Dear editor and reviewers,

Thank you very much for your constructive comments and recommendations. Our responses to comments, list of changes and the marked-up manuscript are as follows. If you feel more explanations or revisions are needed. Please do not hesitate to contact with us.

Best regards,

From the authors

Responses to the editor:

(1) As is suggested by reviewer #2, two watersheds, typical though, are far from enough. If this manuscript is designed to only provide an idea or assumption on "short period calibration of hydrological models", it does not contribute a lot to the hydrological community as many previous researches have already done this using different models though (same comments as reviewer #2). Therefore, I think you may conduct few more case studies, test it and make it convincing. Or if this manuscript focuses on short period calibration of "Physically-based distributed hydrological model", a discussion on the calibration of both conceptual models and physical models is expected. And also their differences in calibration should be discussed. In addition, what can we learn from this idea and how can we use this idea in the calibration of hydrological models for data-scarce basins?

Response:

To make the findings of this study more convincing, we added two more basins to our study: the upstream Yalongjiang Basin with Ganzi station as the outlet, which is a dry basin in the Qinghai-Tibet Plateau, and the Donghe Basin, a wet basin in the upstream region of Three Gorges Reservoir. By doing this, we could test our idea in two wet basins, Jinjiang Basin and Donghe Basin, and two dry basins, upstream Heihe Basin and upstream Yalongjiang Basin, to improve the universality of our conclusions. The main findings of this study are as follows:

1. In the four basins, it is possible that data records with lengths of less than one year could calibrate the model effectively. The models in the two wet basins could be calibrated effectively using a shorter period of records (one month) than the two arid basins (six months).
2. When using one-year data for calibration, for the two wet basins, whether the year is wet or dry have little influences on model calibration. However, in the two dry basins, the influence is significant: data of dry years are less reliable than the ones of wet years.
3. When using 6-month observations for calibration, in all of the four basins the model performances are more sensitive to the timing of the observations than using 1-year data. And this sensitivity is higher for the two dry basins. Data from wet year or wet season are more reliable than the ones from dry year or dry season, especially for the two dry basins.

From these findings, in our opinion, the contributions of this study to utilization of short periods of observations for calibrating physically based distributed hydrological model are as follows:

The paper could inspire more researchers to think about using such dataset to calibrate distributed hydrological models in basins lacking of streamflow data. Many past researches concentrate on this approach for the calibration of lumped conceptual model. However, the possibility for the physically based distributed hydrological model, like SWAT used in this study, has not been discussed widely. Our study confirms that for physically based distributed hydrological models, this approach could also work well.

More importantly, this study is valuable for evaluating the effectiveness of short period data for model calibration in real world application. Our results show that, the phenomenon that some parameter sets are identified behavioral ones based on the comparison between simulation and observations could be considered as one evidence for making the judgement that the short period data. 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 dry years or dry period. It may only be valid when using data with a length of several months and observed in rainy season or wet years. To get more general knowledge about when the observations are most informative for model calibration, more researches similar to our studies should be conducted. Based on our findings, the relationship between wetness level of short period data (i.e., the records were observed in a wet year or dry year and in rainy season or dry season) and their effectiveness for parameter calibration are worthy to be explored in this kind of future studies.

The manuscript has been revised accordingly.

(2) In your responses to reviewer #1, as is concluded that the good model performance for the validation period "may come from the fact that, the three validation years (2006, 2007, and 2008) are all wet years. And two of the three calibration years (2003, 2005) are also wet years. Calibration data may contain sufficient information for parameters identification in wet years." If in this case, the model is useful only when the calibration period and validation period both are wet years or both are dry years. What can we do when only one year's data is available for a data-scarce watershed? We cannot tell if this year is as the same condition as the following/previous years. Therefore, the objective of this study "how the use of limited continuous daily streamflow data might support the application of a physically-based distributed model in data-sparse basins" is not well explained.

Response:

With the two new cases, we have gained deeper insight about the questions addressed by the editor and reviewer #1. In the two wet basins, either using 1-year data of wet year or dry year could achieve performance similar to calibration using three year data in the validation period, which contains both dry years and wet years in the new Donghe Basin case and contains two dry years in the Jinjiang Basin case. For the two dry basins, one wet year data could also work well in validation period (also covers both wet years and dry years). However, one dry year data performs worse than data of wet years. Especially, for the case of Yanlongjiang Basin, model performances of validation period corresponding to extreme dry year 2006, 2007 deviate from calibration using three-year data to an unacceptable extent, i.e. fail to reproduce streamflow in validation period. In summary, one wet year data are useful to reproduce streamflow in both dry and wet basins. One dry year data in wet basins also can works well. But one dry year data in dry basin have the high possibility that it cannot calibrate the model effectively. In such context, to know whether the year is a wet year or a dry year from the annual streamflow frequency is useful to judge the effectiveness of 1-year calibration data. In the case of only one year data is available, it is impossible to obtain yearly streamflow frequency directly. Precipitation data are much easier to be obtained than streamflow data, either from in situ gauging or remote sensing data. Nowadays, many precipitation data product with wide spatial and temporal coverage are available, from which precipitation frequency can be computed. As streamflow is generated from precipitation, it is reasonable to use yearly precipitation frequency in the simulated basin as a surrogate of yearly streamflow frequency to judge whether one specific year is a wet year or dry year.

The manuscript has been revised accordingly to enhance how our findings could support the application of this method in data-sparse basins.

Technical questions:

1) I don't agree that "However, none of these observations has the capability of streamflow data for constraining hydrological model parameters." The soil moisture, ground water level can also be used in hydrological calibration and constraining model parameters (e.g., Immerzeel and Droogers. 2008. Calibration of a distributed hydrological model based on satellite evapotranspiration. Journal of Hydrology, 349(3): 411-424.).

Response:

What we want to express here is that the other type of the observations are useful for model calibration, but they are not as useful as streamflow data. The sentence has been revised to avoid confusion.

2) Through I agree with you that "it is difficult to apply measured values to hydrological models directly", I don't think these parameters are conceptual without explicit physical meaning (e.g., CH_K2, SOL_AWC, SOL_K). They have explicit physical meanings.

Response:

We agree with this comment, the three parameters mentioned by the editor have explicit physical meanings.

3) Check the language throughout this manuscript. For example, Line 14: period(s); Line 22: year(s)…

Response:

Following this comment, the language has been checked to avoid grammatical errors.

4) What are the criteria to distinguish a good model performance? How to tell the model performance (e.g., calibrated using one-month/three-month streamflow data) is as good as the one calibrated using 3-year? The meaning of U combining the P_factor with the R_factor should be clarified. For example, the model performance of "one year 2003 or 2004" is not superior significantly than that of "one month" in Table 5.

Response:

After considering this comment carefully, we realize that our idea of combing P_factor and R_factor to make a single index to quantify the simulation uncertainty is not proper, as the new index U has not explicit meaning and is hard to understand and misleading. In the revised manuscript, the simulation uncertainty will be quantified by P_factor and R_factor directly. P_factor stands for the percentage of observations covered by the uncertainty band, which ranges from 0 to 1. R-factor stands for the relative width of the uncertainty bands, which ranges from 0 to infinity. Higher R-factor combined with lower P_factor represents lower simulation uncertainty.

In the revised manuscript, the performance of each calibration is quantified by three indexes computed for the validation period: the Nash-Sutcliffe Efficiency (NSE) coefficient of simulated streamflow corresponding to the best performed parameter set in the calibration period, the P_factor and the R_factor. We put more weight on NSE and P_factor, as they are most explicit and values are easy to be understood. After NSE and P_factor, then R_factor will be considered and less weight are put. Under such philosophy, the concern about the model performance of "one year 2003 or 2004" is not superior significantly than that of "one month" in Table 5 of Heihe Basin in the original manuscript is answered as follows:

As shown in Figure 10 of the revised manuscript, for the validation period the NSE of "one month"(0.69) is lower than "2003"(0.78) and "2004"(0.72). And the P_factor of "one month"(0.27) is also lower than "2003"(0.65) and "2004"(0.51). Therefore, "2003" and "2004" are superior to "one month"

This strategy of evaluation has been used in the new manuscript and revisions have been made accordingly. In the revised manuscripts, instead of tables, figures are employed to show the result of calibrations more concisely.

6) About the GLUE equation, your approach calculates the posterior distribution at each point first and then gets the 2.5% and 97.5%, while reviewer's approach computes the 2.5% and 97.5% directly. I think both methods are correct.

Response:

Thank you very much for clarifying this issue.

(7) Location of these two watersheds should be identified in Fig. 1. And Figure 1 and Figure 2 can be merged into Figure 1. The hydrological stations used in this study should be included.

Response:

The locations of basins being studied in the research has been shown in Figure 1 of the revised manuscript. Also in this figure, for each basin, a map showing the hydrological stations being simulated has been included.

(8) Figure 6: the unit should follow "HESS author instruction".

Response:

The unit in Figure 5 and 6 (Figure 7a and 7c in the new manuscript) has been revised according to HESS author instruction.

Responses to reviewer #1

The paper shows interesting results on distributed hydrological model calibration, in which the authors demonstrate that the SWAT model can be satisfactorily calibrated using 1-6 month daily discharge observation, that is much shorter than normally used for calibration. It can be a large contribution to hydrological modeling for ungauged or poorly gauged basins where long term observation is not available.

There are two comments and recommendations:

1. The major point of this paper is that a hydrological model can be successfully calibrated even based on a short term observation and wet conditions for both period and basin are preferable for effective calibration. It may be true but I wonder if it happens by chance. The authors discuss meteorological conditions of calibration years (2005-2007 for Jinjiang basin and 2003-2005 for Heihe basin) but do not discuss the conditions of validation years. If the study basins were wet in validation years, it is quite reasonable that short observation for wet period can provide successful calibration, while it is truly surprising if it can provide a good result even for the case that validation years are dry. I would like to recommend the authors to add plots for validation years to cumulative distribution shown in Figs 5 and 6 and discuss more about the conditions of validation years in relation to the conditions of calibration periods.

Response:

Two more basins, one wet and one dry, are added to the study, in order to future prove the effectiveness of our approach. And the results from the two new basins are quite similar to the two old ones, indicating our conclusions are solid. From the finding of the four basins, it is shown that when using one-year data for calibration, for the two wet basins, whether the year is wet or dry have little influences on model calibration, i.e., their information content for model calibration are at similar level as using three-year data. However, in the two arid basins, the influence is significant. Data from wet years have better performances.

The validation years are also added to the cumulative distribution as shown in Figure R1 and R2. For the Heihe Basin (Figure R1), the validation years (2006, 2007 and 2008) are all wet years. Generally, as shown in many studies, model performance in the validation period will decrease, compared with calibration period. However, for the benchmark calibration of the Heihe basin, the performance between validation and calibration period does not show much difference, as shown in Figure 10 of the revised manuscript. This phenomenon may come from the fact revealed by Figure R1 that, the three validation years (2006, 2007, and 2008) are all wet years. And two of the three calibration years (2003, 2005) are also wet years. In this case of dry Heihe Basin, calibration data may contains more information for parameters identification in the data of wet years (2003, 2005) than in the data of dry year (2004).

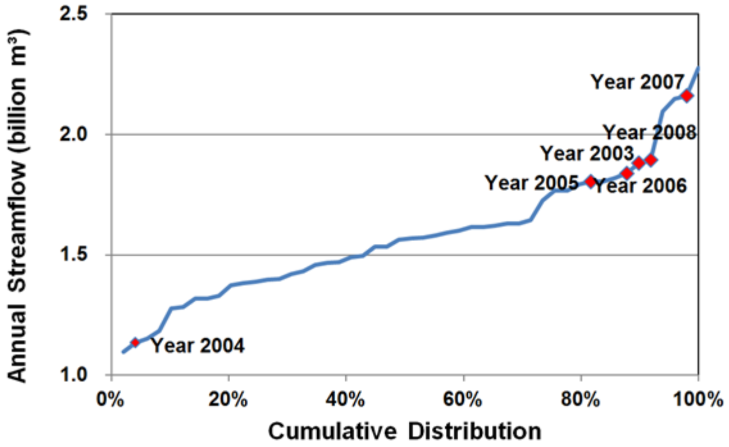


Figure R1 Cumulative distribution of annual streamflow of Heihe Basin (Yingluoxia station) for the period of 1960 to 2008

For the benchmark calibration in the Jinjiang Basin, compared with calibration period, model performance in the validation period decreases. Figure R2 revealed that the two validation years (2008, 2009) are dry years. The three calibrations years is average (2005) and wet year (2006, 2007). The decrease in model performance is consistent with other studies (e.g. Todorovic and Plavsic 2016), in which model efficiency also decrease when calibration period is wetter than validation period. When using only one year data for calibration, the performances of these three single year in validation period are similar to the three-year data used in benchmark calibration.

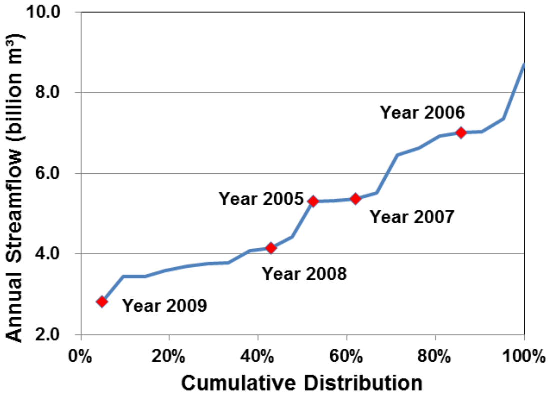


Figure R2 Cumulative distribution of available annual streamflow at Jinjiang Basin (Shilong station) for the period of 1958 to 2009

2. I assume that the wet period of the Heihe basin is from April till September and expected that the calibration based on the six months from Apr. 2004 till Sep. 2004 was capable of giving a good result, but Table 5 shows that no behavioral parameter sets were obtained in this period although many behavioral sets were obtained for the two dry periods from 2003 to 2004 and from 2004 to 2005. This is different from the tendency that is found from other calibrations. It would lead to deeper understanding if the authors could give clear explanations for this exceptional case.

Response:

The Heihe basin is an inland basin in the arid northwest China. Most rainfall occurs in the summer season. About 75% of total annual streamflow come from the wet period from April to September. As shown in Figure R1, 2004 is an extremely dry year. Compared with normal year, either in wet season or in dry season of 2004, the average streamflow decreased. Considering the big contribution to total annual streamflow, the degree of decreases of streamflow in the rainy season has high possibility to be bigger than the dry season. The runoff generation mechanism in this wet season with extremely low streamflow is very different from normal situation which made the model cannot capture the essence of variation in streamflow, therefore none of the randomly generated 10,000 parameter sets can reproduce the hydrograph of this wet season with acceptable accuracy. This is our explanation about why no parameter set was identified as behavioral ones using data of wet season in 2004 and it has been added to the revised manuscript.

References:

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 (1), 16-28, 2016

Response to reviewer #2:

Two watersheds is not enough to conclude your study provides general conclusions! There are groups that use thousands of watersheds, look up large sample hydrology, for example: http://meetingorganizer.copernicus.org/EGU2015/session/18271
http://www.hydrol-earth-syst-sci.net/18/463/2014/hess-18-463-2014.html

Responses:

To make conclusions from this study more general, two more basins, a wet one and a dry one are added to this study and used for evaluating the proposed method. The findings from the new basins are quite similar to the two old ones, indicating the conclusions drown by us are generally solid.

The contribution of the paper is not clear. Such analysis on the quality of calibration data dates back to 1996 (http://www.sciencedirect.com/science/article/pii/0022169495029184) and republication in HESS is not justified!

Response:

The model used in Yapo *et al.*(1996) is a conceptual rainfall-runoff model. In comparison, our study focuses on *physically-based distributed hydrological models*. This is one major difference between our study and Yapo *et al.* (1996). These two types of models face the same problem: lacking of streamflow data for calibrations in ungauged basins. For conceptual rainfall-runoff models, effectiveness of using short period of streamflow data to calibrate the model has been demonstrated in several papers, such as the one Yapo et al.(1996) mentioned by the reviewer, also Perrin et al.(2006), Beven and Tada (2012). However, based on our knowledge, such approach has not been widely tested for physically-based distributed hydrological models. The contributions of this paper are:

Firstly, it is demonstrated that using short period of streamflow data has the possibility to be useful for calibrating physically-based distributed hydrological models, like conceptual model. In the two wet basins, it is possible to only using 1-month data to calibrate the model effectively. In the two dry basins, using 6-month data is possible. These results are different from the common understanding that data of several years are needed for calibrations of such model. The paper could inspire more researchers to think about using such dataset to calibrate distributed hydrological models in basins lacking of streamflow data and test it in more basins. In the past, this approach didn't draw much attention for solving the calibration problem of distributed model in ungauged basins.

Secondly, this study is valuable for evaluating the effectiveness of short period data for model calibration in real world application. Our results show that, the phenomenon that some parameter sets are identified behavioral ones based on the comparison between simulation and observations could be considered as one evidence for making the judgement that the short period data. 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 dry years or dry period. It may only be valid when using data with a length of several months and observed in rainy season or wet years. To get more general knowledge about when the observations are most informative for model calibration, more researches similar to our studies should be conducted. Based on our findings, the relationship between wetness level of short period data (i.e., the records were observed in a wet year or dry year and in rainy season or dry season) and their effectiveness for parameter calibration are worthy to be explored in this kind of future studies.

There is no figure provided of how calibration of SWAT with limited data translates into model simulation! How do I know 1 month of data is enough for calibration if I don't see how the model works graphically? NSE is certainly not enough!

Response:

Thank you for this suggestion. For the Jinjiang Basin, a figure showing the simulated hydrograph corresponding to the calibration using data of one month will be added to the revised manuscript to demonstrate the effectiveness of using such short period of data for calibration. The figure is also shown here (Figure R3 below). For the validation period (2008 to 2009), it is shown that the best simulations of ensemble predictions corresponding to the calibrations using three year data (2005 to 2007) and one month data (July 2006) is quite similar visually. As a response to the concern about NES, we computed the Mean Absolute Error (MAE) of best simulations in low flow period (September to next March) and high flow period (April to September). In high flow period, the MAE of the best simulation corresponding to calibration using three-year data and one-month-data is 66.3 m^3/s and $63.3m^3/s$, respectively. In low flow period, the MAE of the two best simulations is $36.4m^3/s$ and $43.1m^3/s$, respectively. Generally, similar performance level is achieved by the two best simulations. These results could support our conclusion that in the Jinjiang Basin, it is possible that one month data is informative to calibrate the SWAT model effectively.

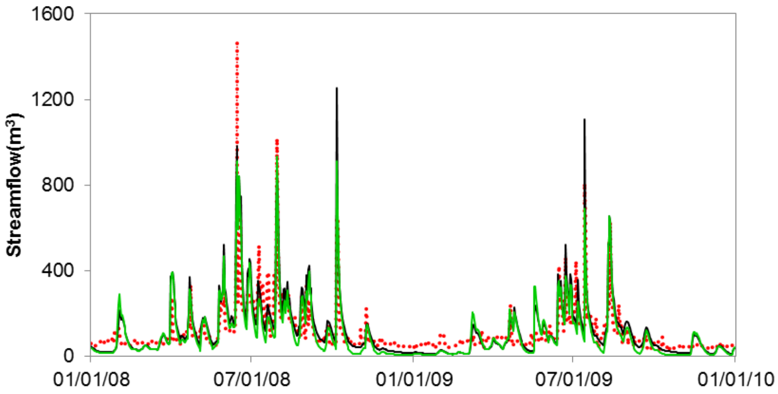


Figure R3 Observed streamflow of 2008 to 2009(dashed red line), best simulations of ensemble prediction for the validation period (2008-2009) corresponding to calibration based on streamflow data of 2005 to 2007 (solid black line) and July 2006(solid green line) in the Jinjiang Basin.

Page 2, lines 1-5: I don't agree with your statement that models like SWAT are able to predict droughts and floods! Droughts and floods respond to climatic forcings and climatic models are used to forecast them, certainly not SWAT!

Response:

As mentioned by the reviewer, hydrological models, like SWAT, use climatic data as forcing data to predict streamflow in the river. When climatic forecasting are available, hydrological models can employ it as input and predict hydrological drought and flood, from the perspective of quantity of streamflow.

Page 2, line 8-9: "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". This refers to conceptual models mostly. Physically based distributed are supposed

to have parameters with clear physical meaning, that can ideally be measured in the field.

Response:

We agree with the reviewer's opinion. 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, such measurements require intensive field survey, which are not available in most researches. Therefore, usually model parameters of physically-based model are also obtained from model calibration based on streamflow data. This explanation has been added to the manuscript.

"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 20 et al., 2002), evapotranspiration (Vervoort et al., 2014; Winsemius et al., 2008), groundwater level(e.g., Khu et al.,2008)." My understanding is that since streamflow measurements are not available, one can alternatively use other variables such as soil moisture, ground water table and evapotranspiration as calibration data. This is certainly not the case, since measuring these variables is much more difficult and costly than streamflow. I suggest you phrase your sentences more carefully to avoid such confusions.

Response:

Our intention is not to make new observations of other variables of hydrological cycles, but to make best use of available data of such variables. There are cases that in some basins, streamflow data is unavailable, but observation data of the other variables are available. In such situation, the available data maybe valuable for model calibration. To avoid confusions, these sentences will be revised based on above understanding.

Page 3, line 4: What do you mean by "changing environment"?

Response:

This term of "changing environment" comes from the paper of Montanari et al. (2013), which introduces the IAHS Scientific Decade 2013–2022 " Panta Rhei—Everything Flows". It means the all factors that influence hydrological system, including both changes in nature (e.g., climate changes) and society (e.g., global population growth).

Page 3, lines 4-6: You argue "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)." I understand that some physical modelers would make such arguments, but it is certainly a debated issue, so I wouldn't make such strong claims. This being said, in a changing climate, even physically based models are not proven to be working properly. The argument made by developers of physically based models is that since they use specific description of the watersheds, their models can handle land-use change (change of physical characteristics of watersheds). This also needs a lot of research still.

Response:

We agree with the reviewer's opinion about the comparison between conceptual and physically based hydrological model. The sentences will be revised in the new version of manuscript to express our understanding more precisely.

Equation 2 is all WRONG! You want to use an objective function of NSE, do, but you can't call it a likelihood function and use it as in the Bayes theorem! There is no scaling in Bayes law! You may call this weight, but not posterior likelihood.

Response:

For GLUE, the term of likelihood is used in a very general sense, as described by developer of GLUE, Keith Beven, in the paper of Beven and Binley (1992): the likelihood function in GLUE works as a fuzzy, belief, or possibilistic measure of how well the model conforms to the observed behavior of the system, and not in the restricted sense of maximum likelihood theory. The likelihood measure quantifies the difference between simulation and observations. The only requirement for a likelihood measure is that it should be assigned as zero for all parameter sets that cannot reproduce the observations and should increase monotonically as the performance rises. Based on the theory of GLUE, in our opinion, using NSE as a likelihood function is proper in our study. Also the Bayes equation is employed in GLUE in a general sense. Equation 2 has been used in many studies related to GLUE, such as Freer and Beven (1996), Beven and Freer (2001).

I have a hard time with equation 3 also! Weights (or as you call them posterior likelihoods) are calculate based on overall performance of the model (t=1:N), but are used at each time step to estimate the cumulative probability of streaamflow. This is not right! You want to estimate the 95% uncertainty range, take the 2.5 and 97.5th percentiles of your streamflow simulation ensemble.

Response:

For application of GLUE, the posterior likelihood is computed based on the overall model performance in the period when observations are available to compare model behavior with observations. Then at each time step, the posterior likelihood is used to estimate cumulative probability of streamflow. This is the way how GLUE estimates simulation uncertainty, and equation 3 can be found in many studies using GLUE.

Page 6, lines 4-6: I don't understand this sentence, and it is critical to evaluate the methodology proposed in this paper: "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."

Response:

What we try to express is that, 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 number of continuous daily observations could achieve similar performance as benchmark calibration, not to determine whether all records with that specific length can calibrate the model effectively..

This explanation will be added to the revised version of paper.

The entire section 2.4 is poorly written and methodology is poorly described, making it really difficult to assess the paper.

Response:

This section has been rewritten in the new manuscript to make it easier understood.

Page 7, line 2: You admit that it is very time consuming to calibrate SWAT using GLUE. Why not using more intelligent calibration approaches like Markov Chain Monte Carlo? It has been shown in the literature that MCMC is orders of magnitude more efficient than GLUE.

Response:

For daily-step model simulation of several years, it is common for lumped conceptual models to finish the simulation within one second. However, for distributed model like SWAT, usually, one such model run may need several minutes. We have no information about prior distribution of model parameters, based on which, parameter sets will be generated randomly. The uniform distribution is used as the prior distribution of each parameter. In such situation, implementing GLUE with Latin hypercube sampling is usually preferred. This strategy has been used in many literatures related to GLUE and SWAT simulation.

Page 7 line 11: "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." I tend to disagree! It depends on your performance metrics, if you use NSE which is sensitive to peak flows, then yes you are right! But if you use metrics such as baseflow index this is not going to hold. Different processes of a model are activated under different forcings, so you can't simply ignore several processes and focus on the one (few) process(es) that are activated at the wet condition.

Response:

We agree with the reviewer that the statement of "because most model modules are activated in high-flow period" is improper and it will be deleted from the manuscript.

Page 7 line 24: Uncertainty bound not band! Correct throughout the manuscript.

Response:

Uncertainty bounds are two lines consisted of the 2.5% and 97.5% quantiles of simulated streamflow in each time step. The uncertainty band is the area bounded by these two lines in hydrograph. Both term of "uncertainty bound" and "uncertainty band" (e.g., Beven and Benley, 1992; Yang et al., 2008) can be used when applying GLUE for uncertainty analysis.

Page 8, lines 2-3: "For the 1-year period, all three calibrations performed similarly to the benchmark calibration, and the dataset for 2006 even outperformed the benchmark" It is interesting and concerning that a shorter calibration period provides a higher performance. It requires explanation as to how it happened! You can't just leave it like that which might spuriously suggest smaller calibration period is sometimes even better! Here are my thoughts: 1. You are not using a consistent period to evaluate your model! 2. Your calibration approach did not converge to the right posterior distribution (as might happen with GLUE) 3. Your data includes some misinformation, meaning not only it doesn't provide any good information to constrain model parameters, but also it misguides the model! In the cases of 1 & 2, extra data can only be redundant and cannot deteriorate the performance of the model!

Response:

The evaluation of each calibration is based on judging model performance in the validation period. The validation period is made to be same for all calibrations in each basin. For the Jinjiang Basin, the validation period is 2008 to 2009. For the Heihe Basin, the validation period is 2006 to 2008.

Streamflow data are obtained from the water administrative department in local government and have used in many studies. The data quality is guaranteed. We apologize that the statement of “the dataset for 2006 even outperformed the benchmark” is misleading. It is only based on the NSE of best performed behavioral parameter set. Another important aspect of the evaluation is simulation uncertainty. It is quantified by the index of “U” as defined in the original manuscript. The U was computed from the P_factor (percentage of observations embraced by the 95% prediction intervals) and R_factor (a measure of the average width of 95% simulation intervals). Our intention of using U was trying to describe simulation uncertainty using one single index. However, after carefully consideration, we feel that the meaning of U value is confusing and decide to using P_factor and R_factor directly to describe simulation uncertainty. As shown in the Figure 6 of the revised manuscript, it is indicated that the simulation uncertainty of calibration using data of 2006 is a little bit higher than benchmark calibration. So we agree that the statement of “the dataset for 2006 even outperformed the benchmark” is improper and it is not our intention to conclude that using shorter period of calibration is better than using long period of observations. The method presented in this study is only expected to be useful for model calibration in data-sparse basins where streamflow data of several years are unavailable. The statement of “the dataset for 2006 even outperformed the benchmark” will be deleted.

Page 8, lines 14-16: “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.” This claim is rather strange to me! One month is enough to capture all the processes? Some processes might not even be activated in one month! Again, this is because you focused all your attention on NSE, and what is most important in NSE is the high peaks. So if you activate the processes that reproduce the high peaks, you get a good performance. This doesn’t mean one month is enough to calibrate a model!

Response:

We apologize that we didn’t give enough details for the evaluations of calibration using 1-month data in the manuscript. As our responses to previous comments, from simulated hydrograph and all model performance indexes (NSE, P_factor and R_factor, MAE for both low and high flow period) in the validation period, it is indicated that the calibrated model corresponding to 1-month calibration data performs similar to the benchmark calibration.

We realize that our expression about the results using 1-month data is too strong and confusing. In the revised manuscript, instead of the original expression, we will conclude that it is possible that 1-month’s continuous daily observations can contain much of the information content of 3-year continuous streamflow data for model calibration.

References:

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.

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.

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.

Montanari, A., Young, G., Savenije, H.H.G., Hughes, D., Wagener, T., Ren, LL., Koutsoyiannis, D., Cudennec, C., Toth, E., Grimaldi, S., Blöschl, G., Sivapalan, M., Beven, K., Gupta, H., Hipsey, M., Schaefli, B., Arheimer, B., Boegh, E., Schymanski, S.J., Baldassarre, G.D., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D.A., Srinivasan, V., Harman, C., Thompson, S., Rogger, M., Viglione, A., McMillan, H., Characklis, G., Pang, Z., Belyaev, V.: "PantaRhei—Everything Flows": Change in hydrology and society—The IAHS Scientific Decade 2013-2022. Hydrolog. Sci. J., 58(6),1256-1275, 2013.

Perrin, C., Oudin, L., Andreassian, V., Rojas-Serna, C., Michel, C., and Mathevet, T.: Impact of limited streamflow data on 30 the efficiency and the parameters of rainfall-runoff models, Hydrolog. Sci. J., 52, 131–151, 2007.

Tada, T. and Beven, K. J.: Hydrological model calibration using a short period of observations, Hydrol. Process., 26, 883-892, 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. and Gupta, H. V. and Sorooshian, S.: Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data, J. Hydrol., 181, 23-48, 1996.

List of Changes

1. Two more case studies are added to the manuscript. The introduction about the two basins is added to Section 2.3. And the results of calibrations in these two basins are in Section 3.3 and 3.5
2. The contributions of this study to future studies are discussed and added the Section 3.6.
3. The Section of 2.4, experiment design, has been rewritten.
4. The evaluations of simulation uncertainty are carried out using two indexes, P_factor and R_factor, instead of U used in the original manuscript.
5. The annual streamflow cumulative distributions of the four basins are shown in Figure 7. And discussion about the relationship between wetness levels of calibration years with model performance in validation years are added to the Section of results and discussion.
6. Instead of tables, the results of calibrations in each basin are shown in the form of figures (Figure6, 9, 10 and11).
7. The locations of 4 studied basins in China are shown in Figure 1. And a table (Table 1) showing the characteristics of the four basins is added.
8. The comparisons of simulation results between calibration using 3-year data and using 1-month data in the Jinjiang Basin are added to the Section 3.2.
9. Abstract and conclusions are revised according to the revisions in the results and discussion part,
10. The grammar of the manuscript has been checked and corresponding revisions are made.

All the other revisions suggested by the editor and reviewers are specified in the responses to the comments in the previous section of this pdf file.

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], Guoqiang 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, a~~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. However, unlike lumped conceptual model, such an approach hasn't 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 observation period on the calibration of the widely applied Soil and Water Assessment Tool model was evaluated in ~~two~~ 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 ~~for~~ of the ~~two~~ 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 ~~both~~ 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 wet Jinjiang Basin and Donghe Basin could be effectively calibrated using a shorter data record (1~~1~~ month), compared with the ~~arid~~ dry Heihe Basin and upstream Yalongjiang Basin (~~6~~ 6 months). Even though the ~~two~~ four basins are very different, when using 1-year or 6-month (covering a whole dry season or rainy season) data, -the results ~~demonstrated~~ show that data from ~~the~~ wet seasons and wet ~~ter~~ years ~~performed better~~are generally more reliable ~~that~~ than data from the dry seasons and dry~~ier~~ years, especially for the two dry basins. The results ~~of this study~~ demonstrated that ~~short periods of observations~~this idea could be a promising ~~solution~~ approach 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~~findings from the discussion in this study are valuable for assessing the effectiveness of short period data for model calibration in real world application.

带格式的: 字体: (中文) Times New Roman, 小四, 上标

带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符

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 withi n a basin. ~~Most~~ Many 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). For physically-based model, 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 survey, which are not available in most researches. Therefore, usually parameters of such model are 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, ~~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~~introducing ~~of~~ limited information from ~~the~~ 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 ~~f~~tried~~ocused~~ to~~on~~ use~~ing~~ 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), 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 has the similar capability ~~of~~ as 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

带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符

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, i~~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). For conceptual models, these researches 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~~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 been ~~addressed~~ discussed 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 how short-period of daily continuous streamflow observations might support the calibration of a physically-based distributed model in data-sparse basins. In real world, such observations ~~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~~ four basins with different climatic conditions (two~~one~~ in ~~a~~ humid regions and the other two in ~~an arid~~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 ~~adpoted~~adopted as the benchmark ~~simulation~~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) ~~method~~ (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 satisfying the requirements of the evaluation. By reducing the number of observations used in the model calibration in a designed manner

带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符

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:

带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符

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:

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

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:

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

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.

带格式的: 字体: (中文) +中文正文 (宋体), (中文) 中文(中国)

2.3 Study basins

We used 4 basins located in China (Figure 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 located in different climatic regions and characteristics of topography, annual precipitation and temperature are quite different.

带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符

~~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 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 ~~rangs~~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 warm dry winters and hot rainy summers. Annual precipitation ranges from 1000 to 1800 mm, most of which falls in summer. The hydrological ~~modeling~~modelling 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 1,089 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. Average annual discharge for the Dong River is 43.31m^3/s. The hydrological model was calibrated and validated by the streamflow data in Wenquan 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 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 ~~10~~8,~~000~~ 843 km^2. The elevation of the study area varies from ~~more than~~around ~~5000m~~ 4700m 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.

The Yalongjiang River is originated in the Tibetan plateau, which is the largest tributary to the Jinshajiang River in the upper Yangtze River. The hydrological modelling was conducted in the upstream region of Ganzi streamflow station, for which the elevation ranges from 3400 to 6021. The area of hydrological simulation by SWAT is 32,535 km^2. Plateau meadow 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 recent 50 years is about 520mm, 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 over a year.

In the four basins, most precipitation happens in summer. Based on annual precipitation, the four basins were divided into two groups. The Jinjiang and Donghe Basin are considered as representatives of wet basins. And Heihe and Yalongjiang Basin represent dry basins. For each basin, the influence of the observational record length on the calibration will be

带格式的: 上标

带格式的: 上标

带格式的: 上标

带格式的: 字体：(中文) +中文正文 (宋体), (中文) 中文(中国)

带格式的: 上标

带格式的: 上标

带格式的: 上标

带格式的: 制表位： 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符

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.

带格式的: 字体: (中文) +中文正文 (宋体), (中文) 中文(中国)

2.4 Experiment design

~~Based on~~ 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 period for the benchmark calibrations conducted in the four basins are shown in Table 2. ~~Heihe Basin, the calibration and validation periods were 2003–2005 and 2006–2008, respectively.~~ As an ~~first~~ initial trial for showing the potential of the method for distributed models~~the distributed model~~, we sought to explore ~~the highest possible performance~~whether ~~using~~ there are ~~certain short lengths of~~ records of certain short length or certain number of continuous daily observations could achieve similar performance as benchmark calibration, not to ~~the general performance of specific lengths~~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 above mentioned two issues, to perform the 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 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 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 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 calibration period shorter than six month. 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 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.

~~. 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.~~

带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符

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~~ m~~m~~ostly~~ y~~ 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~~ For each basin, in order to compare model performance of calibration using short period data ~~and compared~~ with the benchmark calibration in an objective manner~~.~~, t~~T~~he validation periods of these calibrations were made same with the benchmark calibrations~~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~~ identified by the calibration data) for the calibration and validation periods. Two indexes were ~~combined~~ utilized to assess the simulation uncertainty~~.~~: 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

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

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. 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 are put. ~~A new index *U* combining the *P_factor* with the *R_factor* was defined as:~~

~~(5)~~

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

~~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.~~

带格式的: 字体: 倾斜

带格式的: 字体: 倾斜

带格式的: 字体: 倾斜

带格式的: 字体: 倾斜

带格式的: 字体: 倾斜

带格式的: 字体: 倾斜

带格式的: 字体: 倾斜

带格式的: 两端对齐, 段落间距段前: 0 磅, 段后: 0 磅

带格式的: 字体: (中文) +中文正文 (宋体), (中文) 中文(中国)

带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符

~~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~~evaluating the method proposed in this study, the model robustness in the ~~two~~ 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 ~~2~~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 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 f~~For the Jinjiang Basin, too many parameter sets could result reasonable simulations, and therefore the threshold for likelihood was ~~raised~~ 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 to 5. The NSEs of the best simulation in four~~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~~in the four 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~~shown in Figure 6. 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~~ 7a presents the cumulative distribution of the available

带格式的：字体颜色：自动设置

带格式的：字体颜色：自动设置

带格式的：制表位：27.55 字符，左对齐 + 不在 22.57 字符 + 45.13 字符

annual streamflow for Shilong station from 1958 to 2009~~, ~~. For the three year that streamflow data used for benchmark calibration, 2006 is a very wet year and 2005 and 2007 are normal to wet years. The two year of 2008 and 2009 are all dry years. For the benchmark calibration, the model performance in the validation period decreases, compared with calibration period. The decrease in model performance is consistent with other studies (e.g. Todorovic and Plavsic 2016), in which model efficiency also decrease if calibration period is wetter than validation period. When using only 1-year data for calibration, the performances in validation period are similar to benchmark calibration.~~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, n~~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 other four records of 6-month could achieve similar performance as benchmark calibration. 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~~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 similar performance to benchmark calibration, judging from the indexes. Figure 8 shows that simulated streamflow of best performed parameter sets corresponding to the benchmark calibration, 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~~ Thus~~, 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~~ it is possible to calibrate the SWAT model effectively using only 1-month's continuous daily observations of streamflow~~.

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 five 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 three calibration years, as shown in Figure 7b, 2002 is a dry year. 2003 and 2004 are wet years. The two validation year, 2005 and 2006 is a wet year and

带格式的: 字体:(中文) Times New Roman

带格式的: 字体:(中文) Times New Roman

带格式的: 标题 2

带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符

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

带格式的: 字体：(中文) +中文正文 (宋体), (中文) 中文(中国)

3.4~~3~~ Evaluation of the Heihe Basin case

带格式的: 制表位： 30.75 字符，左对齐

The results of the calibration are shown in ~~Table 5~~Figure 10. 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~~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 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 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~~ fails to identify behavioral parameter sets. In the arid Heihe basin, most rainfall occurs in the summer season. About 75% of total annual streamflow come from the wet period from April to September. Compared with normal year, either in wet season or in dry season of in extremely 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 high possibility to be bigger than the dry season. The runoff generation mechanism in this wet season with extremely low streamflow is very different from normal situation, which made the model cannot capture the essence of variation in streamflow, therefore none of the randomly generated 10,000 parameter sets can reproduce the hydrograph of this wet season with acceptable accuracy. ~~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

带格式的: 制表位： 27.55 字符，左对齐 + 不在 22.57 字符 + 45.13 字符

levels of performance 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 in this arid basin.

3.5 Evaluation of the Yalongjiang Basin case

带格式的: 字体: (中文) Times New Roman

带格式的: 标题 2

The cumulative distribution of annual streamflow at Ganzi station(Figure 7d) indicate that, for the calibration period, 2005 is an extremely wet year, 2006 and 2007 are extremely dry years; for the validation period, 2009 is a wet year, 2008 and 2010 are dry years. Figure 11 indicates that, when using 1-year data for calibration, only the wet year 2005 could reach similar level of performance as benchmark calibration. The decreases in model performance when using data of dry year 2006 and 2007 are significant. At the temporal scale of 6-month, the diversity among datasets is high. The 6-month data of 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 extremely dry year of 2006 and 2007, respectively. Similar to the Heihe Basin case, it may be caused by the fact that the runoff generation mechanism in these period differing from 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 calibration period is satisfying, the calibrated model cannot reproduce the streamflow at validation period with acceptable accuracy. When using wettest 3-month, 1-month and 1-week data for calibration, no behavioral parameter set was identified, indicting three short period datasets cannot calibrated the model effectively.

带格式的: 字体: (中文) +中文正文 (宋体), (中文) 中文(中国)

3.~~4~~ 6 ~~Comparison of the two case studies~~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 only a dataset covering several months might contains sufficient information for parameter identification. This study clearly demonstrates the value of fragmentary historical records of past-existed gauging stations or temporary gauging during field surveys for calibrating physically-based distributed hydrological model in data sparse basin, at least for basins with a climate characterized by rainy or relatively rainy summer and dry winter and correspondingly streamflow exhibits an annual cycle of high flow and low flow. The paper could inspire more researchers to think about using such dataset to calibrate distributed hydrological models in basins lacking of streamflow data and test it in more well gauged basins to develop more general understanding about when the measurements are most informative for parameter calibration. In the past, this approach didn't draw much attention for solving the calibration problem of distributed models.

带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符

When applying the method to 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 calibration period, their performance in 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 evidence for making the judgement that the short period data is effective for model calibration. However, such judgement should be made with care. When the number of observations becomes lower, our results show that the possibility of good performance in calibration period accompanied by good simulation in validation period decreases. In most calibrations of the two wet basins, such judgement based on model performance in calibration period is valid. However, when the number of observations is too low (e.g., the calibration in Jinjiang Basin using 1-week observations), it may not valid. In the two dry basins, there are several calibrations showing that good performance in calibration period does not ensure good performance in validation period: when using 1-year data for calibration, performance of dry year data is inferior to wet year. In the Yalongjiang Basin case, the calibrations using 1-year data of dry year 2006, 2007 even fails to reproduce streamflow in 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 (Liden 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 calibration period is good, 6-month data from wet years or wet periods makes more reliable simulations in 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 should be preferred for model calibration.

These findings indicates 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 context, information about 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 exist 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 analyse annual precipitation frequency and intra-annual precipitation regime at basin scale, like the 5-km spatial resolution data used in the case of Donghe Basin. Information about the precipitation frequency can work as surrogates of annual streamflow

带格式的: 默认段落字体

带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符

frequency and intra-annual streamflow regime 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 record is sufficient to obtain robust parameter values, further researches similar to this one, using distributed model in large number of well-gauged basins with differing characteristics, is required. Such study need to generate many samples of short period observations from available streamflow data and then the samples are used 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 guideline on when the short period observations are informative for model calibration.

带格式的: 正文

~~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 extracted from available fragmentary historical observation records or obtained from field campaigns in the target basin~~ or extracted from available fragmentary historical observation records~~. It could be

带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符

considered a solution to the problem of ungauged basins in some situations. Through application of the SWAT model to ~~two~~ 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~~ Jinjiang~~ b~~B~~asins, it was demonstrated surprisingly that the model could be calibrated successfully using only a 1-month dataset, whereas in the two dry ~~Heihe~~ b~~B~~asins, longer datasets (6 months) were required~~.~~ ~~Interestingly, although the two basins are quite different, when using limited numbers of data for the model calibration,~~ and data from wet~~ter~~ years and ~~from~~ wet periods demonstrated ~~best performance~~more reliability than data from dry years and dry periods~~ in both cases~~.~~.~~ The results of this study clearly indicate the ~~value~~ potential of short-term streamflow observations in calibrating distributed hydrological models for ungauged basins. ~~However, i~~In real ~~applications~~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 that some parameter sets are identified behavioral ones based on the comparison between simulation and observations could be considered as one evidence for making the judgement that the short period data . 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 researches similar to our studies should be conducted. Based on our findings, the relationship between wetness level of short period data and their effectiveness for parameter calibration are worthy to be explored in this kind of future studies.~~ 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. 41671018, 91647202), 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.

带格式的: 制表位： 27.55 字符，左对齐 + 不在 22.57 字符 + 45.13 字符

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.

带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符

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.

带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符

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.

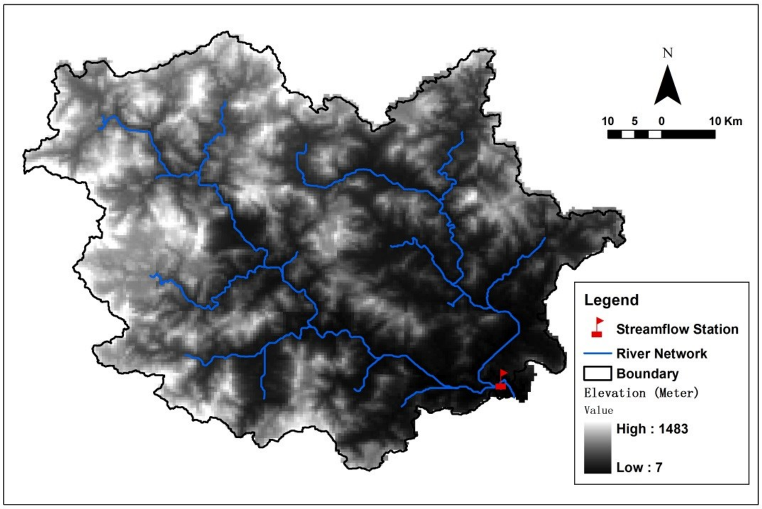
Wu, Y. P. and Liu, S. G. : Automating calibration, sensitivity and uncertainty analysis of complex models using the R package Flexible Modeling Environment (FME): SWAT as an example, Environ. Modell. Softw., 31, 99-109, doi: 10.1016/j.envsoft.2011.11.013, 2012.

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.

带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符

带格式的: 制表位： 27.55 字符，左对齐 + 不在 22.57 字符 + 45.13 字符



带格式的: 字体: 非加粗

带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符

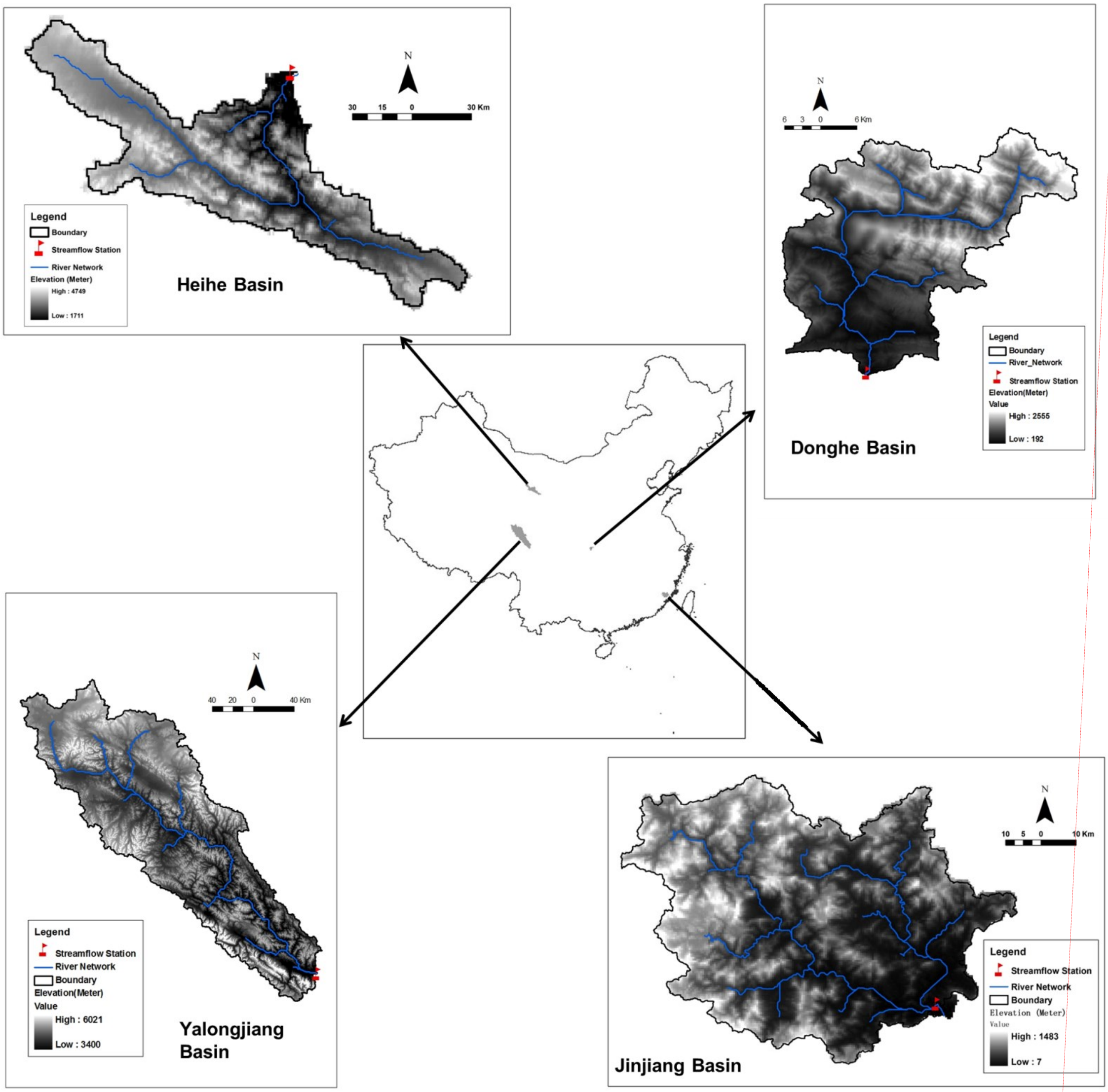


Figure 1~~1~~: Topography, river networks and the streamflow gauging stations of the four basins and their locations in China~~in the Jinjiang Basin, China~~

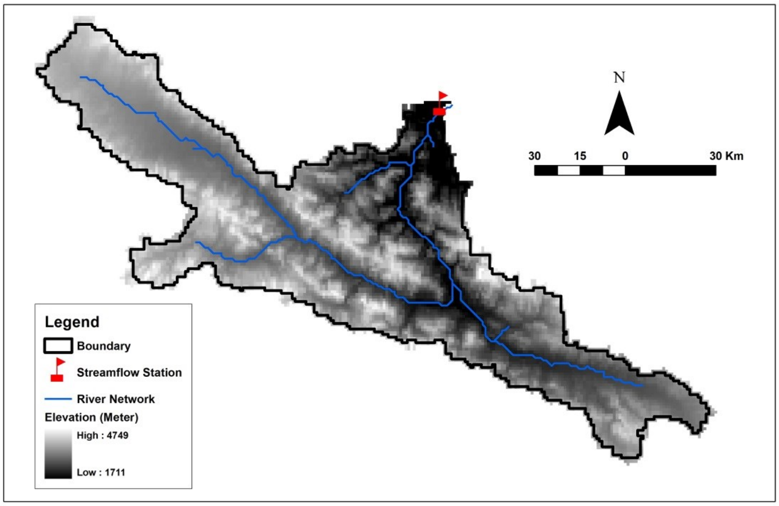
带格式的：字体：(中文) +中文正文 (宋体), (中文) 中文(中国)

带格式的：字体：加粗

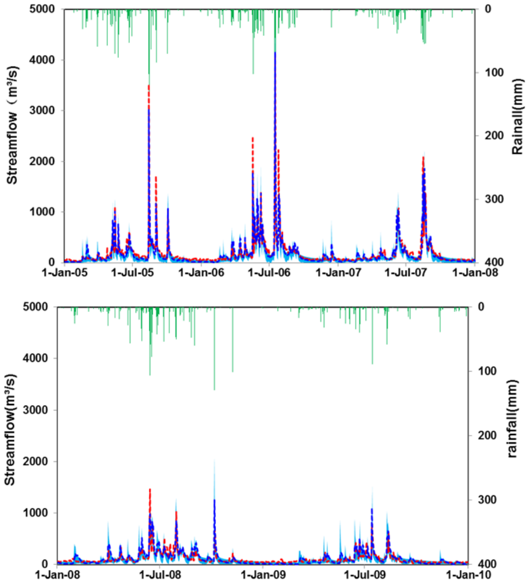
带格式的：字体：(中文) +中文正文 (宋体), 10 磅, (中文) 中文(中国)

带格式的：制表位：27.55 字符，左对齐 + 不在 22.57 字符 + 45.13 字符

带格式的：字体：非加粗



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



带格式的：制表位： 27.55 字符，左对齐 + 不在 22.57 字符 + 45.13 字符

Figure ~~3~~2: 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

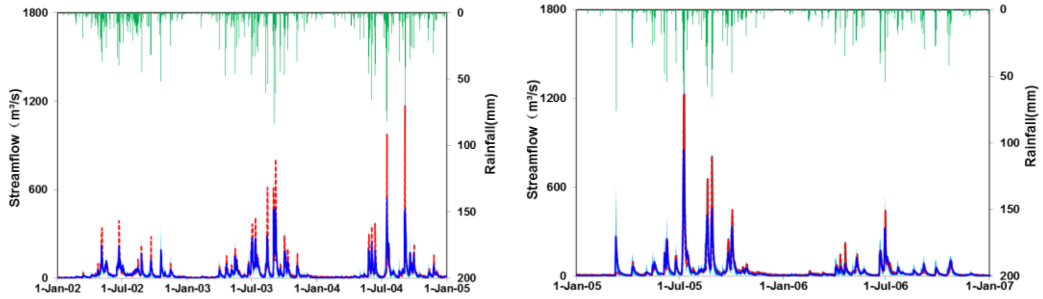


Figure 3: Simulated streamflow for the benchmark calibration of the Donghe Basin in both calibration (2002–2004) and validation period (2005–2006). Dashed lines: observed streamflow; blue band: 95% uncertainty band of ensemble simulation; solid blue lines: best simulation of ensemble prediction; green columns: rainfall

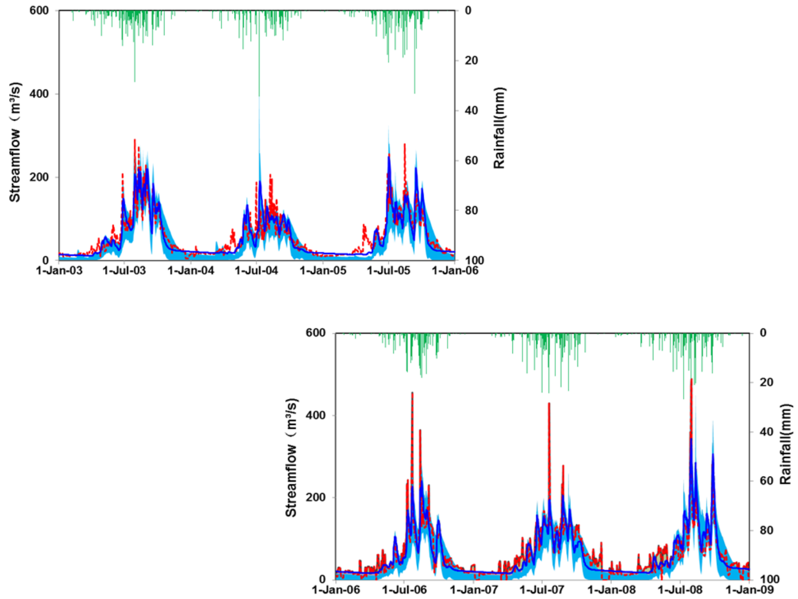


Figure 4: Simulated streamflow for the benchmark calibration of the Heihe Basin in both calibration (2003–2005) and validation period (2006–2008). Dashed lines: observed streamflow; blue band: 95% uncertainty band of ensemble simulation; solid blue lines: best simulation of ensemble prediction; green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

带格式的：字体：加粗

带格式的：字体：10 磅

带格式的：正文，居中

带格式的：字体：10 磅

带格式的：居中

带格式的：字体：加粗

带格式的：字体：10 磅

带格式的：制表位： 27.55 字符，左对齐 + 不在 22.57 字符 + 45.13 字符

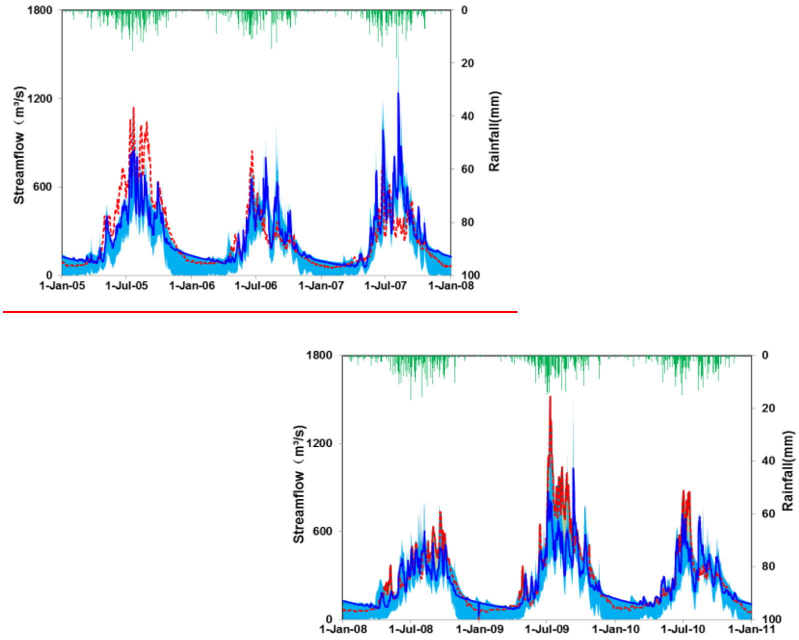


Figure 5: Simulated streamflow for the benchmark calibration of the Yalongjiang Basin in both calibration (2005–2007) and validation period (2008–2010). Dashed lines: observed streamflow; blue band: 95% uncertainty band of ensemble simulation; solid blue lines: best simulation of ensemble prediction; green columns: rainfall

带格式的: 正文

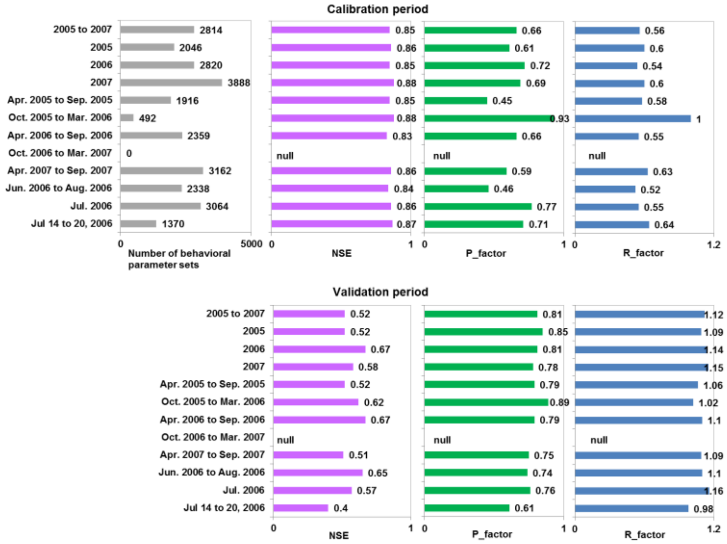
带格式的: 居中

带格式的: 字体: 加粗

带格式的: 字体: 10 磅

带格式的: 居中

带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符



带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符

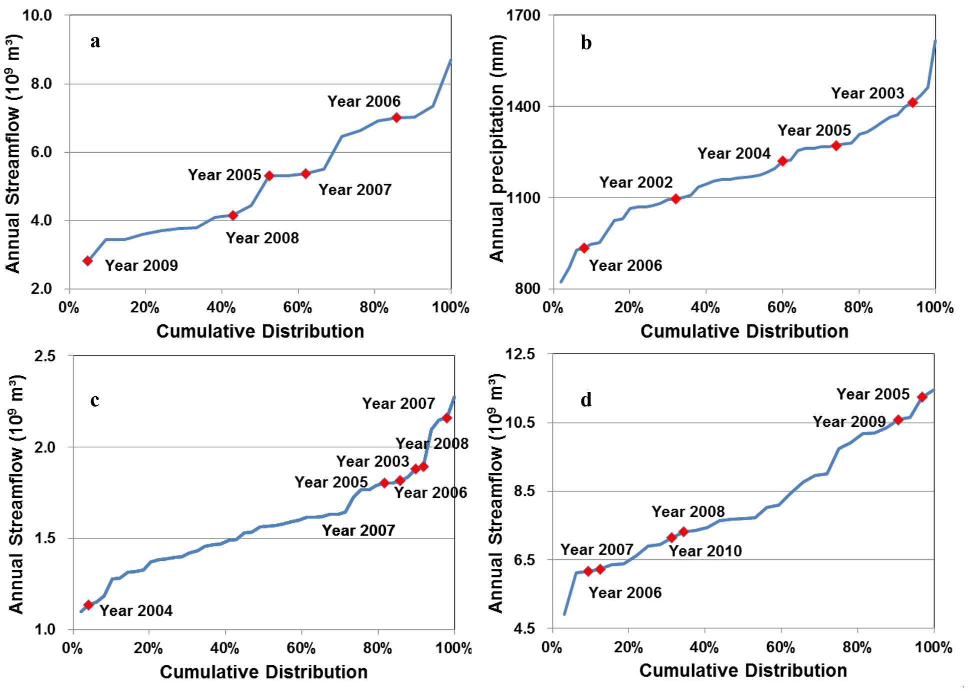


Figure 6: Model performance for the calibrations using short-period data in Jinjiang Basin

带格式的：字体：加粗

带格式的：字体：非加粗

带格式的：制表位： 6.49 字符，左对齐

带格式的：制表位： 27.55 字符，左对齐 + 不在 22.57 字符 + 45.13 字符

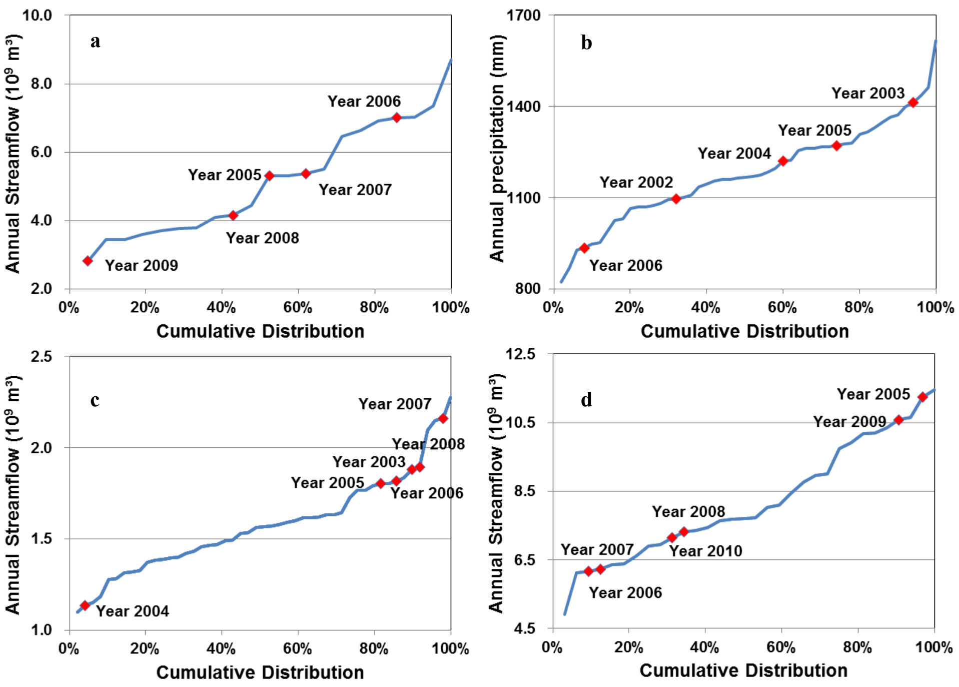
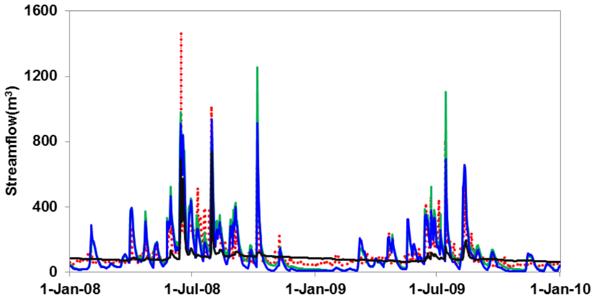


Figure 7: (a) Cumulative distribution of annual streamflow in Jinjiang Basin (at Shilong station) for the period of 1958 to 2009. (b) Cumulative distribution of annual precipitation in Donghe Basin (at Wenquan station), for the period of 1961 to 2010. (c) Cumulative distribution of annual streamflow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



带格式的: 字体: 加粗

带格式的: 字体: 加粗

带格式的: 正文

带格式的: 字体: (中文) +中文正文 (宋体)

带格式的: 居中

带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符

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, green line), one month(July 2006, blue line), one week(July 14 to 20, 2006, black line) and in situ observations (red dashed line).

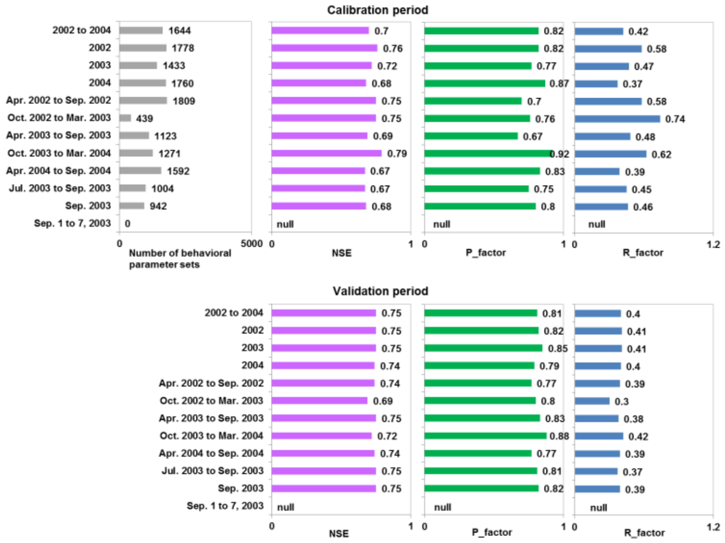


Figure 9: Model performance for the calibrations using short-period data in Donghe Basin.

带格式的: 字体: (中文) Times New Roman, 10 磅, 英语(英国)

带格式的: 题注

带格式的: 字体: (中文) Times New Roman

带格式的: 字体: (中文) +中文正文 (宋体)

带格式的: 居中

带格式的: 字体: (中文) Times New Roman

带格式的: 题注

带格式的: 字体: 非加粗

带格式的: 字体: (中文) Times New Roman

带格式的: 左

带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符

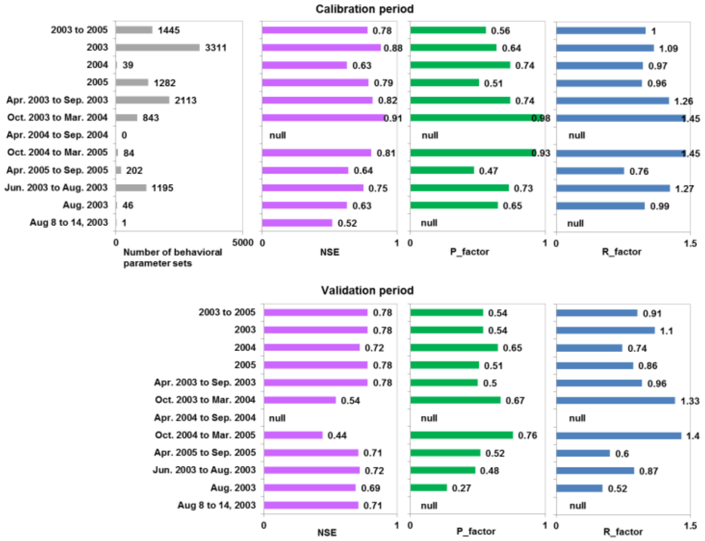


Figure 10: Model performance for the calibrations using short-period data in Heihe Basin.

带格式的：居中

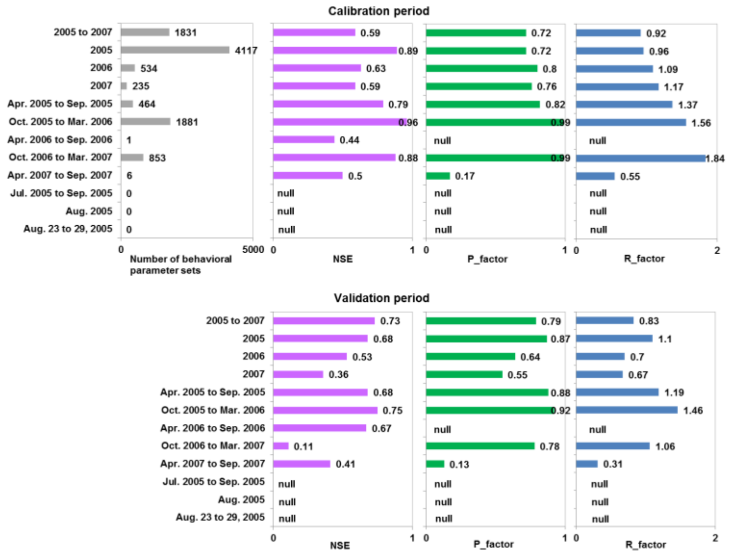
带格式的：字体：加粗

带格式的：左

带格式的：字体：(中文) Times New Roman，加粗

带格式的：字体：(中文) +中文正文 (宋体)，(中文) 中文(中国)

带格式的：制表位： 27.55 字符，左对齐 + 不在 22.57 字符 + 45.13 字符



带格式的: 居中

带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符

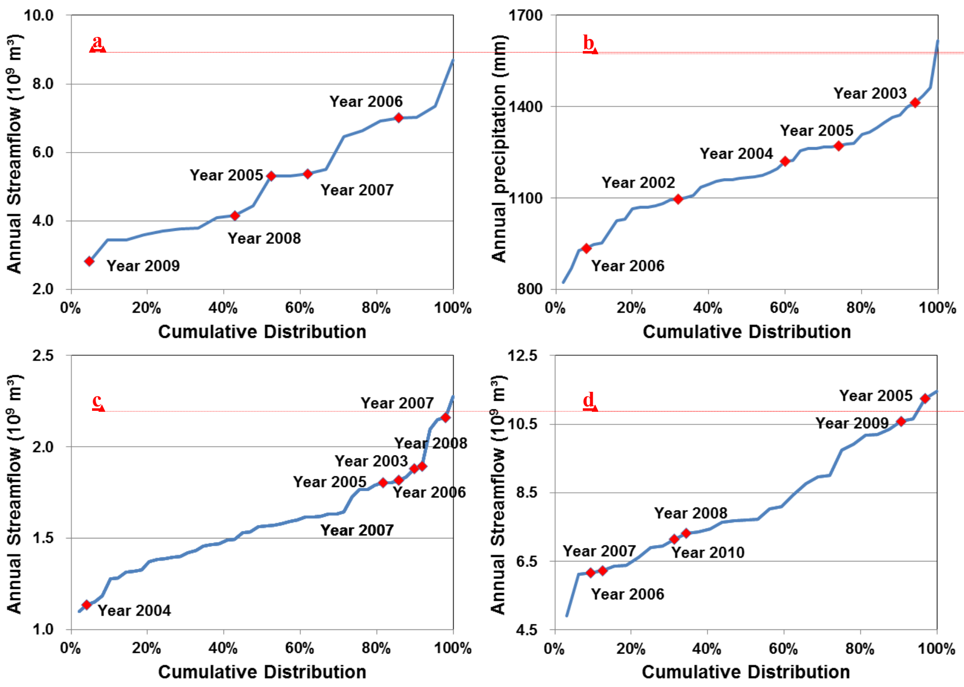


Figure 11: Model performance for the calibrations using short-period data in Yalongjiang Basin

带格式的：字体：加粗

带格式的：字体：加粗，英语(美国)

带格式的：字体：加粗，英语(美国)

带格式的：字体：加粗，英语(美国)

带格式的：字体：加粗，英语(美国)

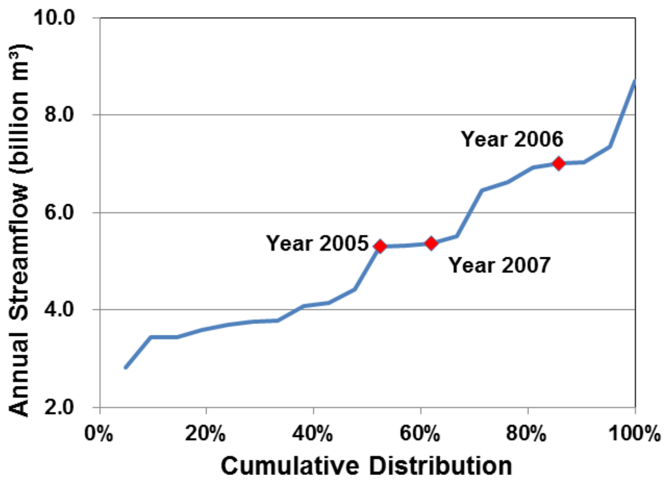
带格式的：字体：加粗

带格式的：字体：加粗

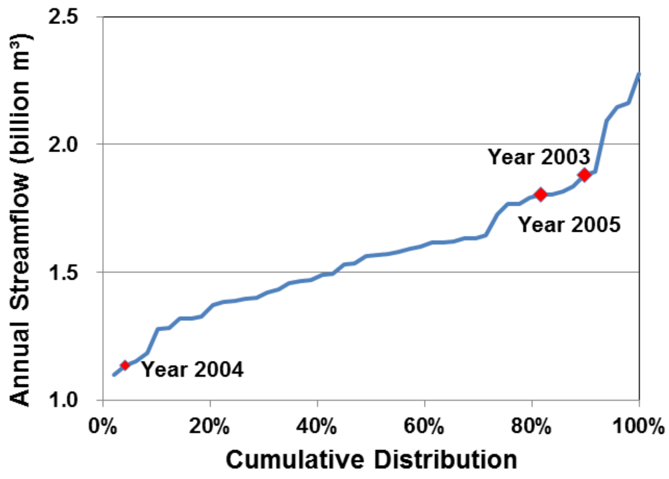
带格式的：正文

带格式的：左

带格式的：制表位： 27.55 字符，左对齐 + 不在 22.57 字符 + 45.13 字符



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

Basin	Streamflow station	Area (km2)	Climate	Annual Rainfall (mm)	Annual average temperature(℃)	Ranges of Elevation(m)
Jinjiang	Shilong	5,629	Subtropical marine monsoon climate	1651	20	50 to 1366

带格式的：字体：非加粗

带格式的：字体：(中文) +中文正文 (宋体), (中文) 中文(中国), (其他) 英语(美国)

带格式的：居中

带格式的：字体：10 磅

带格式的：字体：非加粗

带格式的：字体：(中文) +中文正文 (宋体), (中文) 中文(中国)

带格式的：字体：加粗

带格式的：字体：10 磅

带格式的：字体：(默认) +西文正文 (Times New Roman)

带格式的：居中，行距：单倍行距

带格式表格

带格式的：字体：(中文) +中文正文 (宋体), (中文) 中文(中国)

带格式的：字体：(中文) +中文正文 (宋体), (中文) 中文(中国)

带格式的：字体：(默认) +西文正文 (Times New Roman), (中文) +中文正文 (宋体)

带格式的：字体：(默认) +西文正文 (Times New Roman)

带格式的：字体：(中文) +中文正文 (宋体), (中文) 中文(中国)

带格式的：字体：(中文) +中文正文 (宋体), (中文) 中文(中国)

带格式的：字体：(中文) +中文正文 (宋体), (中文) 中文(中国)

带格式的：行距：单倍行距

带格式的：字体：(中文) +中文正文 (宋体), (中文) 中文(中国)

带格式的：字体：(中文) +中文正文 (宋体), (中文) 中文(中国)

带格式的：左，行距：单倍行距

带格式的：字体：(中文) +中文正文 (宋体), (中文) 中文(中国)

带格式的：居中，行距：单倍行距

带格式的：制表位：27.55 字符，左对齐 + 不在 22.57 字符 + 45.13 字符

Donghe	Wenquan	1,089	Subtropical monsoon climate	1247	18	192 to 2569
Heihe	Yingluoxia	8,843	Continental monsoon climate	423	6	1711 to 4749
Yalongjiang	Ganzi	3,2535	Continental plateau climate	570	8	3400 to 6021

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

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

Table ~~1~~3 Short periods for which corresponding data were used for the calibrations ~~for~~ at the stage one of the evaluation

Length of the period	Jinjiang Basin	Donghe Basin	Heihe Basin	Yalongjiang Basin
One year	2005	2002	2003	2005
	2006	2003	2004	2006
	2007	2004	2005	2007
Six months	April 2005 to September 2005	April 2002 to September 2002	April 2003 to September 2003	April 2005 to September 2005
	October 2005 to March 2006	October 2002 to March 2003	October 2003 to March 2004	October 2005 to March 2006
	April 2006 to September 2006	April 2003 to September 2003	April 2004 to September 2004	April 2006 to September 2006
	October 2006 to March 2007	October 2003 to March 2004	October 2004 to March 2005	October 2006 to March 2007
	April 2007 to September 2007	April 2004 to September 2004	April 2005 to September 2005	April 2007 to September 2007

Table ~~2~~ 4 SWAT model parameters being calibrated

带格式的：字体：（中文）+中文正文（宋体），（中文）中文(中国)

带格式的：行距：单倍行距

带格式的：字体：（中文）+中文正文（宋体），（中文）中文(中国)

带格式的：字体：（中文）+中文正文（宋体），（中文）中文(中国)

带格式的：居中，行距：单倍行距

带格式的：字体：（中文）+中文正文（宋体），（中文）中文(中国)

带格式的：字体：（默认）+西文标题（Times New Roman）

带格式的：行距：单倍行距

带格式的：字体：（中文）+中文正文（宋体），（中文）中文(中国)

带格式的 ...

带格式的 ...

带格式的：居中，行距：单倍行距

带格式的 ...

带格式的 ...

带格式的：行距：单倍行距

带格式的 ...

带格式的 ...

带格式的 ...

带格式的：居中，行距：单倍行距

带格式的 ...

带格式的：字体：10 磅

带格式表格

带格式的：居中

带格式的：居中

带格式的：居中

带格式的：居中

带格式表格

带格式的 ...

带格式的 ...

带格式的 ...

带格式的 ...

带格式的 ...

带格式的 ...

带格式的 ...

带格式的 ...

带格式的 ...

带格式的 ...

带格式的 ...

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

Table ~~3~~ 5 Model performance for the benchmark calibration in the two basins

	Number of behavioral parameter sets	NSE		~~U~~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
~~Heihe Basin~~	~~1445~~	~~0.78~~	~~0.78~~	~~0.44~~	~~0.41~~		
Donghe Basin~~Jinjiang Basin~~	1644~~2814~~	0.70~~0.85~~	0.75~~0.52~~	0.82~~-0.18~~	0.81~~0.28~~	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

~~**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~~

带格式的：左

带格式的：字体：(中文) +中文正文 (宋体)，(中文) 中文(中国)

带格式的：字体：(中文) +中文正文 (宋体)，(中文) 中文(中国)

带格式表格

带格式的：居中

带格式的：左

带格式的：左

带格式的：字体：(中文) +中文正文 (宋体)，(中文) 中文(中国)

带格式的：字体：(中文) +中文正文 (宋体)，(中文) 中文(中国)

带格式的：左

带格式的：字体：(中文) +中文正文 (宋体)，(中文) 中文(中国)

带格式的：字体：(中文) +中文正文 (宋体)，(中文) 中文(中国)

带格式的：字体：(中文) +中文正文 (宋体)，(中文) 中文(中国)

带格式的：字体：(中文) +中文正文 (宋体)，(中文) 中文(中国)

带格式的：左

带格式的：字体：(中文) +中文正文 (宋体)，(中文) 中文(中国)

带格式的：字体：(中文) +中文正文 (宋体)，(中文) 中文(中国)

带格式的：字体：(中文) +中文正文 (宋体)，(中文) 中文(中国)

带格式的：字体：(中文) +中文正文 (宋体)，(中文) 中文(中国)

带格式的：字体：(中文) +中文正文 (宋体)，(中文) 中文(中国)

带格式的：字体：(中文) +中文正文 (宋体)，(中文) 中文(中国)

带格式的：字体：(中文) +中文正文 (宋体)，(中文) 中文(中国)

带格式的：制表位： 27.55 字符，左对齐 + 不在 22.57 字符 + 45.13 字符

	~~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~~	~~-~~	~~-~~	~~-~~	~~-~~	~~-~~

带格式的: 制表位： 27.55 字符，左对齐 + 不在 22.57 字符 + 45.13 字符

	~~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~~	~~-~~	~~-~~

带格式的: 制表位: 27.55 字符, 左对齐 + 不在 22.57 字符 + 45.13 字符