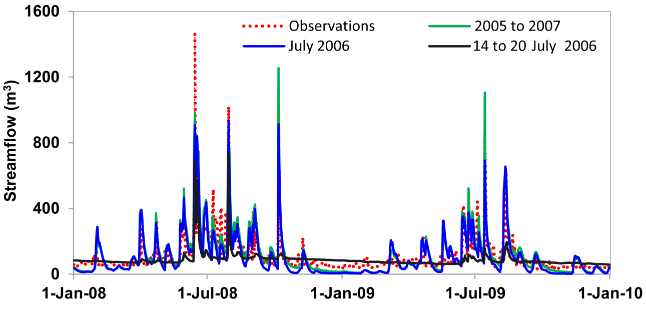

Figure 8. Simulated streamflow of validation period (2008 to 2009) for the Jinjiang Basin case corresponding to the best performed behavioral parameter set derived from calibration using 3-year data (2005 to 2007), one month (July 2006), one week (14 to 20 July 2006) and in situ observations.

creased, indicating that the 2004 dataset was less informative than the other 2 years. The cumulative distribution of annual streamflow at the Yingluoya Station for 1960–2008 (Fig. 7c) indicates that 2004 was an extremely dry year. The other calibration years (2003, 2005) and validation years (2006 to 2008) are all wet years. The limited number of identified behavioral parameter sets derived from calibration using the data of 2004 might only fit the situation of this extremely dry year and they might not perform well in other periods. For the calibrations with the 6-month datasets, only the wet season of 2003, which was the wettest among the 3 years, demonstrated performance comparable with the benchmark calibration. The performances of the other four calibrations were inferior to that of the calibration based on the 3-year dataset. Even the calibration using the dataset of the wet season of 2004 fails to identify behavioral parameter sets. In the arid Heihe Basin, most rainfall occurs in the summer season. About 75 % of total annual streamflow comes from the wet period from April to September. Compared with a normal year, either in wet season or in dry season of in extremely

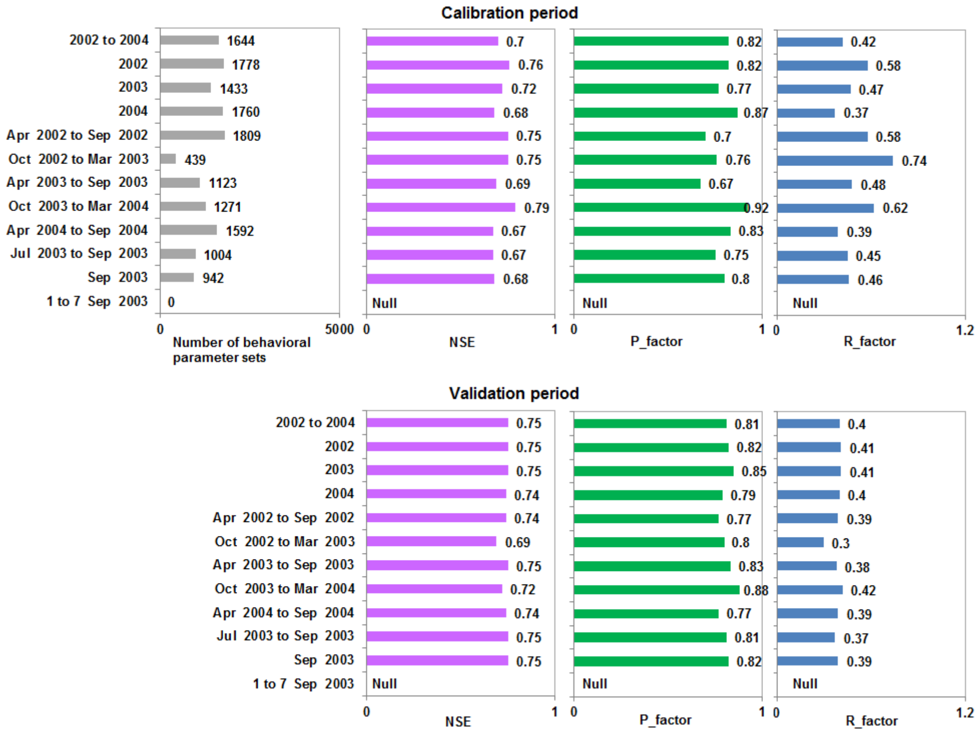

Figure 9. Model performance for the calibrations using short-period data in the Donghe Basin.

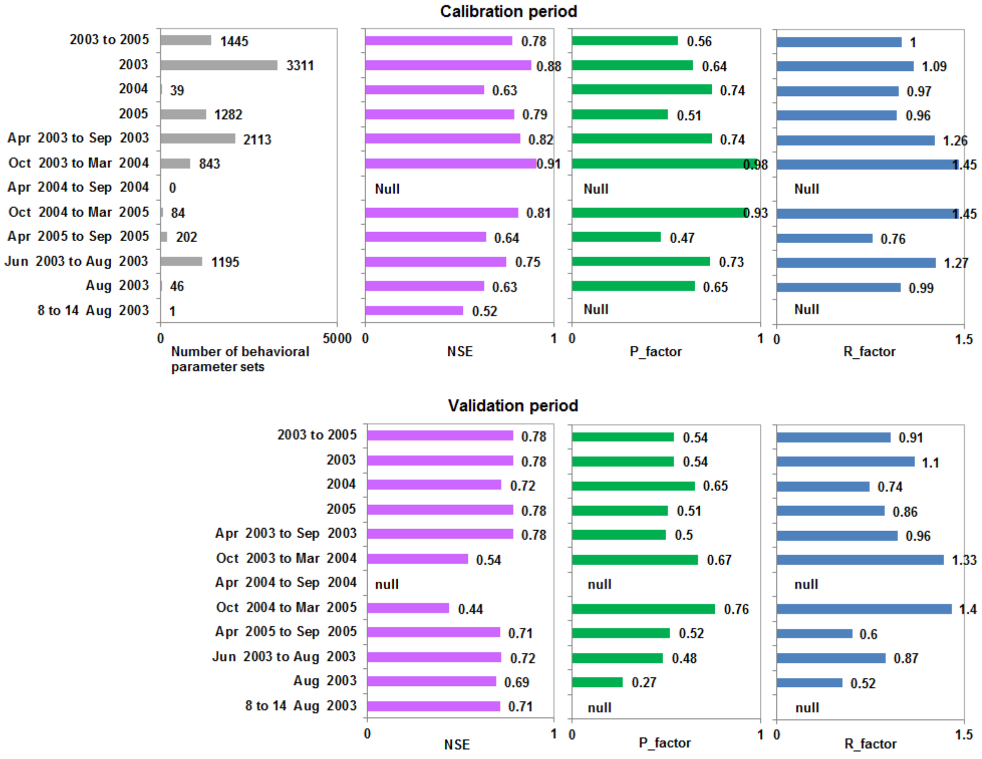

Figure 10. Model performance for the calibrations using short-period data in the Heihe Basin.

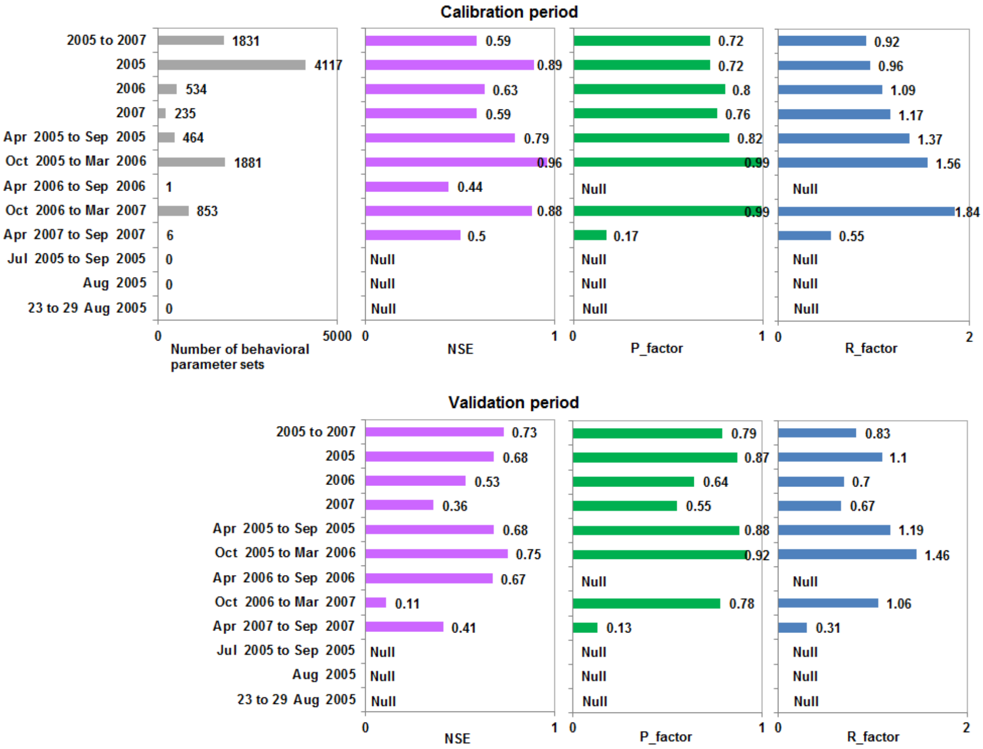

Figure 11. Model performance for the calibrations using short-period data in the Yalongjiang Basin.

dry year 2004, the average streamflow decreased. Considering the big contribution to annual total streamflow, the degree of streamflow decrease in the wet period has a high possibility of being bigger than the dry season. The runoff generation mechanism in this wet season that has extremely low streamflow is very different from a normal situation, which made the model fail to capture the essence of variation in streamflow, and therefore none of the randomly generated 10 000 parameter sets can reproduce the hydrograph of this wet season with acceptable accuracy. Subsets of data for the wet season of 2003 were selected for the second stage of the experiment. The 3-month, 1-month, and 1-week periods with the highest streamflow were June–August, August, and 8–14 August, respectively. None of calibrations based on these datasets achieved similar levels of performance to the benchmark calibration. Based on our evaluation, it is shown that a 6-month dataset could act as a surrogate for the 3-year observational period for model calibration in this arid basin.

3.5 Evaluation of the Yalongjiang Basin case

The cumulative distribution of annual streamflow at Ganzi station (Fig. 7d) indicates that, for the calibration period, 2005 is an extremely wet year, and 2006 and 2007 are extremely dry years; for the validation period, 2009 is a wet year, and 2008 and 2010 are dry years. Figure 11 indicates that, when using 1-year data for calibration, only the wet year 2005 could reach a similar level of performance to the benchmark calibration. The decreases in model performance when using data of the dry years 2006 and 2007 are significant. At the temporal scale of 6 months, the diversity among datasets is high. The 6-month data of the rainy season and dry season in the wettest year 2005 could resemble performance of the benchmark calibration. Only one and six parameter sets are identified as behavioral sets when using rainy season data of the extremely dry years 2006 and 2007, respectively. Similar to the Heihe Basin case, it may be caused by the fact that the runoff generation mechanism in these periods differ from a normal situation, which made the model fail to capture the substantial processes of streamflow variation. For the observations of October 2006 to March 2007, although some behavioral parameters are gained and model performance at the calibration period is satisfying, the calibrated model cannot reproduce the streamflow at the validation period with acceptable accuracy. When using wettest 3-month, 1-month and 1-week data for calibration, no behavioral parameter set was identified, indicating that these three short-period datasets cannot calibrate the model effectively.

3.6 Implications for future applications

The results of this study prove that datasets of continuous daily observations covering periods less than 1 year have the potential to calibrate the SWAT model effectively. In the two wet basins, a 1-month dataset of daily streamflow data could achieve calibration results as good as the benchmark calibration. In the two dry basins, calibration using a 6-month dataset could resemble the performance of calibration using the 3-year dataset. This is in accordance with previous research using lumped conceptual models (Tada and Beven, 2012; Perrin et al., 2007; Seibert and Beven, 2009). Even though the distributed model used in this study is more complex, the results still agree with the findings of Liu and Han (2010), i.e., the information content of the calibration data is more important than the length of the dataset, indicating that only a dataset covering several months might contain sufficient information for parameter identification. This study clearly demonstrates the value of fragmentary historical records of previously existing gauging stations or temporary gauging during field surveys for calibrating a physically based distributed hydrological model in data-sparse basins, at least for basins with a climate characterized by rainy or relatively rainy summers and dry winters, and correspondingly streamflow exhibits an annual cycle of high flow and low flow. The paper could inspire more researchers to think about using such datasets to calibrate distributed hydrological models in basins lacking streamflow data and test it in more well gauged basins to develop a more general understanding about when the measurements are the most informative for parameter calibration. In the past, this approach did not draw much attention for solving the calibration problem of distributed models.

When applying the method to the real world, the biggest challenge is to judge whether the calibrated model can reflect hydrological characteristics of the simulated data-sparse basin. Many calibrations conducted in this study show that if the model could work well in the calibration period, their performance in the validation period is also good. Therefore, the phenomenon that some parameter sets are identified behavioral ones, based on the comparison between simulation and observations, could be considered as one piece of evidence for making the judgement that the short-period data is effective for model calibration. However, such a judgement should be made with care. When the number of observations becomes lower, our results show that the possibility of good performance in the calibration period, accompanied by good simulation in the validation period, decreases. In most calibrations of the two wet basins, such a judgement based on model performance in the calibration period is valid. However, when the number of observations is too low (e.g., the calibration in the Jinjiang Basin using 1-week observations), it may not be valid. In the two dry basins, there are several calibrations showing that good performance in the calibration period does not ensure good performance in the validation period: when using 1-year data for calibration, the performance of dry-year data is inferior to wet-year data. In the Yalongjiang Basin case, the calibrations using 1-year data of dry years 2006 and 2007 even fail to reproduce streamflow in the validation period. In the two dry basins, when using 6-month data for calibration, the diversity of model performance is higher than using 1-year data. This might indicate that drier basins require a greater quantity of data for model calibration, which has been proved by the study using a conceptual model (Lidén and Harlin, 2000), because climatic variability is higher and the runoff generation mechanism is more complex than that in wet basins. Generally in the two dry basins, if the model performance in the calibration period is good, 6-month data from wet years or wet periods make more reliable simulations in the validation period than the ones from dry years or dry periods. Kim and Kaluarachchi (2009) demonstrated that data from high-flow periods have greater control on model calibration because they are more informative with regard to parameter identification. In this context, our suggestion is in line with those made by Yapo et al. (1996) and Melsen et al. (2014): data from wetter periods are preferred for model calibration.

These findings indicate that, to know the “wetness level” of the short-period data, i.e., the records were observed in a wet year or dry year and in a rainy season or dry season, may be helpful to judge whether good simulation could be derived from calibration using a certain short period of observations. In such a context, information about the annual streamflow frequency and intra-annual streamflow regime at the basin outlet is valuable, as the wetness level of the short-period observations can be determined from this information. The coming question is how to get such information in basins lacking streamflow data. Streamflow at a basin outlet is generated by the precipitation data within the basin. A close relationship between streamflow and precipitation exists in a basin. Precipitation data can be obtained more easily than streamflow data, either from in situ gauging or satellite observations. There are publicly available precipitation products (e.g., Global Historical Climatology Network, data available at: https://www.ncdc.noaa.gov/oa/climate/ghcn-daily; Asian Precipitation–Highly Resolved Observational Data Integration Towards the Evaluation of Water Resources, data available at: http://www.chikyu.ac.jp/precip/english) with wide temporal coverage and fine spatial resolution that are sufficient to analyze annual precipitation frequency and intra-annual precipitation regime at basin scale, like the 5 km spatial resolution data used in the case of the Donghe Basin. Information about the precipitation frequency can work as a surrogate of annual streamflow frequency and intra-annual streamflow regime, in order to determine the wetness level of certain short-period streamflow data in real applications, and correspondingly the performance of calibrated model could be indirectly assessed. To develop a general understanding of whether the information content in certain limited calibration data records is sufficient to obtain robust parameter values, further studies similar to this one, using distributed model in large number of well gauged basins with differing characteristics, are required. Such studies need to generate many samples of short-period observations from available streamflow data and then use the samples to calibrate the model. A reasonable sampling strategy is needed for this kind of research. Our study shows that, to some extent, the wetness level of short-period data is related to the performance of calibrated model. Therefore, considering the wetness level of data in the sampling strategies may be valuable to obtain general guidelines on when the short-period observations are informative for model calibration.

4 Conclusions

This study was an initial evaluation of the possibility of calibrating physically based distributed hydrological models using limited streamflow data, which could be extracted from available fragmentary historical observation records or obtained from field campaigns in the target basin. It could be considered a solution to the problem of ungauged basins in some situations. Through application of the SWAT model to four Chinese basins with different climatic and hydrological characteristics, it has been demonstrated that datasets of daily measurements over periods of less than 1 year can constrain simulation uncertainty as effectively as calibration datasets covering several years. In the two wet basins, it was demonstrated surprisingly well that the model could be calibrated successfully using only a 1-month dataset, whereas in the two dry basins, longer datasets (6 months) were required and data from wet years and wet periods demonstrated more reliability than data from dry years and dry periods. The results of this study clearly indicate the potential of short-term streamflow observations in calibrating distributed hydrological models for ungauged basins. In the real world, it is difficult to assess whether good simulations are achievable with limited calibration data because of the lack of model validation data. Our results show that the phenomenon in which some parameter sets are identified as behavioral ones, based on the comparison between simulation and observations, could be considered as one piece of evidence for making the judgement that the short-period data is reliable. However, such judgement should be made with careful consideration as our study also shows that it may not be true when the number of the observations is too low or data are observed in a dry year or a dry period. It may only be valid when using data with a length of at least several months and observed in a rainy season or a wet year. To get more general knowledge about when the observations are most informative for model calibration, more studies similar to our studies should be conducted. Based on our findings, the relationship between the wetness level of short-period data and their effectiveness for parameter calibration are worthy of being explored in these kinds of future studies.

5 Data availability

Please contact the corresponding author to access the data in this study.

Acknowledgements. This study was supported by the National Natural Science Foundation of China (Grant Nos. 41671018, 91647202), the Non-profit Industry Financial Program of Ministry of Water Resources of China (Grant No. 201401036), the National Key Research and Development Program of China (Grant No. 2016YFC0401308), and Fundamental Research Funds for the Central Universities.

Edited by: Y. Chen
Reviewed by: two anonymous referees

References

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R.: Large area hydrologic modeling and assessment – Part 1: Model development, J. Am. Water. Resour. As., 34, 73–89, 1998.

Beven, K.: How far can we go in distributed hydrological modelling?, Hydrol. Earth Syst. Sci., 5, 1–12, doi:10.5194/hess-5-1-2001, 2001.

Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, Hydrol. Process., 6, 279–298, 1992.

Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, J. Hydrol., 249, 11–29, 2001.

Callahan, B., Miles, E., and Fluharty, D.: Policy implications of climate forecasts for water resources management in the Pacific Northwest, Pol. Sci., 32, 269–293, 1999.

Finger, D., Heinrich, G., Gobiet, A., and Bauder, A.: Projections of future water resources and their uncertainty in a glacierized catchment in the Swiss Alps and the subsequent effects on hydropower production during the 21st century, Water Resour. Res., 48, 02521, doi:10.1029/2011WR010733, 2012.

Freer, J. and Beven, K., Bayesian estimation of uncertainty in runoff predication and the value of data: An application of the GLUE approach, Water Resour. Res. 32, 2161–2173, 1996.

Gassman, P. W., Reyes, M. R., Green, C. H., and Arnold, J. G.: Soil and Water Assessment Tool: Historical Development, Applications, and Future Research Directions, T. Asabe, 50, 1211–1250, 2007.

Getirana, A. C. V.: Integrating spatial altimetry data into the automatic calibration of hydrological models, J. Hydrol., 387, 244–255, 2010.

Gupta, H. V., Beven, K. J., and Wagener, T.: Model Calibration and Uncertainty Estimation, in: Encyclopedia of hydrological science, edited by: Anderson, M. G., John Wiley & Sons, Ltd, 2005.

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., Mcdonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., and Ehret, U.: A decade of Predictions in Ungauged Basins (PUB) – a review, Hydrolog. Sci. J., 58, 1198–1255, 2013.

Khu, S. T., Madsen, H., and di Pierro, F.: Incorporating multiple observations for distributed hydrologic model calibration: An approach using a multi-objective evolutionary algorithm and clustering, Adv. Water Resour., 31, 1387–1398, doi:10.1016/j.advwatres.2008.07.011, 2008.

Kim, U. and Kaluarachchi, J. J.: Hydrologic model calibration using discontinuous data: an example from the upper Blue Nile River Basin of Ethiopia, Hydrol. Process., 23, 3705–3717, 2009.

Lidén, R. and Harlin, J.: Analysis of conceptual rainfall–runoff modelling performance in different climates, J. Hydrol., 238, 231–247, 2000.

Liu, J. and Han, D.: Indices for Calibration Data Selection of the Rainfall-Runoff Model, Water Resour. Res., 46, 292–305, 2010.

Madsen, H.: Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives, Adv. Water Resour., 26, 205–216, doi:10.1016/S0309-1708(02)00092-1, 2003.

McEnery, J., Ingram, J., Duan, Q., Adams, T., and Anderson, L.: NOAA'S Advanced Hydrologic Prediction Service: Building Pathways for Better Science in Water Forecasting, B. Am. Meteorol. Soc., 86, 375–385, 2005.

Melsen, L. A., Teuling, A. J., Berkum, S. W. V., Torfs, P. J. J. F., and Uijlenhoet, R.: Catchments as simple dynamical systems: A case study on methods and data requirements for parameter identification, Water Resour. Res., 50, 5577–5596, doi:10.1002/2013WR014720, 2014.

Muleta, M. K. and Nicklow, J. W.: Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model, J. Hydrol., 306, 127–145, doi:10.1016/j.jhydrol.2004.09.005, 2005.

Perrin, C., Oudin, L., Andreassian, V., Rojas-Serna, C., Michel, C., and Mathevet, T.: Impact of limited streamflow data on the efficiency and the parameters of rainfall-runoff models, Hydrolog. Sci. J., 52, 131–151, 2007.

Revilla-Romero, B., Beck, H. E., Burek, P., Salamon, P., de Roo, A., and Thielen, J.: Filling the gaps: Calibrating a rainfall-runoff model using satellite-derived surface water extent, Remote Sens. Environ., 171, 118–131, doi:10.1016/j.rse.2015.10.022, 2015.

Seibert, J. and Beven, K. J.: Gauging the ungauged basin: how many discharge measurements are needed?, Hydrol. Earth Syst. Sci., 13, 883–892, doi:10.5194/hess-13-883-2009, 2009.

Silvestro, F., Gabellani, S., Rudari, R., Delogu, F., Laiolo, P., and Boni, G.: Uncertainty reduction and parameter estimation of a distributed hydrological model with ground and remote-sensing data, Hydrol. Earth Syst. Sci., 19, 1727–1751, doi:10.5194/hess-19-1727-2015, 2015.

Sivapalan, M.: Prediction in Ungauged Basins: a grand challenge for theoretical hydrology, Hydrol. Process., 17, 3163–3170, 2003.

Sun, W. C., Ishidaira, H., and Bastola, S.: Towards improving river discharge estimation in ungauged basins: calibration of rainfall-runoff models based on satellite observations of river flow width at basin outlet, Hydrol. Earth Syst. Sci., 14, 2011–2022, doi:10.5194/hess-14-2011-2010, 2010.

Sun, W. C., Ishidaira, H., and Bastola, S.: Prospects for calibrating rainfall-runoff models using satellite observations of river hydraulic variables as surrogates for in situ river discharge measurements, Hydrol. Process., 26, 872–882, 2012.

Sun, W. C., Ishidaira, H., Bastola, S., and Yu, J. S.: Estimating daily time series of streamflow using hydrological model calibrated based on satellite observations of river water surface width: Toward real world applications, Environ. Res., 139, 36–45, 2015.

Tada, T. and Beven, K. J.: Hydrological model calibration using a short period of observations, Hydrol. Process., 26, 883–892, 2012.

Todorovic, A. and Plavsic, J.: The role of conceptual hydrologic model calibration in climate change impact on water resources assessment, J. Water Clim. Change, 7, 16–28, 2016.

United Nations Offices for Disaster Risk Reduction: 2015 Disasters in Numbers, available at: http://www.unisdr.org/we/inform/publications/47804 (last access: 5 January 2017), 2016.

Vervoort, R. W., Miechels, S. F., van Ogtrop, F. F., and Guillaume, J. H. A.: Remotely sensed evapotranspiration to calibrate a lumped conceptual model: Pitfalls and opportunities, J. Hydrol., 519, 3223–3236, 2014.

Viviroli, D. and Seibert, J.: Can a regionalized model parameterisation be improved with a limited number of runoff measurements?, J. Hydrol., 529, 49–61, 2015.

Vrugt, J. A., Willem, B., Gupta, H. V., and Soroosh, S.: Toward improved identifiability of hydrologic model parameters: The information content of experimental data, Water Resour. Res., 38, 1312, doi:10.1029/2001WR001118, 2002.

Winsemius, H. C., Savenije, H. H. G., and Bastiaanssen, W. G. M.: Constraining model parameters on remotely sensed evaporation: justification for distribution in ungauged basins?, Hydrol. Earth Syst. Sci., 12, 1403–1413, doi:10.5194/hess-12-1403-2008, 2008.

Wohl, E., Barros, A., Brunsell, N., Chappell, N. A., Coe, M., Giambelluca, T., Goldsmith, S., Harmon, R., Hendrickx, J. M. H., and Juvik, J.: The hydrology of the humid tropics, Nature Reports Climate Change, 2, 655–662, 2012.

Wu, Y. P. and Liu, S. G.: Automating calibration, sensitivity and uncertainty analysis of complex models using the R package Flexible Modeling Environment (FME): SWAT as an example, Environ. Modell. Softw., 31, 99–109, doi:10.1016/j.envsoft.2011.11.013, 2012.

Yang, J., Reichert, P., Abbaspour, K. C., Xia, J., and Yang, H.: Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China, J. Hydrol., 358, 1–23, 2008.

Yapo, P. O., Gupta, H. V., and Sorooshian, S.: Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data, J. Hydrol., 181, 23–48, 1996.