Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-192, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 17 May 2016

Hydrology and Earth System Sciences Discussions

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

Wenchao Sun[1,2], Yuanyuan Wang[1,2], Xingqi Cui [1,2], Jingshan Yu[1], Depeng Zuo[1,2] 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)

Abstract. Physically-based distributed hydrological models are widely used for hydrological simulations in various environments. However, as with conceptual models, they are limited in data-sparse basin by the lack of streamflow data for calibration. Short periods of observational data (less than 1 year) may be obtained from the fragmentary historical records of past-existed gauging stations or from temporary gauging during field surveys, which might be of values for model calibration. 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 observation period on the calibration of the widely applied Soil and Water Assessment Tool model was evaluated in two 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 for the two basins. 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 both 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 model of wet Jinjiang Basin could be effectively calibrated using a shorter data record (1 month), compared with the arid Heihe Basin (6 months). Even though the two basins are very different, the results demonstrated that data from the wet season and wetter years performed better that data from the dry season and drier year. The results of this study demonstrated that short periods of observations could be a promising solution to the problem of calibration of physically-based distributed hydrological models in data-sparse basins and further researches similar to this study are required to gain more general understandings about the optimum number of observations needed for calibration when such model are applied to real data-sparse basins.

Key words: Physically-Based Distributed Hydrological Model; Length of observations data; Model calibration: Data-sparse basin

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 term 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. Most parameters of hydrological models are conceptual without explicit physical meaning, which makes it necessary to identify parameter values through model calibration based on streamflow data (Gupta 2005). However, there has been a general decline in the networks designed to monitor streamflow (Wohl et al., 2012), especially in developing countries, because of resource constraints (e.g., financial and human resources), 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 the introduction of limited information from the ungauged basin. 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 focused on using 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), groundwater level(e.g., Khu et al., 2008). These studies have shown promising performances for identifying parameters that describe the processes being measured. However, none of these observations has the capability of 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. Therefore, if such data are to be used for calibration, to know how many observations are needed to calibrate model parameter is important. 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 a 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).

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-192, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 17 May 2016

For conceptual models, these researches indicated that with observations of the order of several scores, reasonable parameter estimates could be derived, 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, 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 been addressed rarely 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 whether physically-based distributed hydrological models could be calibrated effectively using short-period of continuous observations, which might be obtained from fragmentary historical records of past-existed 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., Liden and Harlin, 2000). Therefore, we selected two basins with different climatic conditions (one in a humid region and the other in an arid region) to improve the generality of our findings. The evaluation relies on comparison with the conventional calibration using observations covering several years, which was adpoted as the benchmark simulation. 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) method (Beven and Binley, 1992; Freer and Beven, 1996) was used for the model calibration, for which all the settings during the calibration were verifiable and satisfying 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 USDA. 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 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 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 to 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 options 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 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:

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

where $Q_{obs,i}$ (m^3/s) and $Q_{sim,i}$ (m^3/s) represent the observed and simulated streamflow, respectively, at time step i, and $Q_{obs,avg}$ (m^3/s) 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:

$$L_p[\theta \mid Q_{obs}] = CL[\theta \mid Q_{obs}] L_o[\theta] \tag{2}$$

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-192, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 17 May 2016

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), $L[\theta|Q_{obs}]$ is the NSE that quantifies the performance of reproducing Q_{obs}, and C is a scaling factor 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 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

To make our study more generalized, two Chinese basins with different climatic and geophysical conditions were selected: one is located in a humid region and the other is located in an arid region. For each basin, the influence of the length of the observational record on the calibration will be explored. Then, comparisons between the two basins will be performed to obtain insights that are more generalized.

The Jinjiang Basin (Fig. 1) is located on the west coast of the Taiwan Straits in Fujian Province, China. The area of the basin is 5629 km2. 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 rangs 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 monsoon climatic region, with 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 Heihe Basin (Fig. 2) is in the arid northwest of China. It is the second largest inland basin in China with an area about 128,900 km2. From the southern mountainous region to the northern high-plain area, the elevation decreases from about 5000to 1000 m. The hydrological simulation was executed for the upstream mountainous region of Yingluoya gauging station, encompassing an area of around 10,000 km2. The elevation of the study area varies from more than 5000m 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 arid summers, with average annual precipitation around 400mm.

2.4 Experiment design

Based on the availability of streamflow records, the benchmark calibration for the Jinjiang Basin was based on full daily observations for 2005–2007 and it was validated using data for 2008-2009. For the Heihe Basin, the calibration and validation periods were 2003–2005 and 2006–2008, respectively. As a first trial for the distributed model, we sought to explore the highest possible performance using certain short lengths of records, not the general performance of specific lengths. Under such a precondition, the evaluation considered whether the calibration using a short period of data (i.e., a subset of the calibration data of the benchmark calibration) could achieve performance similar to the benchmark calibration. If so, a subset of the pervious dataset was used for the calibration. This process was repeated until the performance of the subset decreased significantly. In this context, two issues need to be considered carefully: the assessment of the performance of each calibration and the strategic selection of the short period of data.

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, in most time steps of the simulation period, 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, (but mostly in the validation period), because no information is used for the calibration. Subsets of the calibration data from the observations used in the benchmark calibration were extracted strategically, which will be introduced later. These short periods of records were used for the calibration and compared with the benchmark calibration. The validation periods selected for the case studies of the Jinjiang Basin and Heihe Basin were 2008–2009 and 2006–2008, respectively, the same as for 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 constrained by the calibration data) for the calibration and validation periods. Two indexes were combined to assess the simulation uncertainty. The *R_factor* is a measure of the average width of 95% simulation intervals

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

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 σ_{Qobs} is the standard deviation of the streamflow observations. The *P_factor* is the percentage of observations embraced by the 95% prediction intervals. A low value of *R_factor* combined with a high value of the *P_factor* indicates low simulation uncertainty. A new index *U* combining the *P_factor* with the *R_factor* was defined as:

$$U = 1 - \frac{P\_factor}{R\_factor} \tag{5}$$

and computed for both the calibration and validation period. Lower values of U indicate lower simulation uncertainty.

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 possible to follow the studies of conceptual models (e.g. Perrin et al 2007; Seibert and Beven, 2009) that could conduct calibrations many times. To perform the calibration in manageable times, the experiment was conducted in two stages. First, three calibrations using 1-year data record that covered both the rainy and dry seasons, and five calibrations using 6-month data record that covered either a rainy season or a dry season were undertaken. The selected periods for both basins were subsets of the calibration data used for the benchmark calibration and they are listed in Table 1. If there are calibrations using the 6-month data record which could achieve performances similar to the benchmark calibration, stage two of the experiment was initialized, in which the subsets of the best performing 6-month data record were used for calibration. 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, because most model modules are activated in high-flow periods. As our study explored the possibility of optimum performance of certain lengths of records for calibration, the 3-month, 1-month data and one 1-week datasets with highest average streamflow corresponding to the above three time scales were employed to calibrate the model and conduct the evaluation.

3 Results and discussion

3.1 Performances of the two benchmark calibrations

Before we can apply the model for the evaluation proposed in this study, the model robustness in the two 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 2) 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 benchmark calibrations of the two basins, the results of the calibration are summarized in Table 3, and the best simulations and uncertainty bands of the ensemble simulation are shown in Figs. 3 and 4. For the Heihe Basin, the threshold for likelihood was set to 0.5, whereas for the Jinjiang Basin, too many parameter sets could result reasonable simulations, and therefore the threshold was raised to 0.7. The NSEs of the best simulation in two 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 for both basins were successful. The results of these two 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 listed in Table 4. For the 1-year period, all three calibrations performed similarly to the benchmark calibration, and the dataset for 2006 even outperformed the benchmark. Figure 5 presents the cumulative distribution of the available annual streamflow for Shilong station from 1958 to 2009, showing that 2006 had the highest annual streamflow and was a wet year. On the 6-month time scale, the corresponding five calibrations exhibited considerable differences: As a subset of the 2006 dataset, the calibration using data for the period April 2006 to September 2006 achieved better performance than the benchmark calibration. However, no parameter sets were identified as behavioral parameter set 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. Furthermore, it was found that the performance using datasets for the wet season was generally better than using dry season datasets. The second stage of the experiment was undertaken using 3-month (June–August), 1-month (July), and 1-week (July 14–20) datasets with the highest streamflow during April to September 2006. Table 4 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 similar performance to benchmark calibration. Thus, it is indicated that in the Jinjiang Basin, it is possible to calibrate the SWAT model effectively using only 1-month's continuous daily observations of streamflow.

3.3 Evaluation of the Heihe Basin case

The results of the calibration are shown in Table 5. The calibrations using 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 validation period decreased, indicating the 2004 dataset was less informative than the other 2 years. The cumulative distribution of annual streamflow at Yingluoya Station for 1960–2008 (Fig. 6) indicates that 2004 was an extremely dry year. The limited number of identified behavioral parameter sets might only fit the situation of this extremely dry year and they might not perform well in other periods. For the calibrations with 6-month dataset, 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 calibration based on the 3-year dataset. Even the calibration using the dataset of wet season of 2004 failed to identify behavioral parameter sets. This indicated that the variations of the hydrological processes during the wet season of this extreme dry year is difficult to describe using the model, and that of the five 6-months dataset was least representative of the characteristics of the water cycle of the studied basin. 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 August 8–14, respectively. None of calibrations based on these datasets achieved similar levels of performance

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-192, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 17 May 2016

as the benchmark calibration. Based on our evaluation, it is shown that a 6-month dataset could act as a surrogate for 3-year observational period for model calibration.

3.4 Comparison of the two case studies

The evaluations of the selected wet and dry basins revealed similarities and differences that are important to the generalization of the findings of this study, and useful for the evaluation of the value of available limited hydrological data in solving real problems. In both basins, datasets of continuous daily observations covering periods less than 1 year were shown to achieve similar performances to a model calibration based on a 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, which means it is possible that only a dataset covering several months might contains sufficient information for parameter identification. The evaluation established that in the wet Jinjiang Basin, a 1-month dataset of daily streamflow data could achieve calibration results as good as the benchmark calibration. However, for the dry Heihe Basin, datasets covering less than 6 months could not identify parameters effectively. 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 (Liden and Harlin, 2000), because climatic variability is higher and the runoff generation mechanism is more complex than in wet basins. Both cases show that when the length of calibration dataset is the same, data from wetter years and from wet periods perform better than data from drier years and 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), which is that data from wetter periods should always be preferred 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 obtained from field campaigns in the target basin or extracted from available fragmentary historical observation records. It could be considered a solution to the problem of ungauged basins in some situations. Through application of the SWAT model to two 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 wet Jinjiang Basin, it was demonstrated surprisingly that the model could be calibrated successfully using only a 1-month dataset, whereas in the dry Heihe Basin, longer datasets (6 months) were required. Interestingly, although the two basins are quite different, when using limited numbers of data for the model calibration, data from wetter years and from wet periods demonstrated best performance in both cases. The results of this study indicate the value of short-term streamflow observations in calibrating distributed hydrological models for ungauged basins. However, in real applications, it is difficult to assess whether good simulations are achievable with limited calibration data because of the lack of model validation data. Knowing how many observations are needed and in which period the observations are most informative are very important for practical applications in ungauged basins. Therefore, further similar researches using distributed model in other basins with differing characteristics is necessary to develop a general understanding of whether the information content in limited calibration data is sufficient to constrain the model parameters to reflect basin realities.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (Grant Nos. 41201018, 91125015), Non-profit Industry Financial Program of Ministry of Water Resources of China (grant No. 201401036) and Fundamental Research Funds for the Central Universities.

References

Abbaspour, K. C.:SWAT-CUP: SWAT Calibration and Uncertainty Programs - A User Manual, Swiss Federal Institute of Aquatic Science and Technology, 2015.

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

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.

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

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

Cheng, Q. B., Chen, X., Xu, C. Y., Reinhardt-Imjela, C. and Schulte, A.: Improvement and comparison of likelihood functions for model calibration and parameter uncertainty analysis within a Markov chain Monte Carlo scheme, J. Hydrol., 519, 2202-2214, doi: 10.1016/j.jhydrol.2014.10.008, 2014.

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, The, 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, Anderson MG (eds), John Wiley & Sons, Ltd,2006.

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, Hydrol. Sci. J., 58, 1198-1255, 2013.

Khu, S. T. and 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, 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, Hydrology & Earth System Sciences, 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. and 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, 2010.

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.

Sun, W. C. and 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.

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

United Nations Offices for Disaster Risk Reduction, 2015 Disasters in Numbers. http://www.unisdr.org/we/inform/publications/47804, 2016.

Viviroli, D. and Seibert, J.: Can a regionalized model parameterisation be improved with a limited number of runoff measurements? J. Hydrol., 529, Part 1, 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.

Willem Vervoort, R., 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.

Winsemius, H. C. and Hhg, S. and Wgm, B.: Constraining model parameters on remotely sensed evaporation: justification for distribution in ungauged basins? Hydrology & Earth System Sciences, 28, 1403-1413, doi: 10.5194/hess-12-1403-2008, 2009.

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.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-192, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 17 May 2016

Hydrology and Earth System Sciences
Open Access
Discussions

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

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-192, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 17 May 2016

Hydrology and Earth System Sciences
Open Access
Discussions

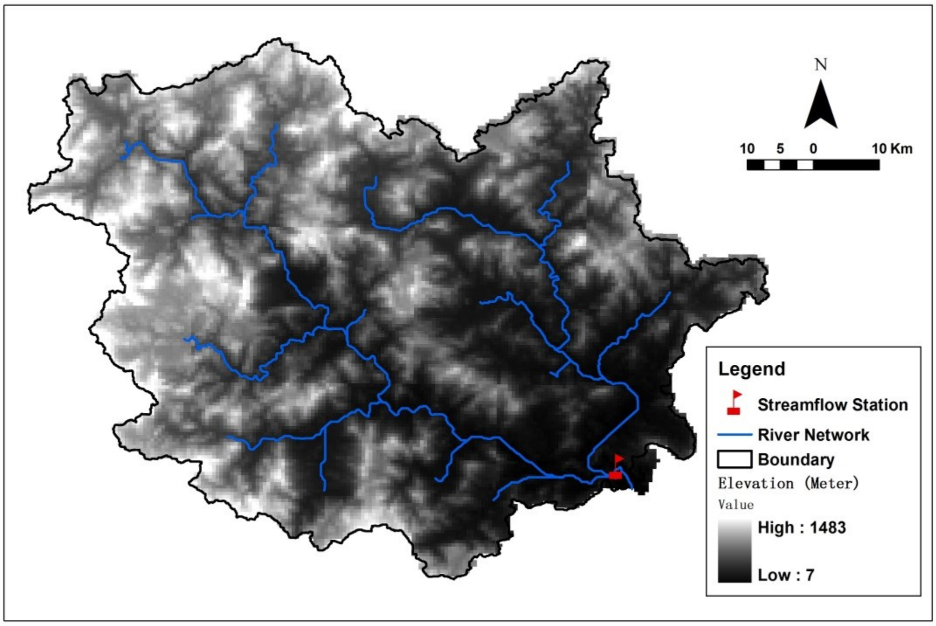

Figure 1: Topography, river network and the streamflow gauging station in the Jinjiang Basin, China

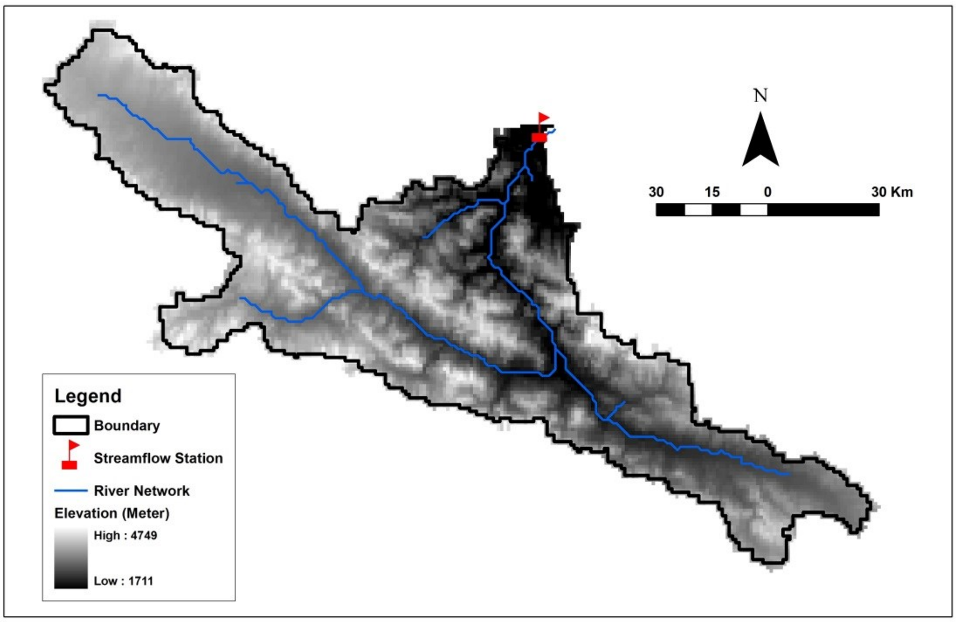

Figure 2: Topography, river network and the streamflow gauging station in the Heihe Basin, China

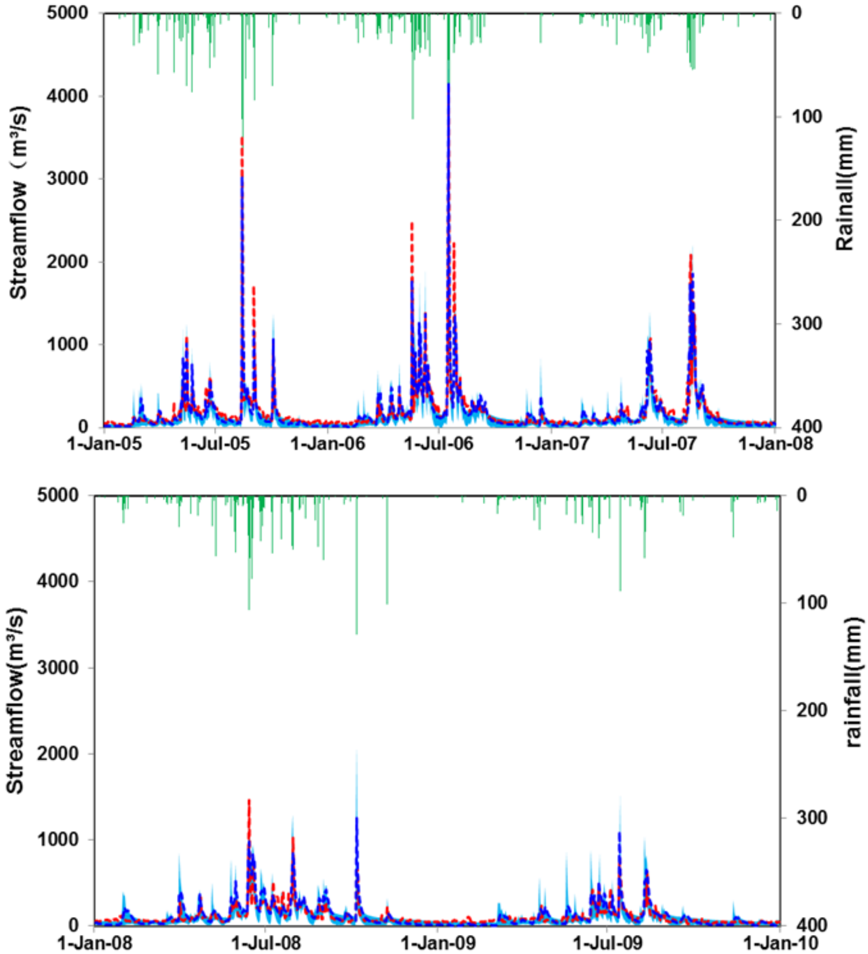

Figure 3: Simulated streamflow for the benchmark calibration of the Jinjiang Basin in both calibration (2005–2007) and validation period (2008–2009). Dashed lines: observed streamflow; blue band: 95% uncertainty band of ensemble simulation; solid blue lines: best simulation of ensemble prediction; green columns: rainfall

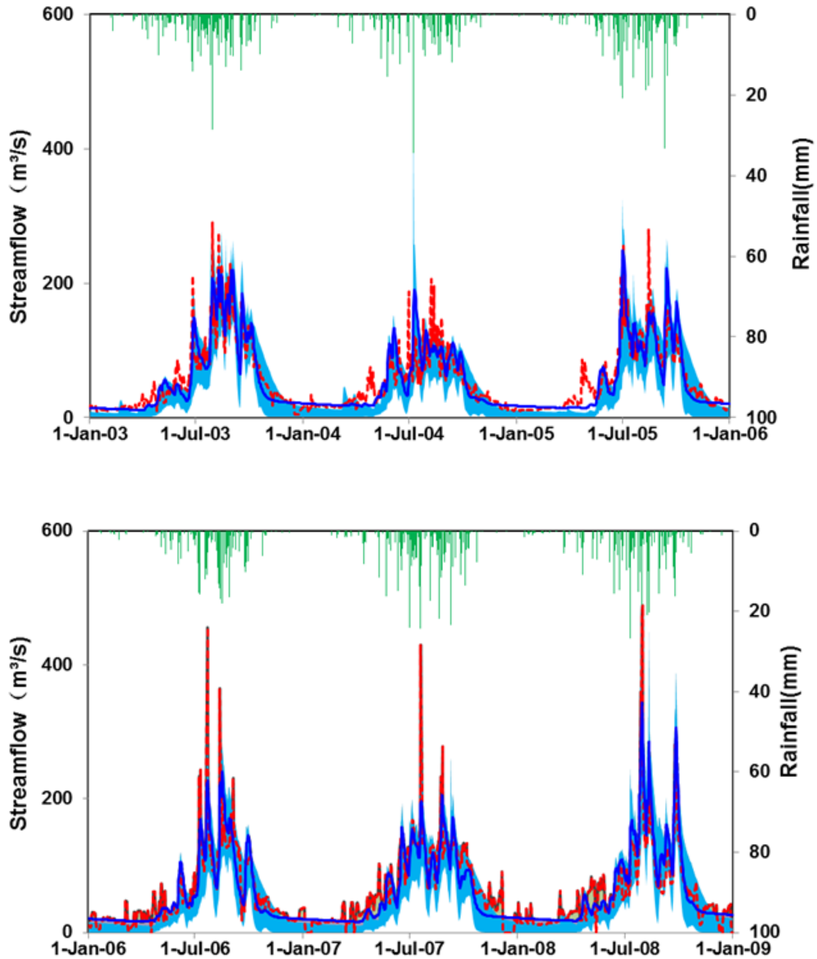

Figure 4: Simulated streamflow for the benchmark calibration of the Heihe Basin in both calibration (2003–2005) and validation period (2006–2008). Dashed lines: observed streamflow; blue band: 95% uncertainty band of ensemble simulation; solid blue lines: best simulation of ensemble prediction; green columns: rainfall

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-192, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 17 May 2016

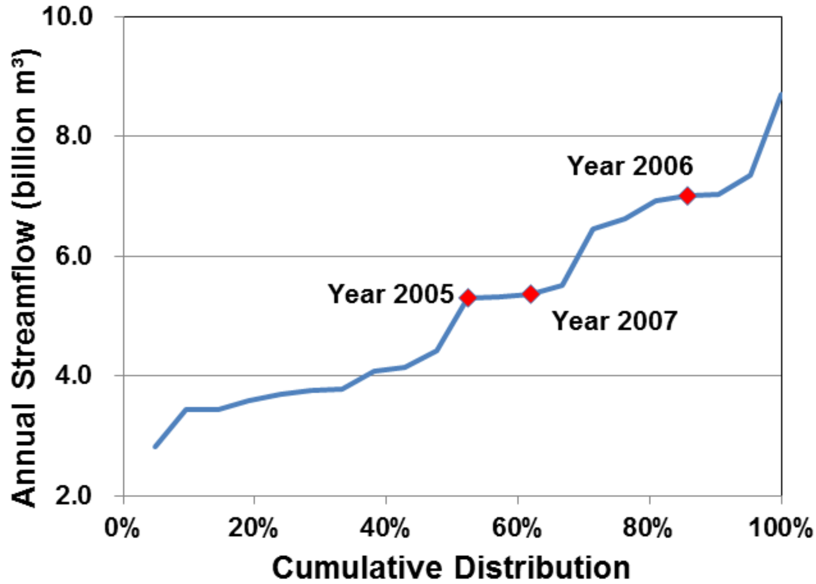

Figure 5: Cumulative distribution of available annual streamflow at Shilong station for the period of 1958 to 2009

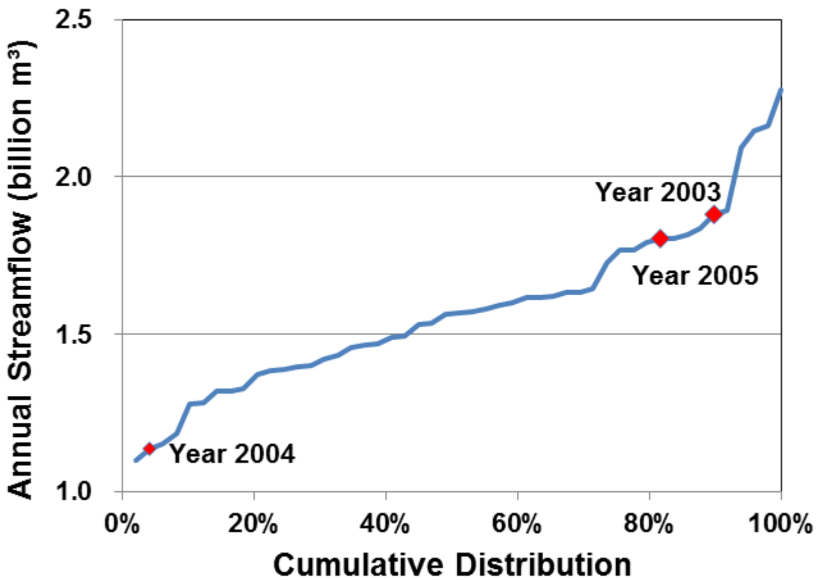

Figure 6: Cumulative distribution of annual streamflow at Yingluoxia station for the period of 1960 to 2008

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-192, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 17 May 2016

Hydrology and Earth System Sciences
Open Access
Discussions

Table 1 Short periods for which corresponding data were used for the calibrations for stage one of the evaluation

Length of the period	Jinjiang Basin	Heihe Basin
One year	2005	2003
	2006	2004
	2007	2005
Six months	April 2005 to September 2005	April 2003 to September 2003
	October 2005 to March 2006	October 2003 to March 2004
	April 2006 to September 2006	April 2004 to September 2004
	October 2006 to March 2007	October 2004 to March 2005
	April 2007 to September 2007	April 2005 to September 2005

Table 2 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/hr)	5–130
SOL_AWC	Available soil water capacity (mm H_2O/mm Soil)	0–1
SOL_K	Soil Saturated hydraulic conductivity (mm/hr)	0–2000

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-192, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 17 May 2016

Hydrology and Earth System Sciences
Discussions
Open Access

Table 3 Model performance for the benchmark calibration in the two basins

	Number of behavioral sets	NSE		U	
		Calibration	Validation	Calibration	Validation
Heihe Basin	1445	0.78	0.78	0.44	0.41
Jinjiang Basin	2814	0.85	0.52	-0.18	0.28

Table 4 Model performance for the calibrations using short-period data in Jinjiang basin

Length of records	Calibration period	Number of behavioral sets	NSE		U	
			Calibration	Validation	Calibration	Validation
One year	2005	2046	0.86	0.52	-0.02	0.22
	2006	2820	0.85	0.67	-0.33	0.29
	2007	3888	0.88	0.58	-0.15	0.32
Six months	Apr. 2005 to Sep. 2005	1916	0.85	0.52	0.22	0.25
	Oct. 2005 to Mar. 2006	492	0.88	0.62	0.07	0.13
	Apr. 2006 to Sep. 2006	2359	0.83	0.67	-0.20	0.28
	Oct. 2006 to Mar. 2007	-	-	-	-	-
	Apr. 2007 to Sep. 2007	3163	0.86	0.51	0.06	0.31
Three months	Jun. 2006 to Aug. 2006	2338	0.84	0.65	0.12	0.33
One month	Jul. 2006	3064	0.86	0.57	-0.40	0.34
One week	Jul 14 to 20, 2006	1370	0.87	0.40	-0.11	0.38

Table 5 Model performance for the calibrations using short-period data in Heihe basin

Length of records	Calibration period	Number of behavioral sets	NSE		U	
			Calibration	Validation	Calibration	Validation
One year	2003	3311	0.88	0.78	0.41	0.51
	2004	39	0.63	0.72	0.24	0.12
	2005	1282	0.79	0.78	0.47	0.41
Six months	Apr. 2003 to Sep. 2003	2113	0.82	0.78	0.41	0.48
	Oct. 2003 to Mar. 2004	843	0.91	0.54	0.32	0.50
	Apr. 2004 to Sep. 2004	-	-	-	-	-
	Oct. 2004 to Mar. 2005	84	0.81	0.44	0.36	0.46
	Apr. 2005 to Sep. 2005	202	0.64	0.71	0.38	0.13
Three months	Jun. 2003 to Aug. 2003	1195	0.75	0.72	0.43	0.45
One month	Aug. 2003	46	0.63	0.69	0.34	0.48
One week	Aug 8 to 14, 2003	1	0.52	0.71	-	-