Responses to Reviewer #1

The reviewer's comments have been considered carefully, and the manuscript has been extensively revised. The following tables summarize the reviewer's comments. For convenience, we have classified them into categories. Comments regarding similar issues and their locations are summarized together. Simplified notation has been used to assist the reviewer in conveniently locating a specific comment: R-Reviewers, and C-Comment. For example, R1-C1 denotes comment 1 made by Reviewer #1. This simplified form is used in all of the following tables.

Category I. The abstract and introduction

	Comment	Reviewer-Location
1	Refining the highlight	R1-C1

Category II. The methodology

	Comment	Reviewer-Location
1	Justification of using two hydrological models	R1-C5
2	Adding a flowchart of the study	R1-C6
3	Adding the evaluation metrics	R1-C7

Category III. The case study and results

	Comment	Reviewer-Location	
1	Adding analysis and description of data used in case study	R1-C8	case study
2	Figure modification	R1-C10	results
3	Modification for the results	R1-C7	

Category IV. The discussion and conclusions

	Comment	Reviewer-Location	
1	Determination of the length of sub-periods	R1-C4	discussion

2	Refining the highlight	R1-C1	conclusion

Category V. Others

	Comment	Reviewer-Location
1	Identifiability of parameters	R1-C2
2	Catchment characteristics unable to be predicted	R1-C3
3	Justification of the parameter continuity assumption	R1-C9

Responses to Reviewer #1:

1. This paper presents methods to estimate the time-varying parameter based on dynamic programming. The authors attempt to combine multiple methods including SSC and ENKF. However, the highlight of this paper is no very clear, which should be refined.

Reply:

Thank you for reviewing our manuscript and for the professional comments. The highlights of this paper are refined as follows:

1. The proposed method combines split-sample calibration (SSC) and ensemble Kalman filter (EnKF) for time-varying parameter estimation. Compared to SSC, the proposed method can find a more continuous parameter trajectory; compared to EnKF, the proposed method allows parameters to retain stable for a pre-determined period, instead of varying at every time-step.

2. The effectiveness of the proposed method is validated with two hydrological models and two real catchment case studies of different conditions.

3. For the case study of the Xun River basin, the proposed method detects the strongest seasonal signal.

The highlights are elaborated on in the abstract as follows:

Although the parameters of hydrological models are usually regarded as constant, temporal variations can occur in a changing environment. Thus, effectively estimating time-varying parameters becomes a significant challenge. **Two methods, including split-sample calibration (SSC) and Data assimilation, have been used to estimate time-varying parameters. However, SSC is unable to consider the parameter temporal continuity, while Data assimilation assumes parameters vary at every time-step.** This study proposed a new method that combines (1) the basic concept of SSC, whereby parameters are assumed to be stable for one sub-period, and (2) the parameter continuity assumption, i.e., the differences between parameters in consecutive time steps are small. **(Pages 2, Lines 3-7)**

The highlights are also elaborated in the conclusions as follows:

1. The proposed method with a suitable length not only produces better simulation performance, but also ensures more accurate parameter estimates than SSC and EnKF in the synthetic experiment using the TMWB model with two parameters. The impact of sub-period lengths on the performance of SSC-DP is significant when the known parameters vary sinusoidally.

2. **The proposed method can be used to deal with complex hydrological models involving a large number of parameters**, demonstrated by the synthetic experiment using the Xinanjiang model with 15 parameters. A sensitivity analysis was performed to reduce the probable computational cost and improve the efficiency of identifying the time-varying parameters.

3. **The proposed method has the potential to detect the relationship between the time-varying parameters and dynamic catchment characteristics.** For example, SSC-DP produced the best simulation performance in the case study of the Wuding River basin and detects that parameters representing soil water capacity and impervious areas changed significantly after 1972, reflecting the soil and water conservation projects carried out from 1958–2000. Additionally, SSC-DP detects the strongest seasonal signal in the case study of Xun River basin, indicating the distinct impacts of seasonal climate variability. **(Page 33 Line 656~673)**

2. The fundamental assumption that the individual parameters may not response to the catchment dynamics due to the linear or nonlinear correlations between parameters (Bardossy, 2007). The effects of identifiability of parameters on this research are suggested to be investigated.

Reply:

We agree with the reviewer that the hydrological model parameters should be treated as parameter vectors instead of independent individual values (Bardossy 2007). The identifiability of parameters is considered in this study:

(1) Parameters are not treated as individuals, but multiple parameters are identified simultaneously. For the two-parameter monthly water balance (TMWB) model, parameters C and SC are estimated simultaneously. While for the Xinanjiang model, the sensitive parameters are calibrated at the same time.

(2) By generating a large number of parameter sets as candidates in each sub-period, the proposed method takes into account the parameter equifinality, while the traditional SSC method only takes the optimal parameter set.

3. The non-stationary change in catchment characteristics may not be predicted. Lots of uncertainty factors would prevent the estimation of future scenarios in catchments.

Reply:

This study focuses on methods to identify time-varying parameters, and the future research is considered to relate time-varying parameters and available information, such as the number of dams and population. Then the time-varying parameters' function can be derived to predict future streamflow under the changing environment.

4. How to generally estimate the stable period, such as decades, years or months, considering catchment characteristics? It is vital for the method in this study. The impact of sub-period lengths on the performance of SSC-DP is significant.

Reply:

Determination of the stable period considers 3 factors:

1. Temporal scale of climate change or human activities. The Wudinghe River basin is taken as a case study. Since 1960s, the soil and water conservation measures were carried out in this basin to reduce the highly erodible loess, such as tree plantation, reservoir construction and land terracing. The human activities lead to a durative and long-term change in the catchment characteristic. Hence, the yearly sub-period is considered.

2. Seasonality. The Xun River basin is taken as a case study. Contrary to the Wudinghe River basin, the relationship between precipitation and runoff of the Xun River basin is rarely affected by human activities during 1991-2001. However, its significant seasonal dynamics can be observed and has been studied in literature (Lan et al. 2020, Lan et al. 2018). In order to diagnose the seasonality, the stable period of 3-month is considered.

3. The simulation accuracy. The length should not be too long to capture the variations in physical processes, while it should be long enough to reduce the uncertainty of calibration. Based on the results of the synthetic experiments, it is suggested that the length should be as long as possible without degrading the simulation performance significantly. For example, in the synthetic experiment with the TMWB model, if the difference between the NSE values of 6-SSC-DP and 3-SSC-DP is small, the preferred length is six months.

The determination of the sub-period length has been described in discussion as follows:

It is suggested that the determination of the sub-period length considers three factors:

(1) The temporal scale of climate change or human activities. For example, the Wudinghe River basin is taken as a case study. The soil and water conservation measures have led to a durative and long-term change in the catchment characteristics since the 1960s. Due to this, the yearly sub-period is preferred.

(2) The seasonality. Contrary to the Wudinghe River basin, the relationship between precipitation and runoff of the Xun River basin is rarely affected by human activities during 1991-2001. However, its significant seasonal dynamics can be observed and has been studied in literature (Lan et al. 2020, Lan et al. 2018). In order to diagnose the seasonality, the stable period of 3-month is adopted.

(3) The simulation accuracy. The length should be neither too long nor too short so as to increase the reliability of the calibration while guaranteeing that variations in real processes are captured. Thus, given that the time scale of the variations is unknown, the proposed SSC-DP can be used with different split-sample lengths. It is suggested that the length should be as long as possible without degrading the simulation performance significantly. For example, in the synthetic experiment with the TMWB model, if the difference between the NSE values of 6-SSC-DP and 3-SSC-DP is small, the preferred length is 6-month. **(Page 31~32 Line 623~641)**

5. The two lumped models were chosen in this study. The number of parameters is different. The sensitivity analysis was further performed to reduce the dimension of parameters in the Xinanjiang model. Hence, the purpose of choosing two different lumped models should be discussed.

Reply:

Two lumped models are chosen to evaluate the applicability of the proposed method to hydrological models with different number of parameters. Furthermore, the parameters of the TMWB model have been identified by EnKF in the work of Deng et al. (2016), but the parameters of the Xinanjiang model are scarcely recognized as time-variant. Hence, the use of the TMWB model is beneficial for comparison.

The purpose of choosing two different lumped models has been added as follows:

There are two important differences between the TMWB and Xinanjiang models: (1) the TMWB model has two parameters, while the Xinanjiang model has fifteen parameters; (2) TMWB is a monthly rainfall-runoff model, whereas the Xinanjiang model can run on hourly or daily step sizes. **(Page 9 Line 157~158)**

6. The titles cannot show the logic framework of the research. The flowchart is suggested to be used to illustrate the framework in this study. The introduction of the manuscript is suggested to present in the appendix.

Reply:

To avoid confusion, the title of Section 3, i.e., “Data and study area”, is replaced by “Synthetic experiment and real catchment case study”.

A flowchart describing the framework of the research is added as follows:

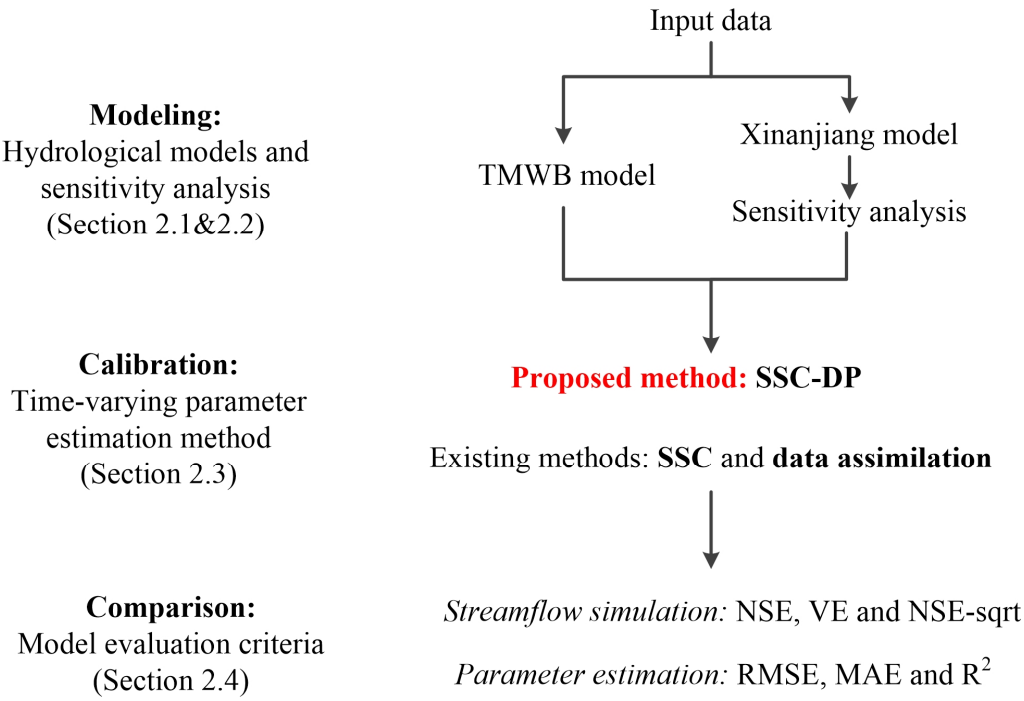


Figure 1 The flowchart of the methodologies

The introduction of the methodologies is presented as follows:

In this section, a SSC-DP method is proposed to identify the time-varying parameters of hydrological models. The two hydrological models considered in this study are the TMWB and Xinanjiang models. Their concepts and differences are presented in Sect. 2.1. A sensitivity analysis is employed to focus efforts on parameters important to calibration and avoid prohibitive computational cost, as outlined in Sect. 2.2. **Three time-varying parameter estimation methods (SSC, SSC-DP, and data assimilation) are presented in Sect. 2.3. The SSC and data assimilation are provided for comparisons with the SSC-DP.** Finally, to evaluate the performance of the time-varying parameter estimation methods, six evaluation criteria are selected and formulated in Sect. 2.4. **The flowchart of the methodologies is shown in Fig. 1. (Pages 7, Lines 123-132)**

7. The sensitive hydrograph phases of model performance criteria, i.e., RMSE, R^2 and NSE are peaks and discharge dynamics, flood peak, and discharge dynamics (Pfannerstill et al., 2014). Three metrics have strong correlations. The results as shown in Figure 5 needs furthermore discussion.

Reply:

Thanks for the comment. This comment involves three aspects:

(1) Three metrics are used to evaluate the streamflow simulations.

NSE coefficient, and two evaluation metrics have been added: relative error (RE) and the NSE on logarithm of streamflow (NSE_{ln}).

In the revised paper, these evaluation metrics are described as follows:

The streamflow simulations given by the proposed method are verified using the NSE, relative error (RE) and NSE on logarithm of streamflow (NSE_{ln}) (Hock, 1999). RE evaluates the error of the total volume of streamflow, while NSE and NSE_{ln} evaluate the agreement between the hydrograph of observations and simulations. NSE is more sensitive to high flows, but NSE_{ln} focuses more on low flows. Higher values of NSE, NSE_{ln} and lower absolute values of RE indicate better streamflow simulations. The NSE, RE and NSE_{ln} are expressed as followed: **(Pages 15~16, Lines 292-301)**

$$NSE = 1 - \frac{\sum_{t=1}^{m}(Q_t - \hat{Q}_t)^2}{\sum_{t=1}^{m}(Q_t - \overline{Q}_t)^2} \tag{15}$$

$$RE = \frac{\sum_{t=1}^{m}(Q_t - \hat{Q}_t)}{\sum_{t=1}^{m} Q_t} \tag{16}$$

$$NSE_{\ln} = 1 - \frac{\sum_{t=1}^{m}(\ln(Q_t) - \ln(\hat{Q}_t))^2}{\sum_{t=1}^{m}(\ln(Q_t) - \ln(\overline{Q}_t))^2} \tag{17}$$

Description of the evaluation results has been added in Revised Manuscript as follows:

- For results of the synthetic experiment with the TMWB model

Figure 6(a) presents the runoff simulation performance for various scenarios. In scenario 1, the NSE values of the three SSC-DP methods are all higher than that of EnKF. The results of NSE_{ln} show no significant differences among various methods. For scenarios 2, 4, and 6, where true parameters have linear trends, the 6-SSC-DP and 12-SSC-DP are superior to the EnKF and 3-SSC-DP in terms of NSE and NSE_{ln}. In scenario3, where the true parameters have periodic variations and change every month, the NSE and NSE_{ln} values of 6-SSC-DP and 12-SSC-DP decrease significantly, because the assumed sub-period length is longer than the time-scale of actual variations. Similarly, in scenario 5, 12-SSC-DP performs worst for NSE and NSE_{ln}, but 6-SSC-DP performs best. In scenario 7 and 8, both 6-SSC-DP and 12-SSC-DP perform better than EnKF. According to the evaluations of NSE and NSE_{ln}, the SSC-DP offers improved accuracy than the EnKF if the proper length is chosen. Another advantage of the SSC-DP is the small RE. For all scenarios, the SSC-DP methods significantly outperform for RE compared with EnKF. Among the SSC-DP methods, the RE of 3-SSC-DP is the smallest. **(Page 22 Line 419~433)**

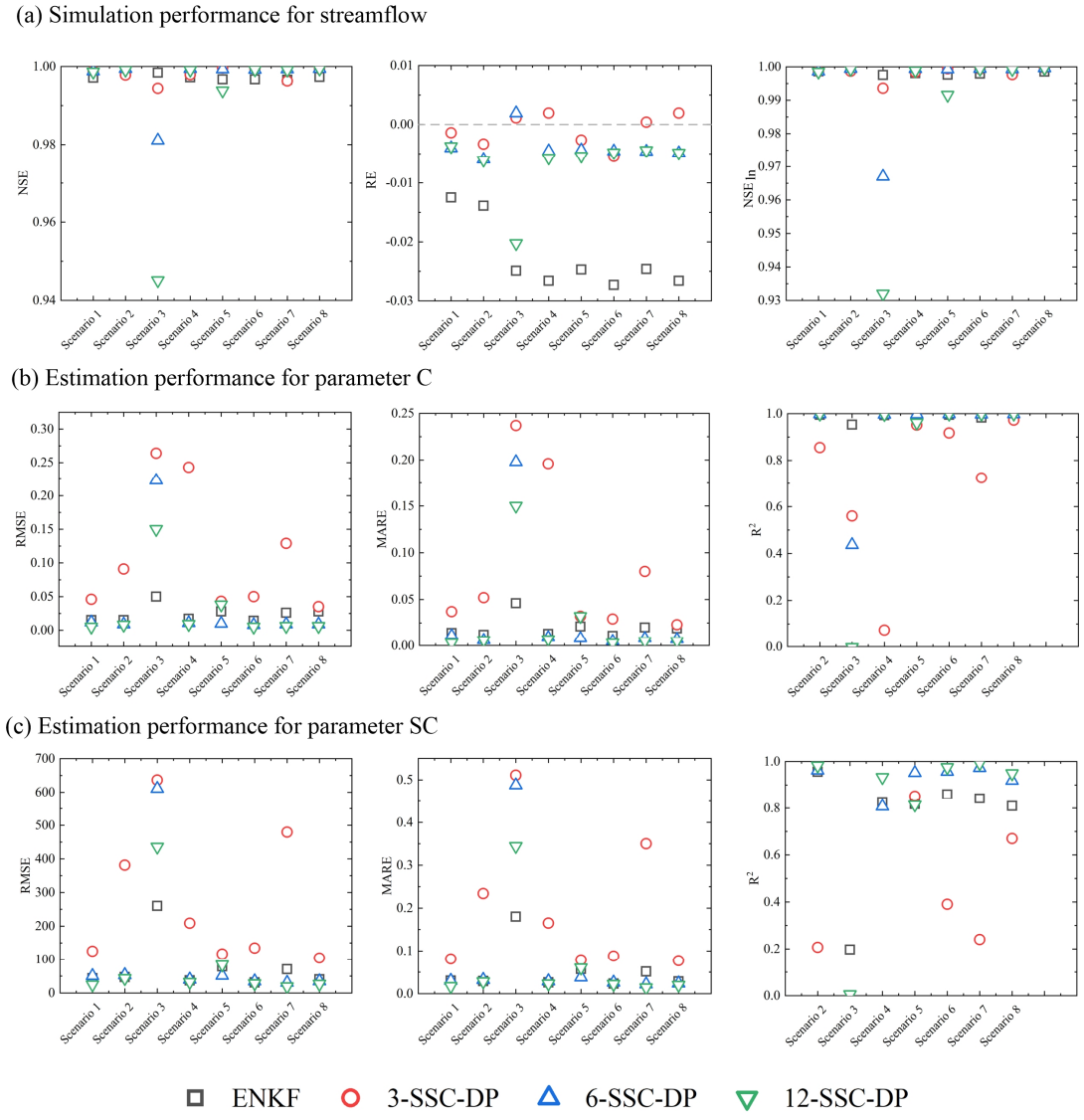


Figure 6 Comparison between the EnKF and SSC-DP methods for (a) streamflow simulation and identification of (b) parameter C and (c) parameter SC.

➢ For results of the synthetic experiment with the Xinanjiang model

The simulated streamflow and identification of time-varying parameters was compared across four methods: 1-SSC, SSC-EnKF, 1-SSC-DP, and 2-SSC-DP. The simulation performance is summarized in Figure 9(a). For all scenarios, the NSE of 2-SSC-DP is the lowest, but it performs better for low flows. The SSC-EnKF produces the highest RE in scenarios 2, 3 and 4, indicating the problem of simulating water balance. The SSC and 1-SSC-DP perform well for all scenarios in terms of NSE, RE and NSE_{ln}. Wherein, the SSC performs better than the 1-SSC-DP with regard to RE, while 1-SSC-DP is slightly superior to SSC in scenario 3 with higher NSE_{ln}. **(Page 24 Line 472~479)**

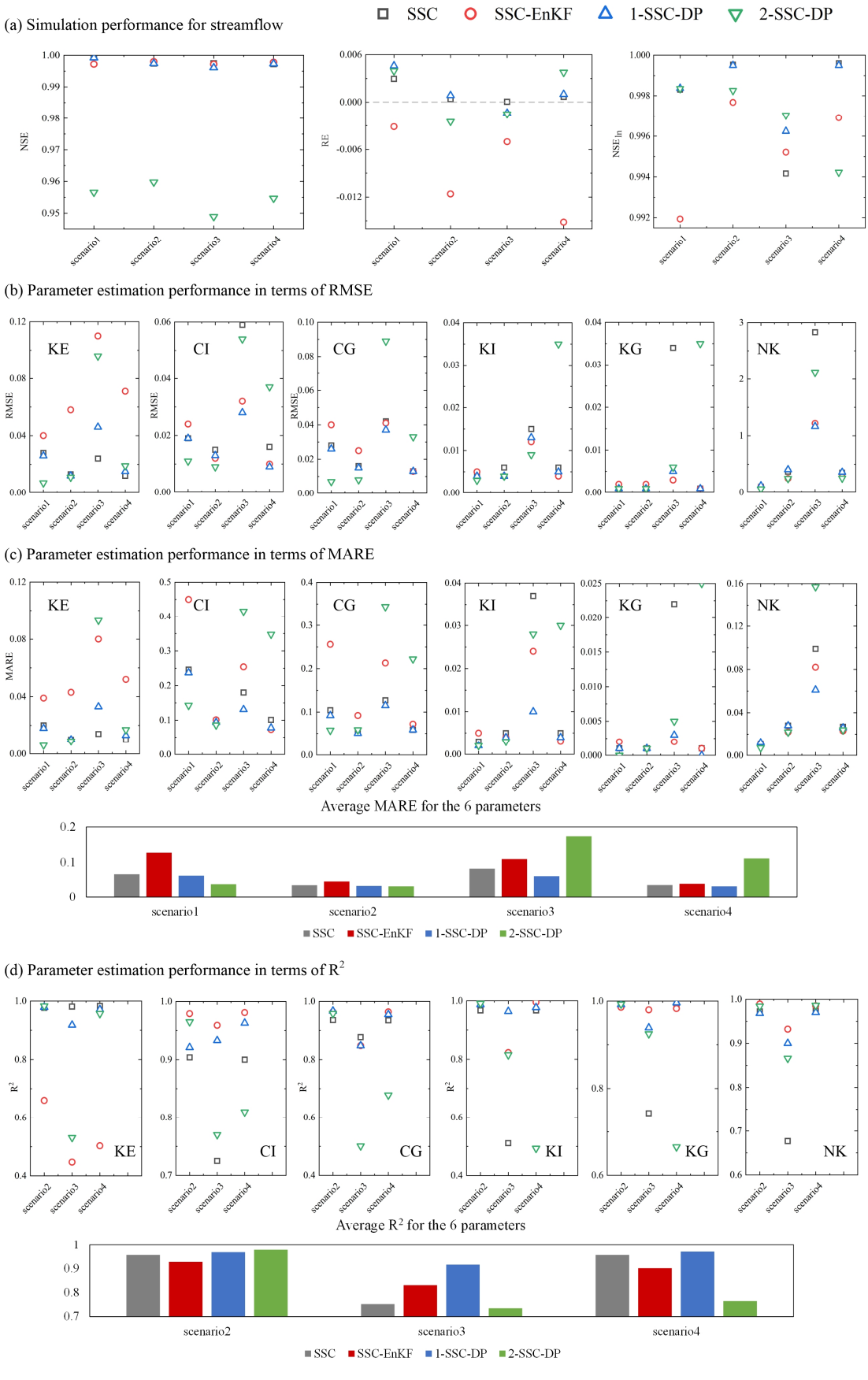


Figure 9 Comparison among the SSC, SSC-EnKF and SSC-DP methods for (a)

streamflow simulation and parameter identification in terms of (b) RMSE, (c) MARE and (d) R^2.

- For results of case study in Wuding River basin

The simulation performance is presented in Figure 12. The values of the NSEs are relatively low, because the streamflow in dry regions is difficult to simulate. It can be seen that the 12-SSC-DP gives the best simulation results among different methods with the highest NSE, NSE_{ln} and small RE. Although the 12-SSC produces relatively high NSE, it performs worst simulations for low flows. The SSC-EnKF has relatively high NSE_{ln}, but the RE of it is the largest. Overall, the 12-SSC-DP significantly improves the simulation performance of the Xinanjiang model in the Wuding River basin. **(Page 26 Line 506~515)**

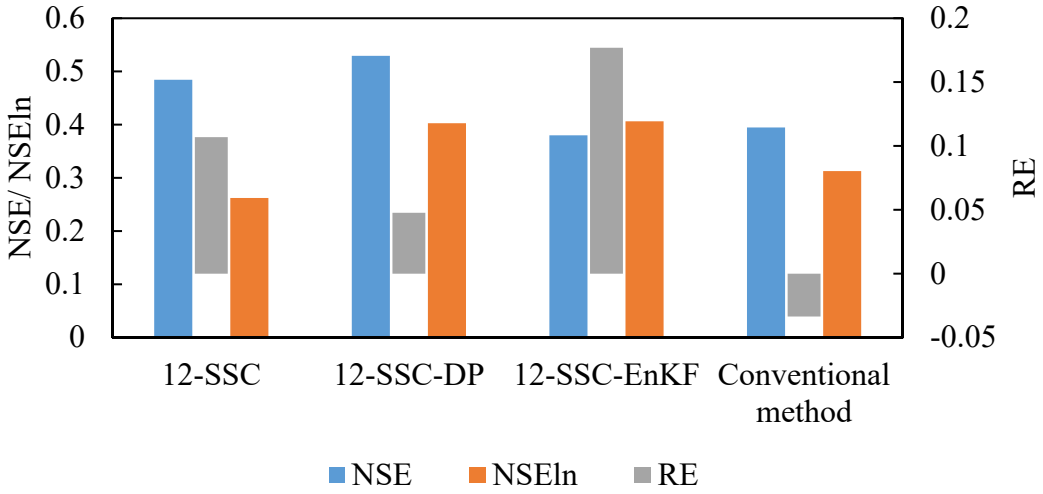


Figure 12 Simulation performance for streamflow in the Wuding River basin.

- For results of case study in Xun River basin

The simulation performance is presented in Figure 15. All methods performed well, with NSE values of 92.5 %, 93.0 %, 95.0 %, and 94.8 % for the conventional method, 3-SSC-EnKF, 3-SSC, and 3-SSC-DP, respectively. 3-SSC and 3-SSC-DP also perform well for NSE_{ln} compared with 3-SSC-EnKF and the conventional method. However, as regards to RE, the values are 0.0007 and 0.0324 for 3-SSC-DP and 3-SSC-DP, respectively. It indicated that the 3-SSC-DP can better simulate water balance than the 3-SSC in the Xun River basin. **(Page 28~29, Lines 560~567)**

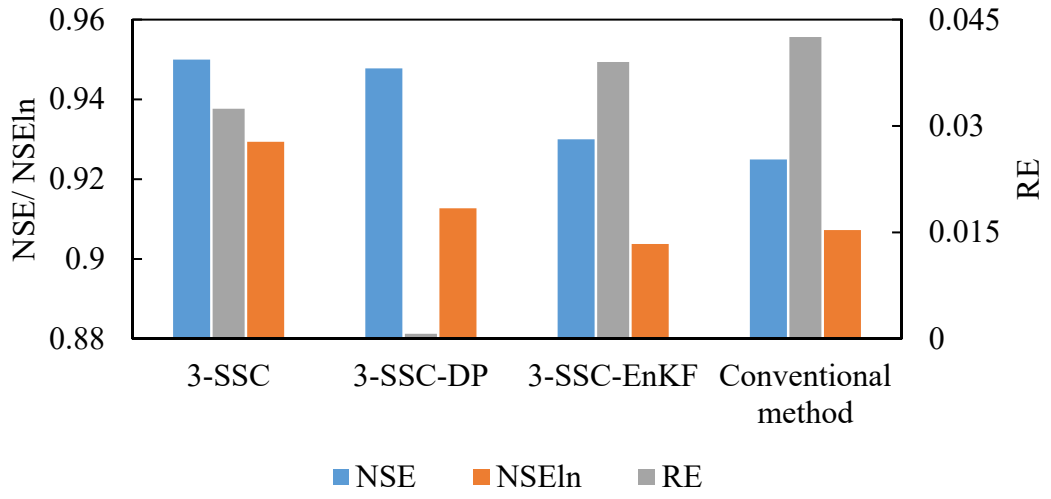


Figure 15 Simulation performance for streamflow in the Xun River basin.

(2) Three metrics are used to evaluate the parameter estimations.

The estimated parameters are evaluated by the RMSE (Alvisi et al., 2006), MARE (Khalil et al., 2001) and R^2 (Kim et al., 2007). RMSE and MARE quantify the accuracy of the estimated parameters, but RMSE is more sensitive to high values than MARE. R^2 records the overall agreement between the true and estimated parameters. Smaller values of RMSE, MARE and higher values of R^2 indicate stronger parameter identification ability. For the *p*-th parameter, the formulations are as follows:

$$RMSE_p = \sqrt{\frac{1}{m}\sum_{t=1}^{m}(\theta_{t,p} - \hat{\theta}_{t,p})^2} \tag{18}$$

$$MARE_p = \frac{1}{m}\sum_{t=1}^{m}\frac{\left|\theta_{t,p} - \hat{\theta}_{t,p}\right|}{\theta_{t,p}} \tag{19}$$

$$R^2_p = \frac{\sum_{t=1}^{m}\left(\hat{\theta}_{t,p} - \bar{\hat{\theta}}_p\right)\left(\theta_{t,p} - \bar{\theta}_p\right)}{\sqrt{\sum_{t=1}^{m}\left(\hat{\theta}_{t,p} - \bar{\hat{\theta}}_p\right)^2\left(\theta_{t,p} - \bar{\theta}_p\right)^2}} \tag{20}$$

where θ_t and $\hat{\theta}_t$ are the true parameter and its estimated value at the *t*-th time step, respectively; $\bar{\theta}_p$ and $\bar{\hat{\theta}}_p$ are the mean value of the true parameters and its estimated values, respectively; and m is the length of the data during the whole period. **(Pages 16-17, Lines 302-313)**

Description of the evaluation results has been added in Revised Manuscript as follows:

- For results of the synthetic experiment with the TMWB model

Figures 6 (b) and (c) focuses on the ability of the four methods to identify time-varying parameters. It can be seen that the RMSE and MARE values of the 3-SSC-DP are larger than those of other methods in most cases. That is because the sub-period length that serves as a calibration period for MCMC is too short (i.e., three months) that the estimated parameters are associated with higher uncertainties.

Regarding the synthetic true parameters are constant values (scenario 1), 12-SSC-DP gives the best performance with the lowest RMSE, MARE and highest R^2. The observations and estimated parameters are presented in Figure 7 (b). It shows that the estimated parameters obtained by EnKF vary at every time step, resulting in larger deviations from the observations than 6-SSC-DP and 12-SSC-DP.

When the synthetic true parameters vary linearly (scenarios 2, 4, and 6), 12-SSC-DP produces best estimations in comparison with EnKF, 3-SSC-DP, and 6-SSC-DP. The performances of 6-SSC-DP and EnKF are similar.

When the synthetic true parameters vary sinusoidally from month to month, EnKF gives the best estimations in scenario 3. The poor performances of 6-SSC-DP and 12-SSC-DP can be explained by the sub-period length being much longer than the actual one. When the parameters vary periodically at six-month intervals (scenario 5), 6-SSC-DP yields the best performance with the lowest RMSE, MARE and highest R^2. The differences of estimation performances among 3-SSC-DP, 12-SSC-DP and EnKF are small. The estimated parameters for scenario 5 have been plotted in Fig. 7(a). Although 3-SSC-DP and 12-SSC-DP have different lengths of sub-periods, they can also detect the correct seasonal signal of the parameters. For the annual variation in parameters (scenario 7), 12-SSC-DP and 6-SSC-DP produce better results than EnKF. Similar results can be seen in scenario 8 where C has a combined variation from year to year. In summary, the results indicate that the SSC-DP with a suitable length can estimate more accurate parameters than EnKF. **(Page 22~24 Line 434~459)**

- For results of the synthetic experiment with the Xinanjiang model

Figures 9(b) and (c) compare the time-varying parameter estimation performance among the four methods. In scenarios 1 and 2, 2-SSC-DP produces the lowest RMSE, MARE and R2, followed by the 1-SSC-DP. The 1-SSC-DP is slightly superior to the 1-SSC and significantly outperforms the SSC-EnKF for the two scenarios.

When the synthetic true parameters vary sinusoidally from month to month (scenario 3), the estimated parameters are plotted in Fig. 10. It can be seen that 1-SSC-DP successfully detects seasonal signal in every parameter. The SSC-EnKF performs well for R2, but it has high MARE. Although the average MARE of the SSC and 2-SSC-DP are lower than that of SSC-EnKF, the R^2 of them are relatively low. Therein, from Fig. 10, the estimated parameters by the 1-SSC fluctuate generally periodically, but the variations are dramatic, resulting in lowest R^2 for CI, KI, KG and NK. The

estimated parameters of the 2-SSC-DP fluctuate more slowly, but the sub-period length is too long. In scenario 4, 1-SSC performs better than the SSC-EnKF and 2-SSC-DP, but is still slightly inferior to the 1-SSC-DP. Overall, the 1-SSC-DP achieves higher-quality and more robust parameter estimations performances than the other methods. **(Page 25 Line 480~494)**

(3) The figure 5 is replaced by Figure 6 in the Revised Manuscript. The results as shown in Figure 6 have been presented in the reply (2) of R1-C7.

8. The streamflow, climate and underlying surface conditions in the two study areas were not analyzed in this study. However, it is critical to the estimation of time-varying parameters.

Reply:

Figure 5 has been modified as follows:

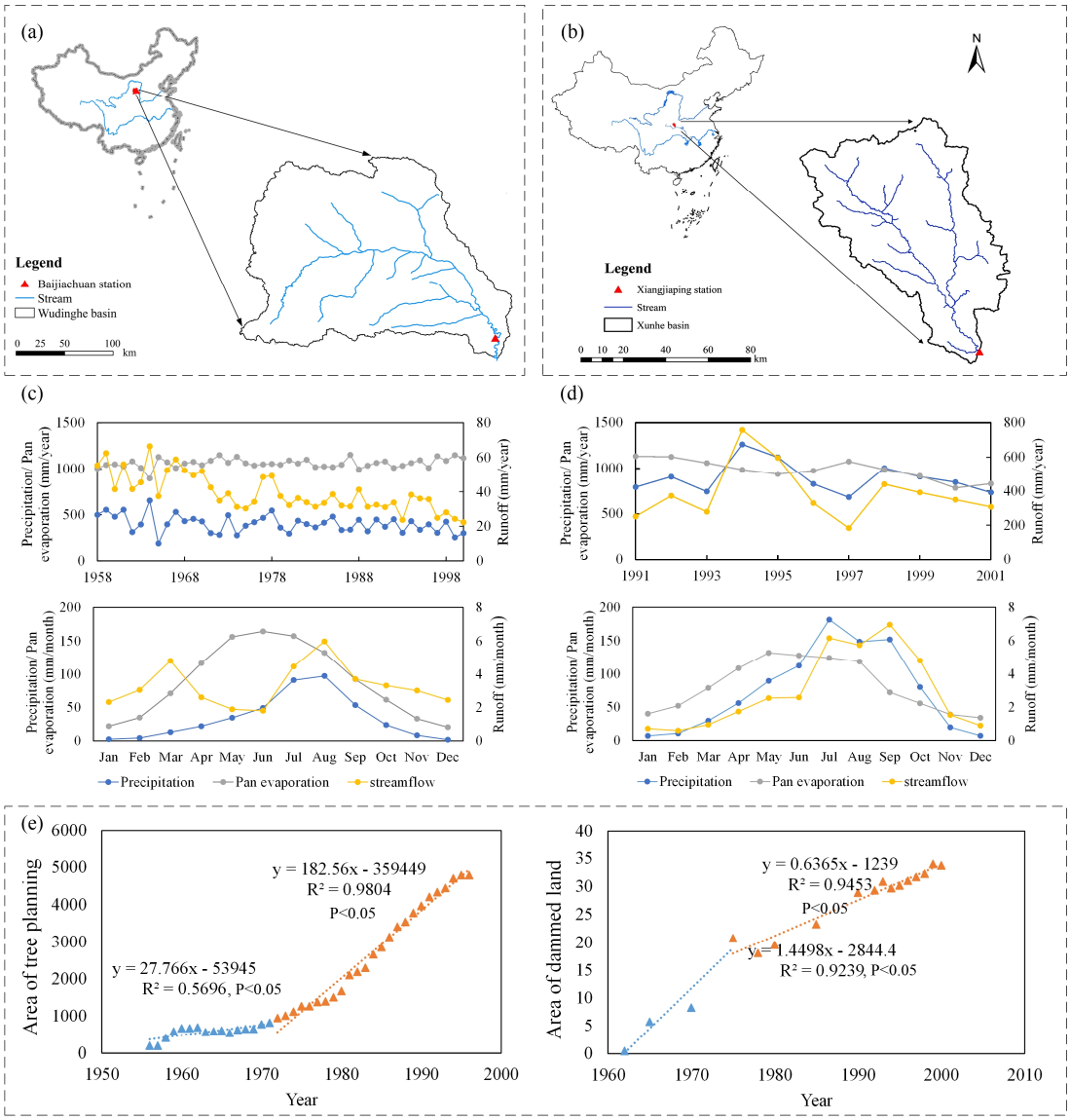


Figure 5 Location of (a) Wuding River basin and (b) Xun River basin. The plots (c) and (d) show the average yearly and monthly variations of precipitation, pan evaporation and streamflow in the Wuding River basin and Xun River basin, respectively. The plot (e) shows the temporal variations in the soil and water conservation measures.

The details of the Wuding River basin have been added as follows:

As illustrated in Fig. 5(a), the station furthest downstream, Baijiachuan, drains an area of 29,662 km^2 (98 % of the total basin) and records the daily runoff data. The data of the daily precipitation and streamflow in the Wuding River basin were obtained from the local Hydrology and Water Resources Bureau of China, the quality of which has been checked by the official authorities, and there are no gaps among these data for all the hydrological stations. It can be seen from Fig. 5(c) that the annual streamflow in the Wudinghe River basin has a distinct decreasing trend, while seasonal variations are not significant, but the annual precipitation and pan evaporation generally have no trend, suggesting the impacts of human activities on rainfall–runoff relationships. **(Page 20 Line 374~381)**

The details of the Xun River basin have been added as follows:

It can be observed from Fig. 5(d) that no trend is found in annual precipitation, pan evaporation and streamflow, suggesting that the relationship between precipitation and runoff of the Xun River basin is rarely affected by human activities during 1991-2001. However, there exhibit strong seasonal patterns in these three climatic and hydrological variables, suggesting that seasonal variations in hydrological parameters should be considered. **(Page 21 Line 404~409)**

9. In lines 175-176, the assumption that the continuity condition aims to minimize the difference between the estimated parameters for sub-periods i and i+1 unreasonable. The differences between two consecutive sub-periods represent the time-varying changes of the catchment. The continuity conditions for enhancing the model performance should focus on the model structure, such as state variables.

Reply:

Thanks for the comment. The main hypothesis of parameter continuity is justified as follows:

1. The hypothesis of parameter continuity can be found in the model prediction process of the ensemble Kalman filter (EnKF). Therein, the values of the parameters at the time step t+1 are forecasted by perturbing those of parameters from the time step t. The equation is as follows:

$$\vartheta_{t+1}^{k-}=\vartheta_{t}^{k+}+\delta_{t}^{k},\ \delta_{t}^{k}\sim N\left(0,R_{t}\right) \tag{1}$$

where ϑ_{t+1}^{k-} is the forecasted parameter vector at the time step t+1,while ϑ_{t}^{k+} is the well-calibrated parameter vector. δ_{t}^{k} is the white noise following a Gaussian distribution with zero mean and specified covariance of R_t which is very small. That is, the fluctuations between parameters of adjacent sub-periods can be little.

2. Some conceptual hydrological parameters reflect the catchment characteristics, such as soil water storage capacity in the Xinanjiang model. While climate change and human activities exert influence on catchment characteristics, the soil water storage capacity can hardly change dramatically in a very quick time, such as an hour.

Hence, it is reasonable to consider parameter continuity in estimating time-varying parameters.

This point has been added in the Revised Manuscript as follows:

Some conceptual hydrological parameters reflect the catchment characteristics. While climate change and human activities exert influence on these catchment

characteristics, they can hardly change dramatically in a very quick time, such as the soil water storage capacity. **(Page 5 Line 79~82)**

10. Minor comment. The resolution of Figure 5 is low and information is not presented.
Reply:

The Figure 5 is replaced by Figure 6 in the Revised Manuscript to be easier to read as follows:

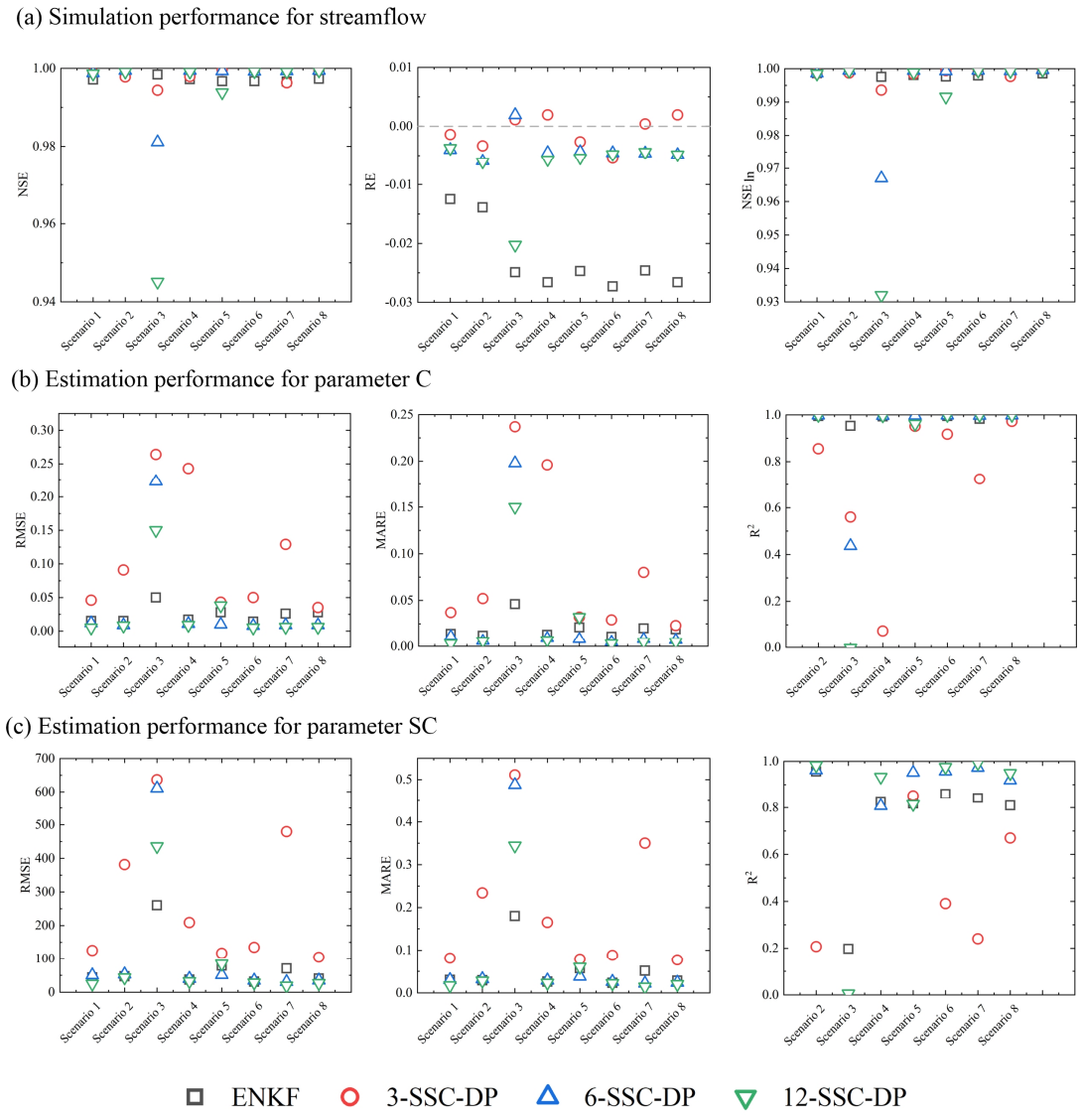


Figure 6 Comparison between the EnKF and SSC-DP methods for (a) streamflow simulation and identification of (b) parameter C and (c) parameter SC.

Bardossy, A. (2007) Calibration of hydrological model parameters for ungauged catchments. Hydrology and Earth System Sciences 11(2), 703-710.

Lan, T., Lin, K., Xu, C.-Y., Tan, X. and Chen, X. (2020) Dynamics of hydrological-model parameters: mechanisms, problems and solutions. Hydrology and Earth System Sciences 24(3), 1347-1366.

Lan, T., Lin, K.R., Liu, Z.Y., He, Y.H., Xu, C.Y., Zhang, H.B. and Chen, X.H. (2018) A Clustering Preprocessing Framework for the Subannual Calibration of a Hydrological Model Considering Climate-Land Surface Variations. Water resources research 54(0).

Deng, C., Liu, P., Guo, S., Li, Z. and Wang, D. (2016) Identification of hydrological model parameter variation using ensemble Kalman filter. Hydrology and Earth System Sciences 20(12), 4949-4961.

Responses to Reviewer #2

The reviewer's comments have been considered carefully, and the manuscript has been extensively revised. The following tables summarize the reviewer's comments. For convenience, we have classified them into categories. Comments regarding similar issues and their locations are summarized together. Simplified notation has been used to assist the reviewer in conveniently locating a specific comment: R-Reviewers, and C-Comment. For example, R2-C1 denotes comment 1 made by Reviewer #2. This simplified form is used in all of the following tables.

Category I. The methodology

	Comment	Reviewer-Location
1	Justification of using conceptual hydrological models	R2-C2
2	Adding units for Xinanjiang model parameters	R2-C6
3	Justification of using the sensitivity analysis	R2-C7
4	Adding the evaluation metrics	R2-C9

Category II. The case study and results

	Comment	Reviewer-Location	
1	Quantified analysis of the soil and water conservation measures on the Wuding River basin	R2-C10	
2	Quantified analysis of the seasonal variations on the Xun River basin	R2-C11	
3	Quality of the data used	R2-C10; R2-C11; R2-C14; R2-C18	
4	Figure modification	R2-C1;	results
5	Replacing tables with figures	R2-C1	
6	Modification for the results	R2-C12; R2-C20	
7	Statistical significance of this analysis	R2-C14	
8	Explanation for the relative bad performance on the Wuding River basin	R2-C14	
9	Modification of the attribution analysis	R2-C17	
10	Adding a hydrograph plot for the Xun River basin	R2-C19	

Category III. Others

	Comment	Reviewer-Location
1	Justification of the parameter continuity assumption	R2-C8
2	Explanation and modification of words and phrasing	R2-C3; R2-C4; R2-C5; R2-C13; R2-C15; R2-C16; R2-C21

Responses to Reviewer #2:

1. OVERALL RECOMMENDATION

The manuscript addresses the important topic of calibration of rainfall-runoff model parameters, and presents results obtained on two different catchments with two models. Even if the introduction includes relevant references and the methods are well presented, the paper lacks important discussions on the rainfall-runoff model performances, observed time series quality, attribution of observed/simulated changes, consideration of only two catchments, and several obtained results are over-interpreted. Finally, several figures and tables must be significantly improved. Therefore, I think the manuscript requires major revision before publication.

Reply:

Thank you for reviewing our manuscript and for the professional comments, which are carefully followed in making revisions.

2. GENERAL COMMENTS

(1) The tables 6 to 8 might be presented as figures to be more easily interpreted. Figure 5, 7 and 10 are very difficult to read, and must be significantly improved.

Reply:

(1) Table 6 and Figure 5 have been modified and replaced by Figure 6 in Revised Manuscript as follows:

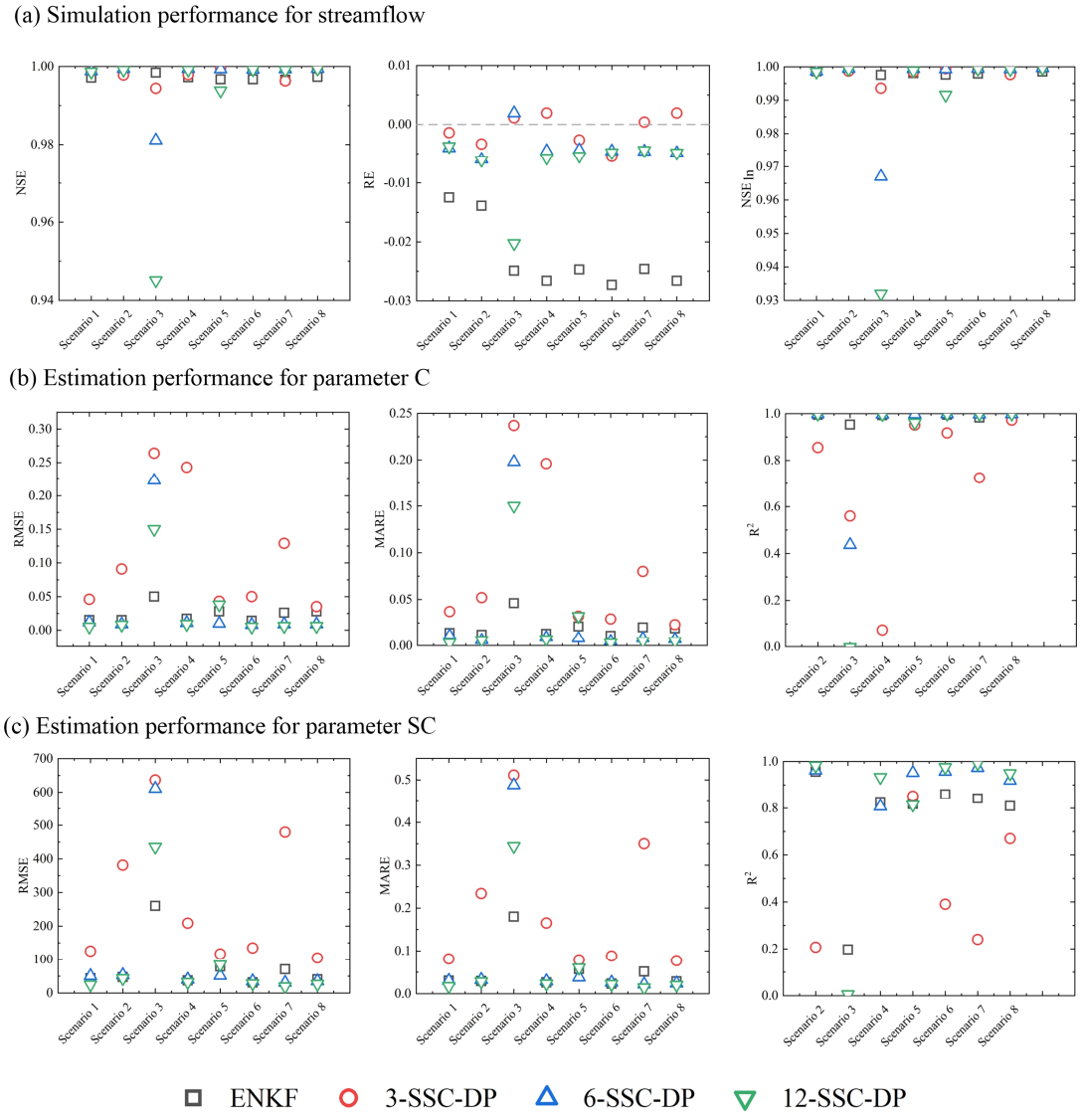


Figure 6 Comparison between the EnKF and SSC-DP methods for (a) streamflow simulation and identification of (b) parameter C and (c) parameter SC.

(2) Table 7 has been presented as Figure 8 in Revised Manuscript as follows:

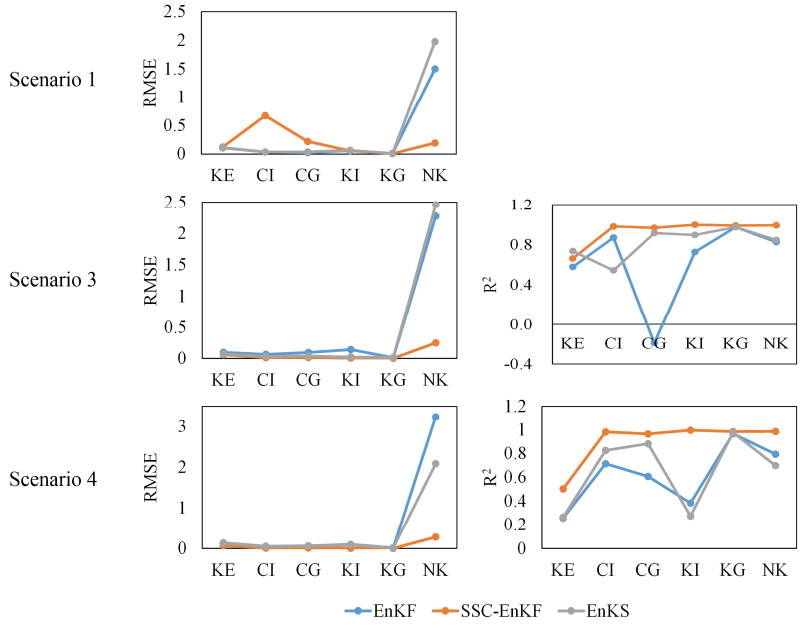


Figure 8 Comparison among EnKF, SSC-EnKF, and EnKS in the synthetic experiment with the Xinanjiang model

(3) Table 8 and Figure 7 have been modified and replaced by Figure 9 in Revised Manuscript as follows:

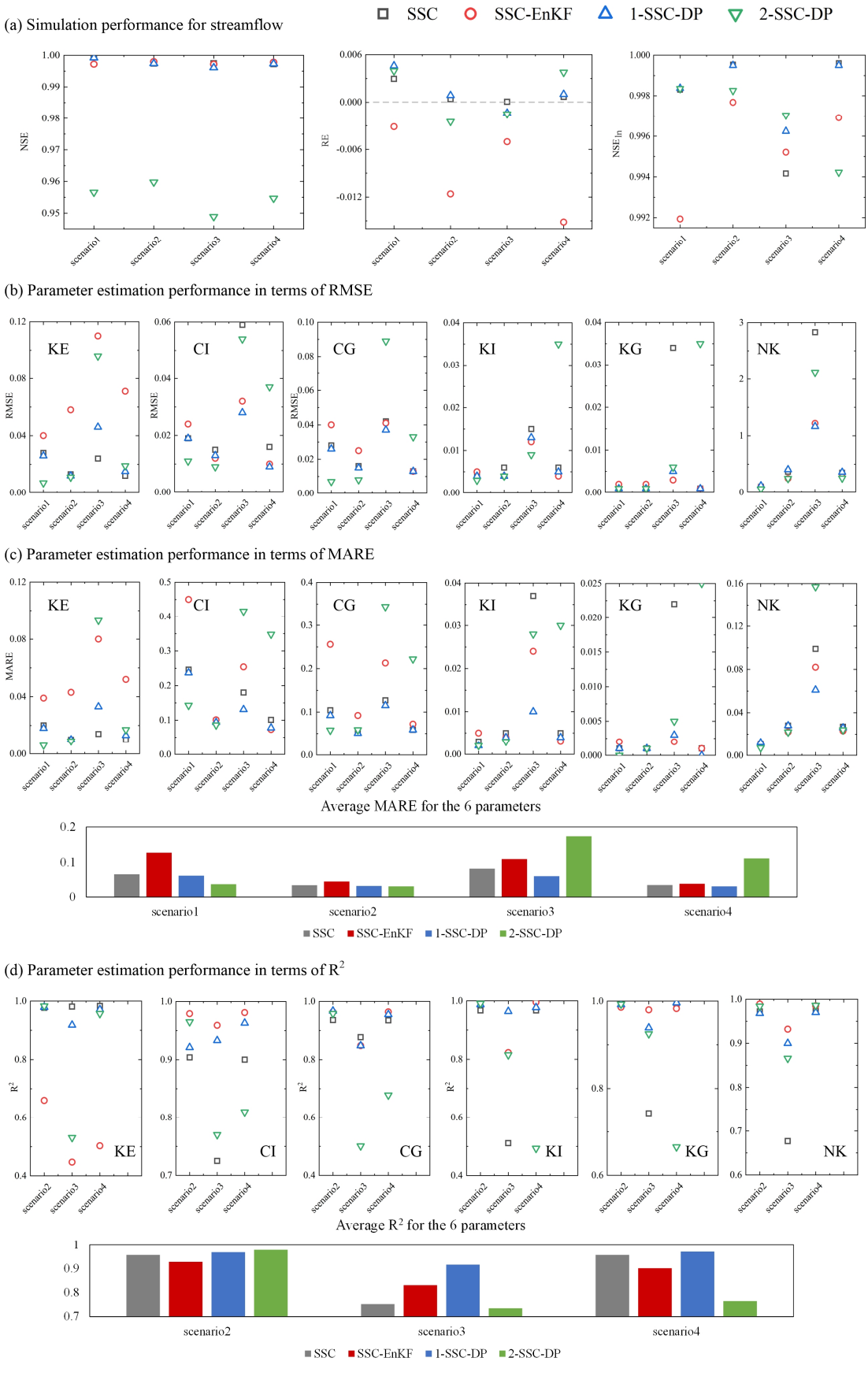


Figure 9 Comparison among the SSC, SSC-EnKF and SSC-DP methods for (a)

streamflow simulation and parameter identification in terms of (b) RMSE, (c) MARE and (d) R^2.

(4) Figure 10 is modified and replaced by Figure 13 in Revised Manuscript as follows:

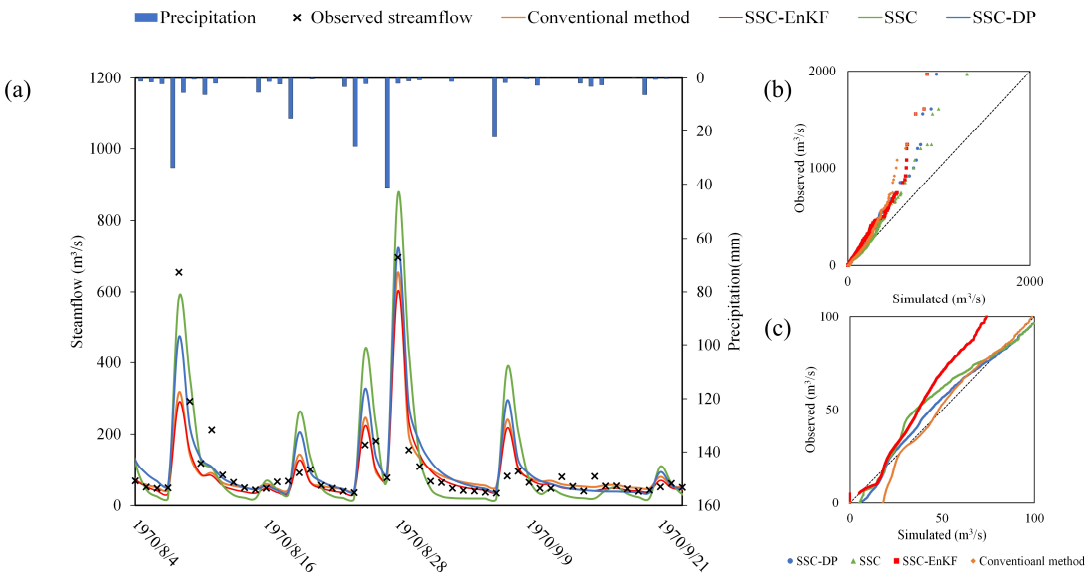


Figure 13 The simulated and observed streamflow using the conventional method, SSC-EnKF, SSC, and SSC-DP for the Wuding River basin. (a) Streamflow simulation hydrograph; (b) The quantile-quantile plot for all streamflow; (c) The quantile-quantile plot for streamflow lower than 100 m^3/s.

(2) Line 28 to 29: several studies highlighted the difficulty of conceptual rainfall-runoff models in the context of climate change impact studies.

Reply:

Thanks. We agree that conceptual rainfall-runoff models can be difficult to simulate the variations in discharge in response to climate changes in some cases (Merz et al. 2011, Fowler et al. 2020). That is, the simulation accuracy reduces when the conceptual model is applied in situations where the climatic conditions, e.g., dry periods, are not consistent with that of the calibration period, e.g., wet periods. Some literatures have made improvements to enhance parameter transferability between various climatic conditions. One approach is to allow the parameters of the conceptual model to change (Stephens et al. 2019, Deng et al. 2019), which can efficiently improve the accuracy of the conceptual model and simulate the response of runoff in a changing environment.

(3) Line 35 to 36: the terms “constants” and “stable” must be defined: constant/stable in space and/or in time?

Reply:

The terms are defined as “constant in time scale” and “temporally stable”. The statement at line 35 to 36 will be modified in the Revised Manuscript: Parameters are

usually regarded as constants in time scale, because of the general idea that catchment conditions are temporally stable.

(4) Line 43 to 44: in this context, it may be needed to define what is called "climate conditions".

Reply:

Here, the "climate conditions" means "wet/dry periods". To avoid confusion, the sentence at line 42 to 44 of the Revised Manuscript is modified by replacing "climate conditions" with "wet/dry periods": Fowler et al. (2016) pointed out that the parameter set obtained by mathematical optimization based on wet periods may not be robust when applied in dry periods.

(5) Line 122: the terms "behavioural" must be clearly defined or not used in this context.

Reply:

The "behavioural" means "important to calibration metrics and predictions". To avoid confusion, the "behavioural" is replaced by "sensitive" in the line 122 of the Revised Manuscript.

(6) Line 137, line 150 and Table 1 and 2: please presents parameter units.

Reply:

Thanks for reminder. The parameter units have been added on Line 137, line 150 and in Tables 1 and 2 as follows:

The parameter *SC* represents the field capacity **(mm)**. **(Pages 8, Lines 143~144)**

Table 1 Parameters of the TMWB model

Parameter	Physical meaning	Range and units
C	Evapotranspiration parameter	0.2-2.0 (–)
SC	Catchment water storage capacity	100-2000 (mm)

The meaning, range and units of all the parameters in the Xinanjiang model are listed in Table 2. **(Page 9 Line 154~155)**

Table 2 Parameters of the Xinanjiang model

Category	Parameter	Physical meaning	Range and units
Evapotranspiration	WM	Tension water capacity	80-400 (mm)
	X	WUM=X×WM, WUM is the tension water capacity of lower layer	0.01-0.8 (–)
	Y	WLM=Y×WM, WLM is the tension water capacity of deeper layer	0.01-0.8 (–)
	K	Ratio of potential evapotranspiration to pan evaporation	0.4-1.5 (–)

	C	The coefficient of deep evapotranspiration	0.01-0.4 (–)
Runoff production	B	The exponent of the tension water capacity curve	0.1-10 (–)
	IMP	The ratio of the impervious to the total area of the basin	0.01-0.15 (–)
Runoff separation	SM	The areal mean of the free water capacity of the surface soil layer	10-80 (mm)
	EX	The exponent of the free water capacity curve	0.6-6 (–)
	CG	The outflow coefficients of the free water storage to groundwater	0.01-0.45 (–)
	CI	The outflow coefficients of the free water storage to interflow	0.01-0.45 (–)
Flow concentration	N	Number of reservoirs in the instantaneous unit hydrograph	0.5-10 (–)
	NK	Common storage coefficient in the instantaneous unit hydrograph	1-20 (–)
	KG	The recession constant of groundwater storage	0.6-1 (–)
	KI	The recession constant of the lower interflow storage	0.9-1 (–)

(7) Section 2.2: the need to reduce the number of Xinanjian free parameters using a sensitivity analysis must be investigated more deeply in the paper. In the current version of the paper, this model is considered with different number of free parameters depending on the modeling experiments. Why not calibrating the 15 free parameters of this rainfall-runoff model for all experiments?

Reply:

Here we add a synthetic experiment with the Xinanjiang model, where the true values of KE, CI, CG, KI, KG, and NK have periodic variations with changes every month (720h) and those of the insensitive parameters remain temporally constant. The 1-SSC-DP is applied to this experiment and all 15 free parameters are calibrated without a sensitivity analysis. The estimated parameters are plotted in Figure R1.

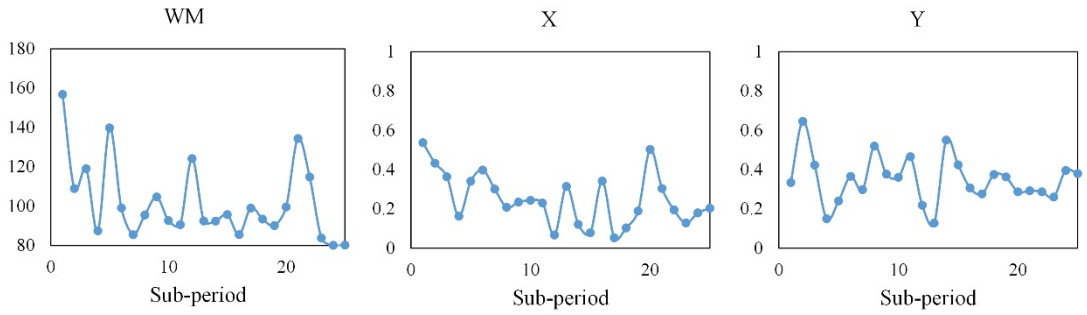


Figure R1 The parameters estimated without a sensitivity analysis

From the figure, it can be seen that except the estimated KE, CI, CG, KI, KG, and NK, the estimations of the insensitive parameters, such as WM, X and Y, are also recognized to vary significantly during the calibration period. This is inconsistent with the true values in the synthetic experiment, and the attribution analysis between time-varying parameters and watershed characteristics will be mistaken in practical use, which also occurs in the data assimilation method. Hence a sensitivity analysis is needed to find which parameters are really important for calibration.

This point has been highlighted in the Revised Manuscript as follows:

A sensitivity analysis is employed to focus efforts on parameters important to calibration and avoid prohibitive computational cost, as outlined in Sect. 2.2. **(Page 7 Line 126~127)**

(8) Section 2.3.1: one of the main hypotheses of this paper is the important "fluctuations" of the model parameter values over adjacent sub-periods, hypothesis that is not justified by the literature review, and that is not illustrated with the obtained results. This point must be discussed more deeply in the paper.

Reply:

Thanks for the comment. The main hypothesis of parameter continuity is justified as follows:

1. The hypothesis of parameter continuity can be found in the model prediction process of the ensemble Kalman filter (EnKF). Therein, the values of the parameters at the time step t+1 are forecasted by perturbing those of parameters from the time step t. The equation is as follows:

$$\vartheta_{t+1}^{k-}=\vartheta_{t}^{k+}+\delta_{t}^{k},\ \delta_{t}^{k}\sim N\left(0,R_{t}\right) \tag{1}$$

where ϑ_{t+1}^{k-} is the forecasted parameter vector at the time step t+1,while ϑ_{t}^{k+} is the well-calibrated parameter vector. δ_{t}^{k} is the white noise following a Gaussian distribution with zero mean and specified covariance of R_{t} which is very small. That is, the fluctuations between parameters of adjacent sub-periods can be little.

2. Some conceptual hydrological parameters reflect the catchment characteristics, such as soil water storage capacity in the Xinanjiang model. While climate change and human activities exert influence on catchment characteristics, the soil water storage capacity can hardly change dramatically in a very quick time, such as an hour.

Hence, it is reasonable to consider parameter continuity in estimating time-varying parameters. This point has been added in the Revised Manuscript as follows:

Some conceptual hydrological parameters reflect the catchment characteristics. While climate change and human activities exert influence on these catchment

characteristics, they can hardly change dramatically in a very quick time, such as the soil water storage capacity. **(Page 5 Line 79~82)**

(9) Evaluation criteria: Why only use the NSE criterion as only evaluation criteria, and no other criteria, such as KGE and its components? NSE appears to be nondiscriminating between considered calibration methods. Using other calibration criteria- looking at different time step and/or different error characteristics such as bias on the highest streamflow values – might be interesting in this context.

Reply:

As well as NSE coefficient, two evaluation metrics have been added in Revised Manuscript: relative error (RE) and the NSE on logarithm of streamflow (NSE_{ln}).

In the revised paper, these evaluation metrics are described as follows:

The streamflow simulations given by the proposed method are verified using the NSE, relative error (RE) and NSE on logarithm of streamflow (NSE_{ln}) (Hock, 1999). RE evaluates the error of the total volume of streamflow, while NSE and NSE_{ln} evaluate the agreement between the hydrograph of observations and simulations. NSE is more sensitive to high flows, but NSE_{ln} focuses more on low flows. Higher values of NSE, NSE_{ln} and lower absolute values of RE indicate better streamflow simulations. The NSE, RE and NSE_{ln} are expressed as followed: **(Pages 15~16, Lines 292-301)**

$$NSE = 1 - \frac{\sum_{t=1}^{m}(Q_t - \hat{Q}_t)^2}{\sum_{t=1}^{m}(Q_t - \overline{Q}_t)^2} \tag{15}$$

$$RE = \frac{\sum_{t=1}^{m}(Q_t - \hat{Q}_t)}{\sum_{t=1}^{m} Q_t} \tag{16}$$

$$NSE_{\ln} = 1 - \frac{\sum_{t=1}^{m}(\ln(Q_t) - \ln(\hat{Q}_t))^2}{\sum_{t=1}^{m}(\ln(Q_t) - \ln(\overline{Q}_t))^2} \tag{17}$$

A description of the evaluation results has been added as follows:

- For results of the synthetic experiment with the TMWB model

Figure 6(a) presents the runoff simulation performance for various scenarios. In scenario 1, the NSE values of the three SSC-DP methods are all higher than that of EnKF. The results of NSE_{ln} show no significant differences among various methods.

For scenarios 2, 4, and 6, where true parameters have linear trends, the 6-SSC-DP and 12-SSC-DP are superior to the EnKF and 3-SSC-DP in terms of NSE and NSE_{ln}. In scenario3, where the true parameters have periodic variations and change every month, the NSE and NSE_{ln} values of 6-SSC-DP and 12-SSC-DP decrease significantly, because the assumed sub-period length is longer than the time-scale of actual variations. Similarly, in scenario 5, 12-SSC-DP performs worst for NSE and NSE_{ln}, but 6-SSC-DP performs best. In scenario 7 and 8, both 6-SSC-DP and 12-SSC-DP perform better than EnKF. According to the evaluations of NSE and NSE_{ln}, the SSC-DP offers improved accuracy than the EnKF if the proper length is chosen. Another advantage of the SSC-DP is the small RE. For all scenarios, the SSC-DP methods significantly outperform for RE compared with EnKF. Among the SSC-DP methods, the RE of 3-SSC-DP is the smallest. **(Page 22, Lines 419~433)**

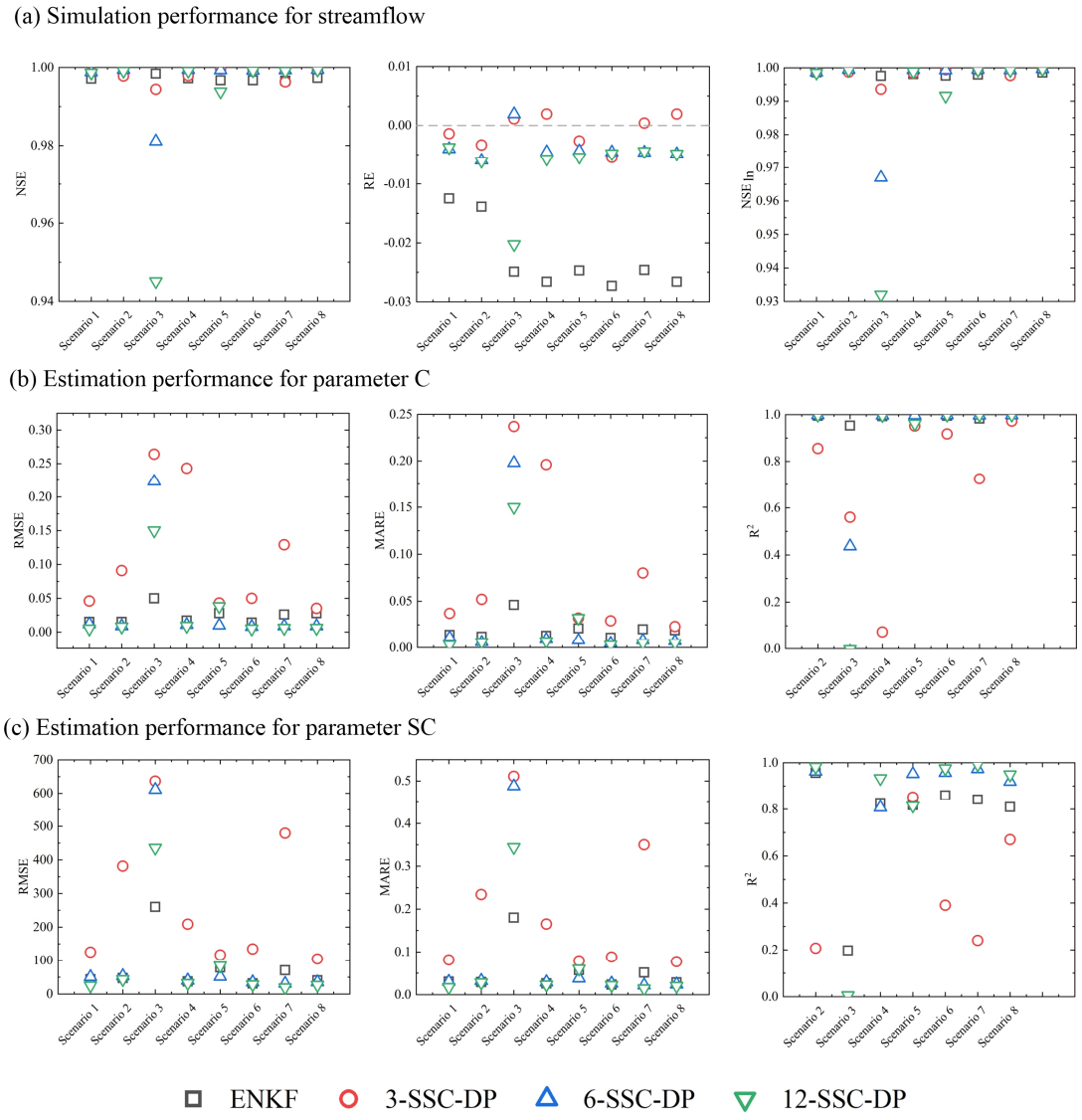


Figure 6 Comparison between the EnKF and SSC-DP methods for (a) streamflow simulation and identification of (b) parameter C and (c) parameter SC.

➢ For results of the synthetic experiment with the Xinanjiang model

The simulated streamflow and identification of time-varying parameters was compared across four methods: 1-SSC, SSC-EnKF, 1-SSC-DP, and 2-SSC-DP. The simulation performance is summarized in Figure 9(a). For all scenarios, the NSE of 2-SSC-DP is the lowest, but it performs better for low flows. The SSC-EnKF produces the highest RE in scenarios 2, 3 and 4, indicating the problem of simulating water balance. The SSC and 1-SSC-DP perform well for all scenarios in terms of NSE, RE and NSE_{ln}. Wherein, the SSC performs better than the 1-SSC-DP with regard to RE, while 1-SSC-DP is slightly superior to SSC in scenario 3 with higher NSE_{ln}. **(Page 24, Lines 472~479)**

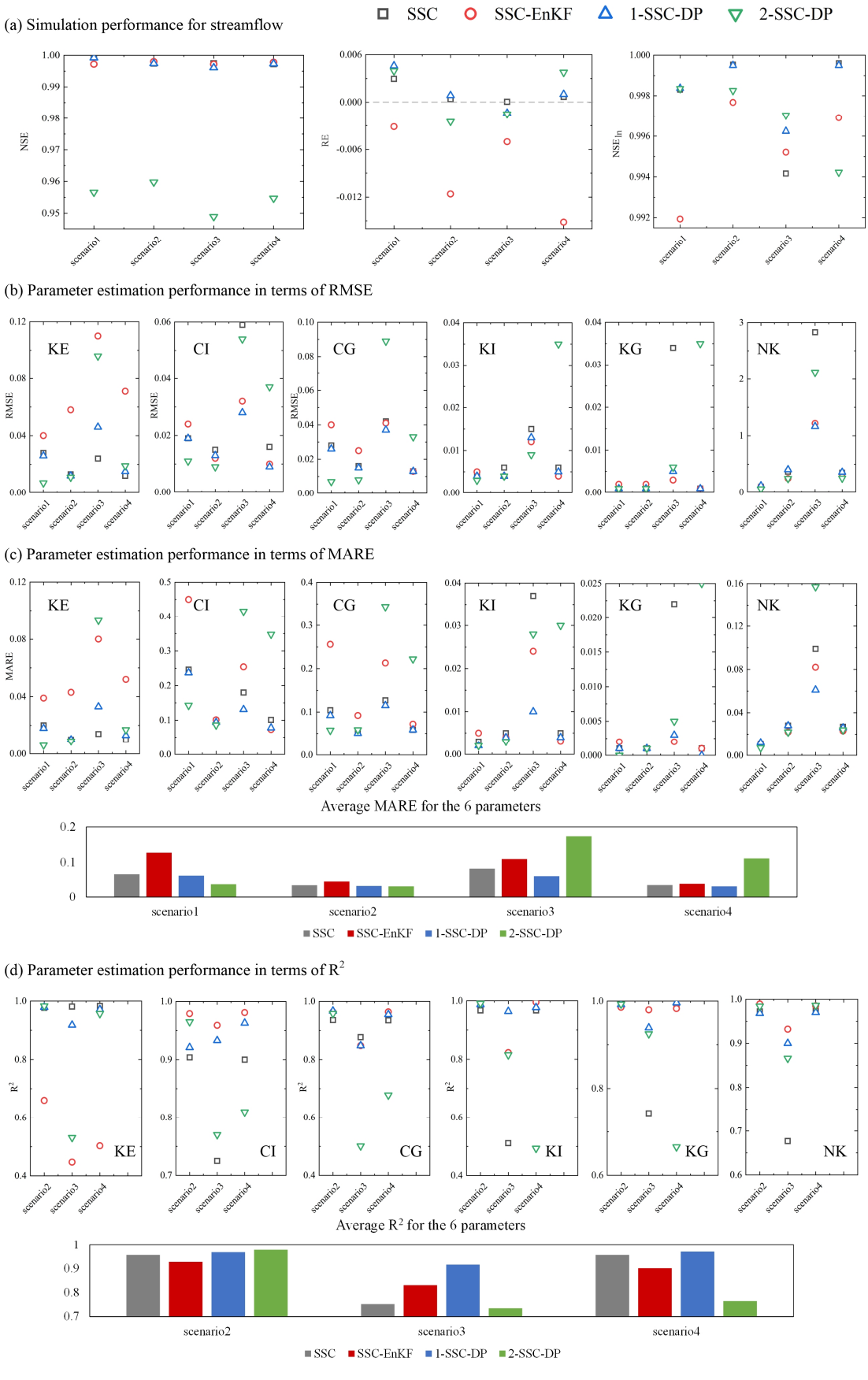


Figure 9 Comparison among the SSC, SSC-EnKF and SSC-DP methods for (a)

streamflow simulation and parameter identification in terms of (b) RMSE, (c) MARE and (d) R^2.

- For results of case study in Wuding River basin

The simulation performance is presented in Figure 12. The values of the NSEs are relatively low, because the streamflow in dry regions is difficult to simulate. It can be seen that the 12-SSC-DP gives the best simulation results among different methods with the highest NSE, NSE_{ln} and small RE. Although the 12-SSC produces relatively high NSE, it performs worst simulations for low flows. The SSC-EnKF has relatively high NSE_{ln}, but the RE of it is the largest. Overall, the 12-SSC-DP significantly improves the simulation performance of the Xinanjiang model in the Wuding River basin. **(Page 26, Lines 508~515)**

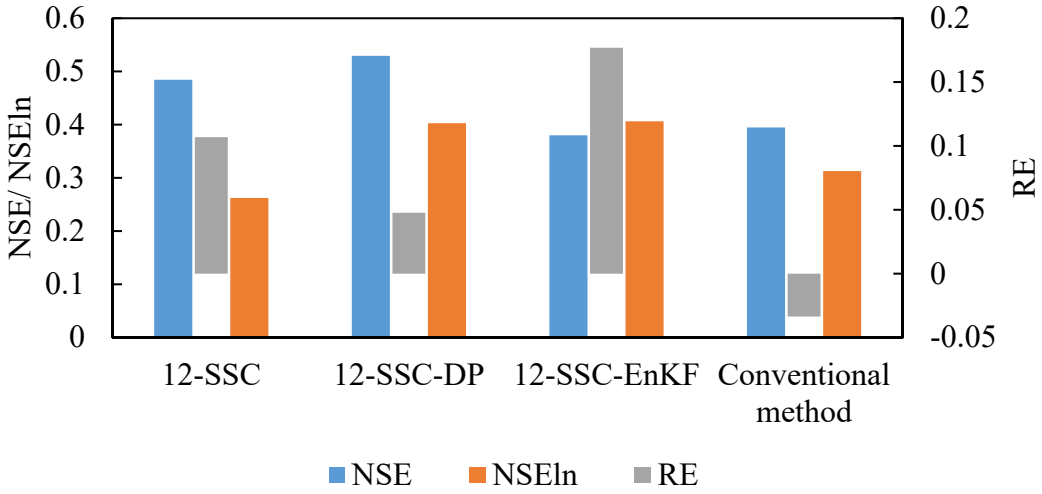


Figure 12 Simulation performance for streamflow in the Wuding River basin.

- For results of case study in Xun River basin

The simulation performance is presented in Figure 15. All methods performed well, with NSE values of 92.5 %, 93.0 %, 95.0 %, and 94.8 % for the conventional method, 3-SSC-EnKF, 3-SSC, and 3-SSC-DP, respectively. 3-SSC and 3-SSC-DP also perform well for NSE_{ln} compared with 3-SSC-EnKF and the conventional method. However, as regards to RE, the values are 0.0007 and 0.0324 for 3-SSC-DP and 3-SSC-DP, respectively. It indicated that the 3-SSC-DP can better simulate water balance than the 3-SSC in the Xun River basin. **(Page 28~29, Lines 560~567)**

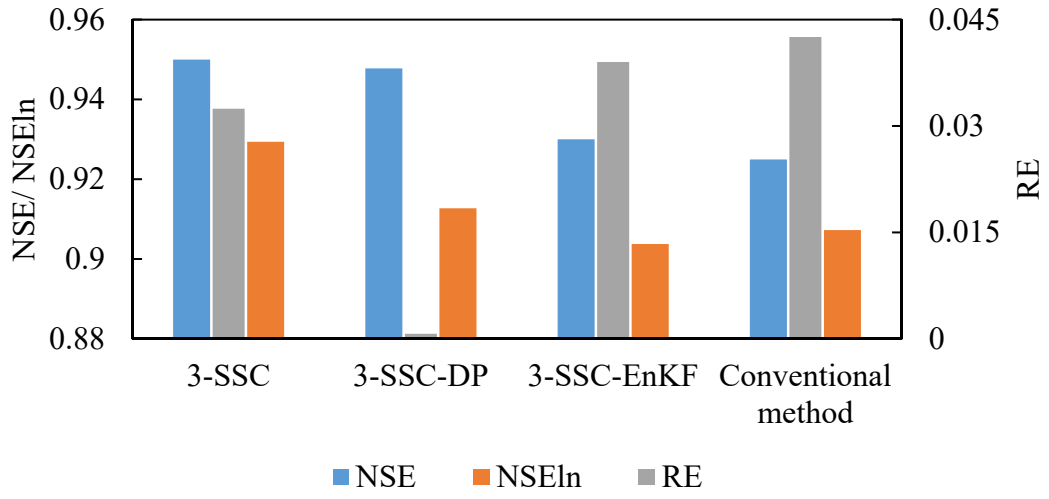


Figure 15 Simulation performance for streamflow in the Xun River basin.

(10) Section 3.2 (Wuding river basin), lines 364 to 369: the changes of the studied catchment characteristics seem to be decisive for the interpretation of the results obtained on this watershed. Nevertheless, no quantitative results / analysis of these changes are given in the paper: what is the percentage of the catchment that has been afforested? What are the number and the capacity of the built reservoirs? When are they built? Finally, an important point not discussed in the paper is the stationary and then quality of the precipitation and streamflow time series studied and used for the model calibration. This point is crucial in this context and need to be discussed.

Reply:

This comment involves two aspects:

(1) For the first aspect, a quantitative analysis of the changes in the Wuding river basin has been added in the Revised Manuscript, including the areas of tree planning and check dams for soil and water conservations. This point is described in the Revised Manuscript as follows:

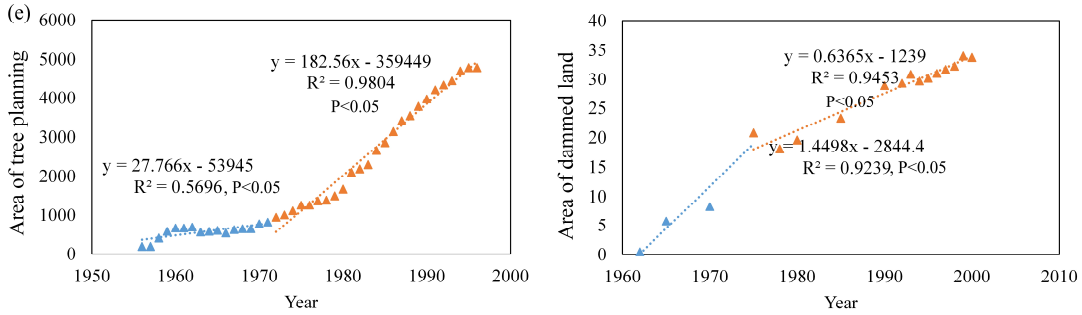


Figure 5(e) Temporal variations in the soil and water conservation measures.

Soil and water conservation measures, such as construction of the check dams and afforestation, have been undertaken since the 1960s. The areas of two soil and water conservation measures are plotted in Fig. 5(e), the data of which were collected from Zhang et al. (2002). The areas of tree planning have an increasing trend, but the slope gets much larger after 1972. It indicates that the greater efforts have been made for afforestation since the turning point. Similarly, the areas of dammed lands also increase, but the rate gets slower after 1972. These two soil and water conservation measures had changed the underlying surface of the watershed, and impacted the relationship between precipitation and runoff (Gao et al., 2017; Jiao et al., 2017). **(Page 20 Line 382~390)**

(2) The reviewer concerns the quality of the precipitation and streamflow used. The data of the daily precipitation and streamflow in the Wuding River basin are obtained from the local Hydrology and Water Resources Bureau of China, the quality of which has been checked by the official authorities, and there are no gaps among these data for all the hydrological stations. This point has been clarified in the Revised Manuscript. **(Page 20 Line 374~381)**

(11) Section 3.3 (Xun River basin): same remarks as the Wuding river basin: what about potential changes on this basin? Are precipitation and streamflow time series of good quality?

Reply:

This comment involves two aspects:

(1) The seasonal variations of the mean monthly precipitation, pan evaporation and streamflow are shown in Fig. 5(d). It shows that Xun River basin exhibits strong seasonal patterns in these climatic and hydrological variables. This point is added in the Revised Manuscript as follows:

It can be observed from Fig. 5(d) that no trend is found in annual precipitation, pan evaporation and streamflow, suggesting that the relationship between precipitation and runoff of the Xun River basin is rarely affected by human activities during 1991-2001. However, there exhibit strong seasonal patterns in these three climatic and hydrological variables, suggesting that seasonal variations in hydrological parameters should be considered. **(Page 21 Line 404~409)**

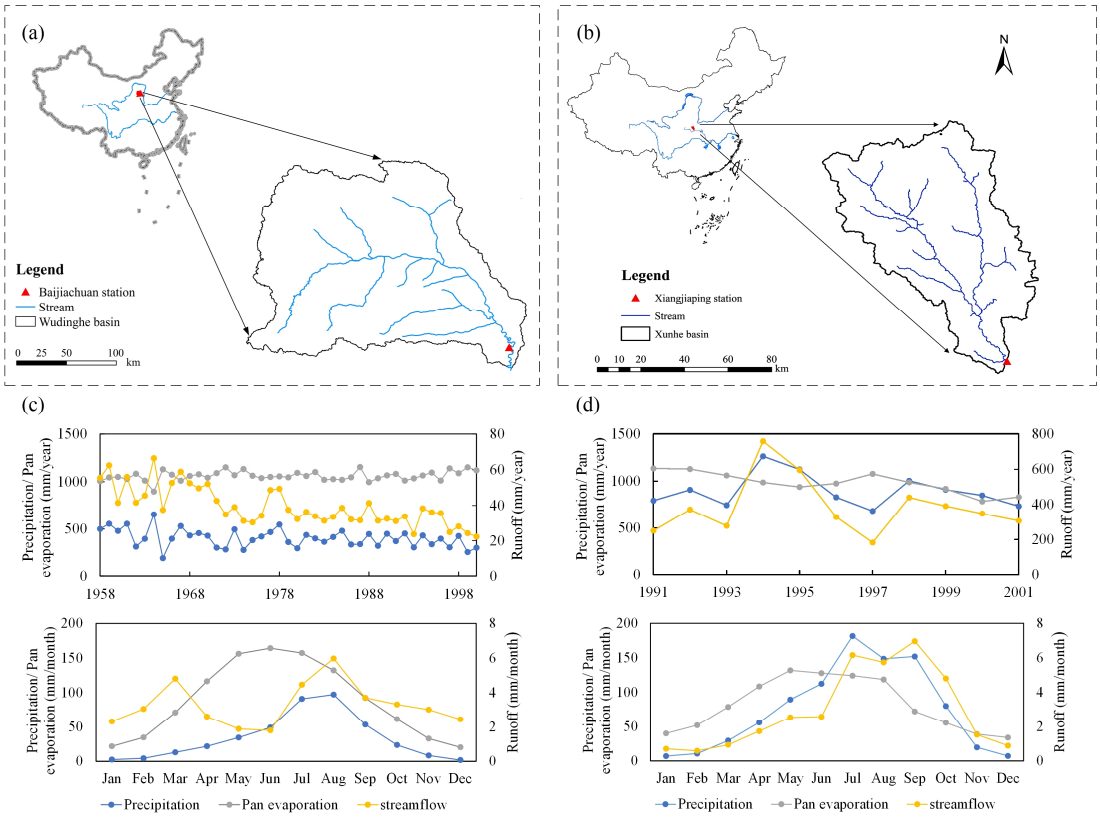


Figure 5 Location of (a) Wuding River basin and (b) Xun River basin. The plots (c) and (d) show the average yearly and monthly variations of precipitation, pan evaporation and streamflow in the Wuding River basin and Xun River basin, respectively.

(2) The data of the precipitation and streamflow in the Xun River basin are also obtained from the local Hydrology and Water Resources Bureau of China, the quality of which has been checked by the official authorities, and there are no gaps among these data for all the hydrological stations. This point has been clarified in the Revised Manuscript. **(Page 21 Line 401~403)**

(12) Section 4.1: the seasonal signal of the parameter values (cf. figures 6 and 8) must be more significantly discussed in the paper.

Reply:

More discussion concerning the seasonal signal of the parameter values is added in the Revised Manuscript as follows:

(1) For the synthetic experiment with the TMWB model

When the synthetic true parameters vary sinusoidally from month to month, EnKF gives the best estimations in scenario 3. The poor performances of 6-SSC-DP and 12-SSC-DP can be explained by the sub-period length being much longer than the actual one. When the parameters vary periodically at six-month intervals (scenario 5), 6-SSC-DP yields the best performance with the lowest RMSE, MARE and highest R2. The differences of estimation performances among 3-SSC-DP, 12-SSC-DP and EnKF are

small. The estimated parameters for scenario 5 have been plotted in Fig. 7(a). Although 3-SSC-DP and 12-SSC-DP have different lengths of sub-periods, they can also detect the correct seasonal signal of the parameters. For the annual variation in parameters (scenario 7), 12-SSC-DP and 6-SSC-DP produce better results than EnKF. Similar results can be seen in scenario 8 where C has a combined variation from year to year. In summary, the results indicate that the SSC-DP with a suitable length can estimate more accurate parameters than EnKF. **(Pages 23~24, Lines 447-459)**

(2) For the synthetic experiment with the Xinanjiang model

When the synthetic true parameters vary sinusoidally from month to month (scenario 3), the estimated parameters are plotted in Fig. 10. It can be seen that 1-SSC-DP successfully detects seasonal signal in every parameter. The SSC-EnKF performs well for R2, but it has high MARE. Although the average MARE of the SSC and 2-SSC-DP are lower than that of SSC-EnKF, the R2 of them are relatively low. Therein, from Fig. 10, the estimated parameters by the 1-SSC fluctuate generally periodically, but the variations are dramatic, resulting in lowest R2 for CI, KI, KG and NK. The estimated parameters of the 2-SSC-DP fluctuate more slowly, but the sub-period length is too long. In scenario 4, 1-SSC performs better than the SSC-EnKF and 2-SSC-DP, but is still slightly inferior to the 1-SSC-DP. Overall, the 1-SSC-DP achieves higher-quality and more robust parameter estimations performances than the other methods. **(Pages 25, Lines 484-494)**

(13) Line 456 to 458: this conclusion must be significantly moderated: the “SSC-DP” calibration method is by definition better to select more continuous parameter values.

Reply:

This conclusion has been moderated in the Revised Manuscript as follows:

Overall, the 1-SSC-DP achieves higher-quality and more robust parameter estimation performances than the other methods. **(Page 25 Line 480~494)**

(14) Section 4.2: this data analysis is crucial in this context. It might be relevant to present it in the data section. Moreover, this analysis must be significantly improved: what about potential errors (random or systematic) in the observed precipitation and streamflow series? What about potential break in the streamflow series due to rating curve changes? What is the statistical significance of this analysis? The analysis of only one catchment requires to look carefully the studied time series in the context of attribution of changes. The relative bad performance of the rainfall-runoff model on this catchment (NSE=0.41) must be discussed. In particular, the systematic streamflow underestimation for the different calibration methods must be discussed.

Reply:

This comment involves four aspects:

(1) For the first aspect, the hydrological data are collected from the local Hydrology and Water Resources Bureau of China, the systematic errors of which have been

checked by the official authorities. Additionally, random errors are considered in the synthetic experiment. The results show that 5% random errors have little influence on the SSC-DP.

(2) The reviewer concerns about the potential break in the streamflow series due to rating curve changes. The streamflow data used have been checked to guarantee their continuity, that is, no break (the discharge equal to zero) has been found except on two discontinuous days. Since the daily streamflow is also very low near the break, the values of the streamflow are reasonable.

(3) It is found that all the analyses of the linear regression are significant. The statistical significance of the analysis has been added in the Revised Manuscript and Fig. 11 as follows:

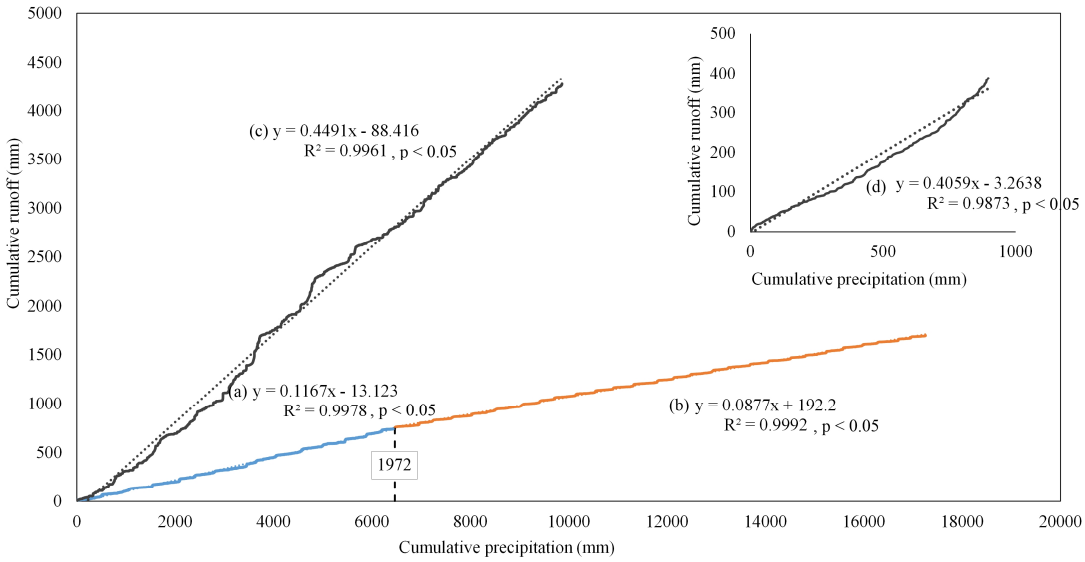


Fig. 11 Double mass curves between daily runoff and precipitation for (a) Wuding River basin from 1958–1972; (b) Wuding River basin from 1973–2000; (c) Xun River basin from 1991–2001. Subgraph (d) represents the double mass curve between the mean daily runoff and precipitation from 1991–2001.

The two linear slopes (p-value < 0.05) of the curves are different before and after 1972, demonstrating the relationship between precipitation and runoff changes under the soil and water conservation measures. **(Page 25, Lines 497~500)**

(4) The reviewer also concerns the bad simulation performance in the Wuding River basin. It is because the streamflow in dry regions is difficult to simulate. The main reason is the deficiencies of the model structure. This point is added to the Revised Manuscript. **(Page 26, Lines 508~510; Pages 26~27, Lines 522-523)**

(15) Line 492 to 495: the “unreasonable model states” between sub-periods might be illustrated in the paper.

Reply:

The statement about "the unreasonable model states" is an incorrect description and has been deleted in the Revised Manuscript. The parameters over each sub-period are calibrated separately using the SSC method. Several sets of parameters can lead to similar simulation performance in each sub-period, i.e., parameter equifinality. This equifinality causes uncertainty in simulating fluxes and streamflow.

Here, the time-series of the estimated groundwater discharge have been plotted in Fig. R2.

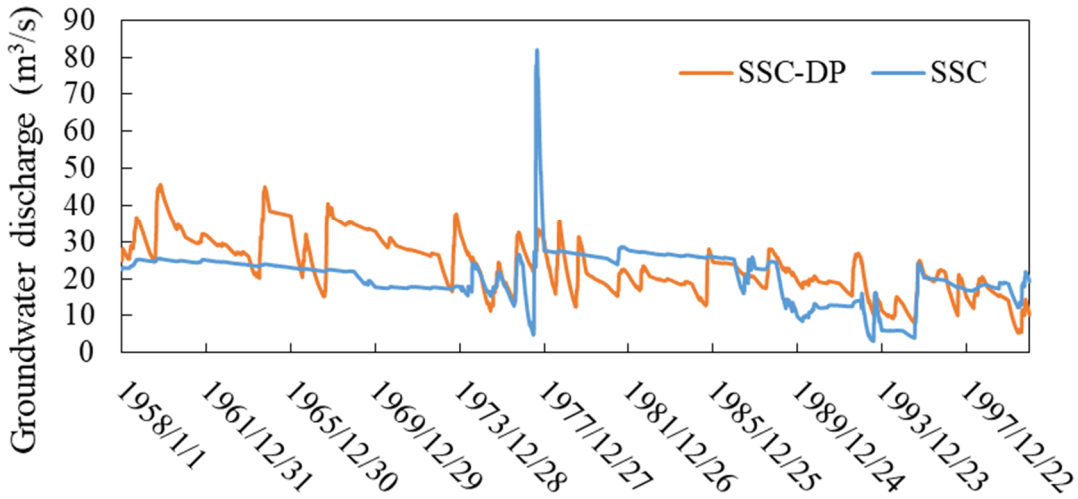


Fig. R2 The estimated groundwater discharge of the Wuding River basin

From the Fig. R2, the estimated groundwater discharge by the SSC fluctuates dramatically on December 27, 1977, which seems unreasonable, while the estimations by the SSC-DP have no dramatically fluctuations. Hence, the SSC-DP outperforms the SSC for the Wuding River basin.

(16) Line 500 to 501: this conclusion must be moderated, since results have been obtained on one catchment only.

Reply:

This sentence has been moderated in the Revised Manuscript:

It can be inferred that the 12-SSC-DP is more applicable to the simulation of streamflow in the Wuding River basin. **(Page 27 Line 527~528)**

(17) Line 511 to 514: this attribution analysis must be moderated (see previous remarks on attribution analysis).

Reply:

The attribution analysis has been moderated in the Revised Manuscript:

The results show that *WM* remains constant before and after 1972, but *WUM* varies

significantly over this period, indicating that the distribution of soil water capacity may change, i.e., *WUM* decreases but *WLM* increases. A Person correlation analysis is applied to investigate the relationship between the areas of tree planning and *WUM* as well as *WLM*. It is found that there is a significant negative correlation (Pearson correlation efficient ρ=-0.38, P<0.05) between the areas of tree planning and *WUM*. While *WLM* has a nonsignificant positive correlation (ρ=0.26, P>0.05) with the areas of tree planning. It can be inferred that less severe soil erosion occurred, because the upper layers became thinner while the lower layer, where vegetation roots dominate, became thicker (Jayawardena and Zhou, 2000). Additionally, *IMP* is significantly correlated with the areas of tree planning (ρ=-0.33, P<0.05). Except for afforestation, the areas of the dammed lands are significantly correlated with *WLM* (ρ=0.46, P<0.05), suggesting that the construction of the check dams also has influence on the soil water capacity of the Wuding river basin. Other parameters, *KE*, *KI*, *KG*, *N* and *NK* have little differences before and after 1972. The variations in *WLM* and *IMP* slowed down after the turning point, similar to the results of Deng et al. (2016). **(Pages 27~28, Lines 530-545)**

(18) Line 520 to 522: again, what about potential error in the rating curve in this context?

Reply:

The potential error in the rating curve is considered in this study from two aspects:
(1) The streamflow data are managed by the local Hydrology and Water Resources Bureau. There is a strict specification for hydrometry for drawing the rating curve. Hence, the streamflow accuracy is guaranteed.
(2) In the synthetic experiment, the uncertainty of observations has been considered, and the results show that 5% random errors have little influence on the SSC-DP.

(19) Line 526 to 539: why not presenting a Figure such as Figure 10 to illustrate rainfall-runoff simulations on this catchment?

Reply:

Thanks for the reminder. The Figure and the description have been added in the Revised Manuscript as follows:

Figure 16 illustrates the hydrograph and quantile-quantile plots for the simulations in the Xun river basin. It is evident that the peak flows estimated by the 3-SSC are higher than those of 3-SSC-DP, and 3-SSC-DP simulate better the flows ranging from 100 m^3/s to 200 m^3/s. **(Page 28~29 Line 560~570)**

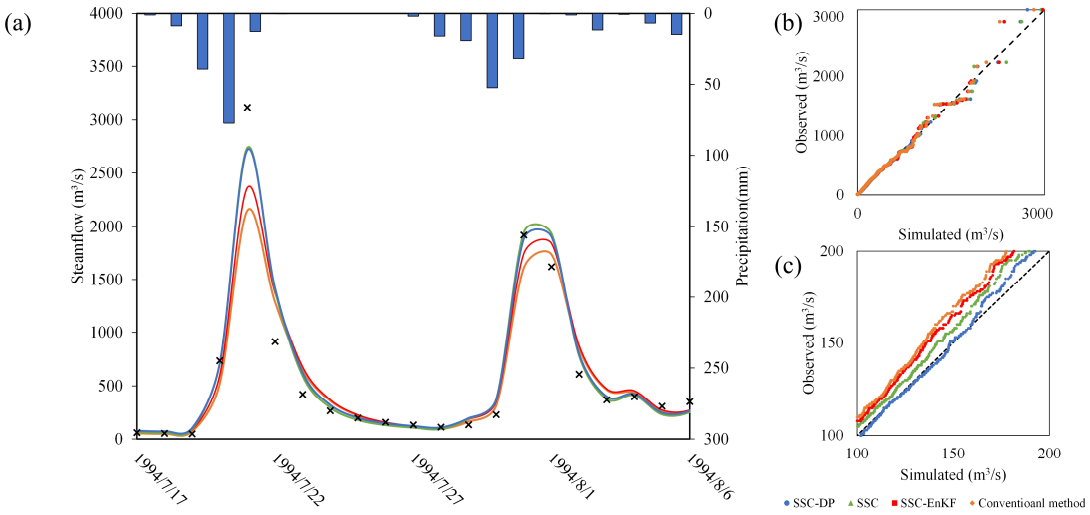


Figure 16 The simulated and observed streamflow using the conventional method, SSC-EnKF, SSC, and SSC-DP for the Xun River basin. (a) Streamflow simulation hydrograph; (b) The quantile-quantile plot for all streamflow; (c) The quantile-quantile plot for streamflow ranging from 100 m^3/s to 200 m^3/s.

(20) Line 531 to 532: is this out-performance significant?

Reply:

To give a more comprehensive evaluation, two metrics, relative error (RE) and the NSE on logarithm of streamflow (NSE_{ln}), are added. This sentence on line 531 to 532 has been modified as follows:

As regards to RE, the values are 0.0007 and 0.0324 for 3-SSC-DP and 3-SSC-DP, respectively. It indicated that the 3-SSC-DP can better simulate water balance than the 3-SSC in the Xun River basin. **(Pages 28~29, Lines 564-567)**

(21) Line 534 to 539 and line 637 to 648: again, these attribution conclusions must be moderated, because they are drawn from only two basins, without any investigation of potential systematic errors in the observed time series.

Reply:

The statement is an incorrect description and has been deleted in the Revised Manuscript. Here, the estimations of groundwater discharge are plotted in the Fig. R3. It can be seen that the estimations are similar for SSC and SSC-DP, which is different from that in the Wuding case study. It suggests that the SSC-DP gives more robust simulation performance for both case studies.

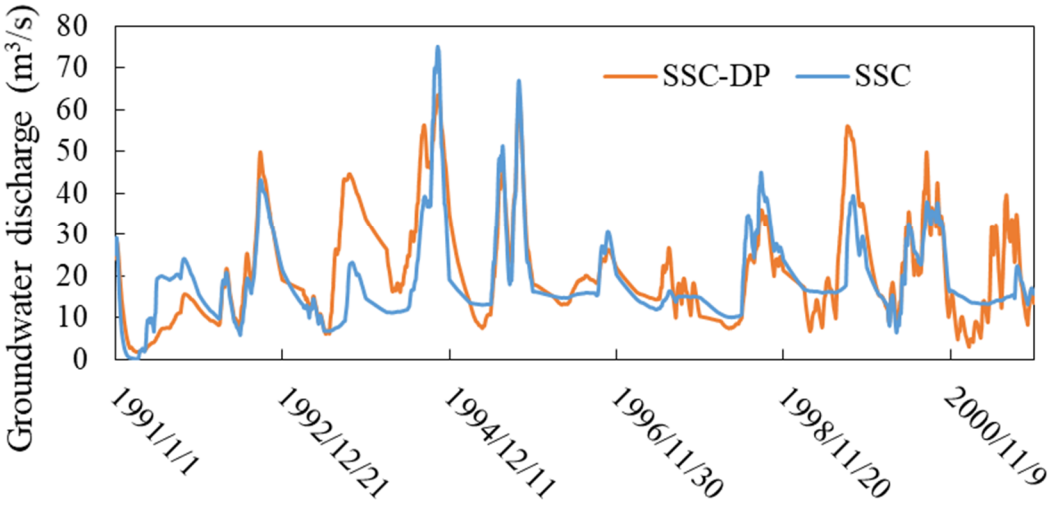


Fig. R3 The estimated groundwater discharge of the Xun River basin

Merz, R., Parajka, J. and Bloeschl, G. (2011) Time stability of catchment model parameters: Implications for climate impact analyses. Water resources research 47(W02531).

Fowler, K., Knoben, W.J.M., Peel, M.C., Peterson, T.J., Ryu, D., Saft, M., Seo, K.-W. and Western, A. (2020) Many Commonly Used Rainfall-Runoff Models Lack Long, Slow Dynamics: Implications for Runoff Projections. Water resources research 56(5).

Stephens, C.M., Marshall, L.A. and Johnson, F.M. (2019) Investigating strategies to improve hydrologic model performance in a changing climate. Journal of Hydrology 579.

Deng, C., Liu, P., Wang, W., Shao, Q. and Wang, D. (2019) Modelling time-variant parameters of a two-parameter monthly water balance model. Journal of Hydrology 573, 918-936.

A list of changes made to original manuscript in detail

1) Page 2 Line 3~7, the statements of "Following a survey of existing estimation methodologies, this paper describes a new method that combines" were modified as "Two methods, including split-sample calibration (SSC) and Data assimilation, have been used to estimate time-varying parameters … This study proposed a new method that combines".
2) Page 3 Line 31, "(Deng et al., 2019; Stephens et al., 2019)" were added.
3) Page 3 Line 37~38, "constants" and "stable" were modified as "constants in time scale" and "temporally stable", respectively.
4) Page 3 Line 44~46, "one climate condition" was modified as "wet periods"; "different conditions" was modified as "dry periods".
5) Page 5 Line 79~82, "because changes in the watershed characteristics occur over a prolonged period" is modified as "Some conceptual hydrological parameters reflect the catchment characteristics … they can hardly change dramatically in a very quick time, such the soil water storage capacity".
6) Page 7 Line 123~124, "In this section, a SSC-DP method is proposed to identify the time-varying parameters of hydrological models" was added.
7) Page 7 Line 126~127, "To avoid the prohibitive computational cost of the Xinanjiang model's calibration procedure, sensitivity analysis is employed to select behavioral parameters with less uncertainty" was modified as "A sensitivity analysis is employed to focus efforts on parameters important to calibration and avoid prohibitive computational cost".
8) Page 7 Line 128~130, "Three time-varying parameter estimation methods (SSC, SSC-DP, and data assimilation) are then used to determine the variations in these behavioral parameters, as described in Sect. 2.3." was modified as "Three time-varying parameter estimation methods (SSC, SSC-DP, and data assimilation) are presented in Sect. 2.3. The SSC and data assimilation are provided for comparisons with the SSC-DP".
9) Page 7 Line 132, "The flowchart of the methodologies is shown in Fig. 1" was added.
10) Page 8 Line 144, the unit of parameter SC was added.
11) Page 9 Line 154~155, "The 15 parameters in the Xinanjiang model are defined in Table 2" was modified as "The meaning, range and units of all the parameters in the Xinanjiang model are listed in Table 2".

12) Page 9 Line 157~158, “the TMWB model is much simpler and has fewer parameters than the Xinanjiang model” was modified as “the TMWB model has two parameters, while the Xinanjiang model has fifteen parameters”.

13) Page 15~17 Line 292~313, ” The streamflow simulations and parameter estimations given by the proposed time-varying parameter estimation approach are verified using the NSE…A Taylor diagram is used to summarize the standard deviation, RMSE, and R2 in a polar plot, providing a graphical representation of the performance of SSC-DP” was changed into “The streamflow simulations given by the proposed method are verified using the NSE, relative error (RE) and NSE on logarithm of streamflow (NSEln)…and *m* is the length of the data during the whole period”.

14) Page 17 Line 315, “Data and study area” was changed into “Synthetic experiment and real catchment case study”.

15) Page 20 Line 374~381, “The data of the daily precipitation and streamflow in the Wuding River basin were obtained from the local Hydrology and Water Resources Bureau of China …… suggesting the impacts of human activities on rainfall–runoff relationships” was added.

16) Page 20 Line 382~390, “The erosion of loess, vegetable degradation, and human activities mean that the Wuding River basin suffers severe soil erosion…Several studies have reported the anthropogenic impacts of this area and demonstrated the changing relationship between precipitation and runoff” was changed into “Soil and water conservation measures, such as construction of the check dams and afforestation, have been undertaken since the 1960s…These two soil and water conservation measures had changed underlying surface of the watershed, and impacted the relationship between precipitation and runoff”.

17) Page 21 Line 401~403, “The data in the Xun River basin were also obtained from the local Hydrology and Water Resources Bureau of China, and there are no gaps among these data for all the hydrological stations” was added.

18) Page 21 Line 404~409, “As a tributary of the Han River…Given that the majority of rainfall (approximately 70–80 % of the total) occurs in the summer, seasonal variations should also be considered” was changed into “It can be observed from Fig. 5(d) that no trend is found in annual precipitation, pan evaporation and streamflow… suggesting that seasonal variations in hydrological parameters should be considered”.

19) Page 22 Line 419~433, “Table 6 presents the runoff simulation performance for various scenarios…SSC-DP offers improved accuracy if the proper length is chosen” was changed into “Figure 6(a) presents the runoff simulation performance for various scenarios…Among the SSC-DP methods, the RE of 3-SSC-DP is the smallest”.

20) Page 22~24 Line 434~459, “Figure 5 focuses on the ability of the four methods to identify time-varying parameters…Thus, the estimated parameters are associated with higher uncertainties” was changed into “Figures 6 (b) and (c) focuses on the ability of the four methods to identify time-varying parameters…In summary, the results indicate that the SSC-DP with a suitable length can estimate more accurate parameters than EnKF”.

21) Page 24 Line 468, “Table 7” was changed into “Fig.8”.

22) Page 24 Line 473~479, “The simulation performance is summarized in Table 8…That is because, when there are more sub-periods (but the length of each sub-period is not too short), the performance tends to be better” was changed into “The simulation performance is summarized in Figure 9(a)…Wherein, the SSC performs better than the 1-SSC-DP with regard to RE, while 1-SSC-DP is slightly superior to SSC in scenario 3 with higher NSEln”.

23) Page 25 Line 480~494, “Figure 7 compares the time-varying parameter estimation performance among the four methods…However, it performs worse than 1-SSC-DP and SSC-EnKF in scenarios 3 (period) and 4 (combination)” was changed into “Figures 9(b) and (c) compares the time-varying parameter estimation performance among the four methods.…Overall, the 1-SSC-DP achieves higher quality and more robust parameter estimations performances than the other methods”.

24) Page 25 Line 496, “Figures 9(a) and (b)” was changed into “Figures 11(a) and (b)”.

25) Page 25 Line 498, “(p-value < 0.05)” was added.

26) Page 26 Line 506~515, “The simulation performance of the conventional method is presented in Fig. 10(a) (NSE = 40.7 %)…It is evident that 12-SSC and 12-SSC-DP can significantly improve the simulation performance of the Xinanjiang model in this semi-arid region” was changed into “The simulation performance is presented in Figure 12…Overall, the 12-SSC-DP significantly improve the simulation performance of the Xinanjiang model in the Wuding River basin”.

27) Page 26~27 Line 522~523, “The underestimation mainly derives from the deficiencies of the model structure” was added.

28) Page 27 Line 523, “Models” was changed into “Methods”.

29) Page 27 Line 527~528, “It can be inferred the 12-SSC-DP is more applicable to the simulation of streamflow in semi-arid regions” was modified as “It can be inferred the 12-SSC-DP is more applicable to the simulation of streamflow in the Wuding River basin”.

30) Page 27~28 Line 529~545, “The estimated time-varying parameters estimated by 12-SSC-DP are plotted in Fig. 11…The variations in WLM and IMP slowed down after the turning point, similar to the results of” was changed into “The estimated

time-varying parameters estimated by 12-SSC-DP are plotted in Fig.14…The variations in WLM and IMP slowed down after the turning point, similar to the results of".

31) Page 28 Line 547, "Figures 9(c) and (d)" was changed into "Figures 11(c) and (d)".

32) Page 28~29 Line 560~570, "All methods performed well…The superior performance in the Wuding River basin suggests that SSC-DP is more useful when simulating streamflow in dry regions (or periods)" was changed into "The simulation performance is presented in Figure 15…3-SSC-DP simulate better the flows ranging from 100 m3/s to 200 m3/s".

33) Page 31~32 Line 623~641, "As reported in Sect. 4…the preferred length is six months" was changed into "It is suggested that the determination of the sub-period length considers three factors…if the difference between the NSE values of 6-SSC-DP and 3-SSC-DP is small, the preferred length is 6-month".

34) Page 33 Line 656~673, "One synthetic experiment used the TMWB model with two parameters and eight scenarios … between the temporal variations of parameters and the changing environment in real catchments." was changed into "The proposed method with a suitable length not only produces better simulation performance … indicating the distinct impacts of seasonal climate variability".

35) Page 39, the units were added to Table 1.

36) Page 40, the units were added to Table 2

37) Page 44, Figure 1 was added.

38) Page 48, Figure 5 was modified.

39) Page 49, Table 6 was deleted and Figure 6 was modified.

40) Page 51, Table 7 was deleted and Figure 8 was added.

41) Page 52, Table 8 was deleted and Figure 9 was modified.

42) Page 54, Figure 11 was modified.

43) Page 55, Figure 12 was added.

44) Page 56, Figure 13 was added.

45) Page 58, Figure 15 was added.

44) Page 59, Figure 16 was added.

A time-varying parameter estimation approach using split-sample calibration based on dynamic programming

Xiaojing Zhang[a,b], Pan Liu[a,b*]

[a]State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China

[b]Hubei Provincial Key Lab of Water System Science for Sponge City Construction, Wuhan University

*Corresponding author. Email: liupan@whu.edu.cn;

Tel: +86-27-68775788; Fax: +86-27-68773568

Abstract: Although the parameters of hydrological models are usually regarded as constant, temporal variations can occur in a changing environment. Thus, effectively estimating time-varying parameters becomes a significant challenge. Two methods, including split-sample calibration (SSC) and Data assimilation, have been used to estimate time-varying parameters. However, SSC is unable to consider the parameter temporal continuity, while Data assimilation assumes parameters vary at every time-step. This study proposed a new method that combines (1) the basic concept of split-sample calibration (SSC), whereby parameters are assumed to be stable for one sub-period, and (2) the parameter continuity assumption, i.e., the differences between parameters in consecutive time steps are small. Dynamic programming is then used to determine the optimal parameter trajectory by considering two objective functions: maximization of simulation accuracy and maximization of parameter continuity. The efficiency of the proposed method is evaluated by two synthetic experiments, one with a simple two-parameter monthly model and the second using a more complex 15-parameter daily model. The results show that the proposed method is superior to SSC alone, and outperforms the ensemble Kalman filter if the proper sub-period length is used. An application to the Wuding River basin indicates that the soil water capacity parameter varies before and after 1972, which can be interpreted according to land use and land cover changes. Further application to the Xun River basin shows that parameters are generally stationary on an annual scale, but exhibit significant changes over seasonal scales. These results demonstrate that the proposed method is an effective tool for identifying time-varying parameters in a changing environment.

Keywords: hydrological model; time-varying parameter; calibration; dynamic programming

1. Introduction

Conceptual models describe the physical processes that occur in the real world by means of certain assumptions and empirically determined functions (Toth and Brath, 2007). In spite of their simplicity, conceptual models are effective in providing reliable runoff predictions for widespread applications (Quoc Quan et al., 2018; Refsgaard and Knudsen, 1996), such as real-time flood forecasting, climate change impact assessments (Deng et al., 2019; Stephens et al., 2019), and water resources management. Conceptual hydrological models typically have several inputs, a moderate number of parameters, state variables, and outputs. Among these, the parameters play an important role in accurate simulation and should be related to the catchment properties. However, parameter values often cannot be obtained by field measurements (Merz et al., 2011). An alternative approach is to calibrate parameters based on historical data.

Parameters are usually regarded as constants in time scale, because of the general idea that catchment conditions are temporally stable. Constant parameters become inaccurate in differential split sample test (DSST) conditions (Klemes, 1986). For example, parameters calibrated based on data from a wet (or dry) period may fail to simulate runoff in a dry (or wet) period for the same catchment. Broderick et al. (2016) used DSST to assess the transferability of six conceptual models under contrasting climate conditions. They found that performance declines most when models are calibrated during wet periods but validated in dry ones. Fowler et al. (2016) pointed out that the parameter set obtained by mathematical optimization based on wet periods may not be robust when applied in dry periods. Additionally, the catchment properties can

change over time, such as in the case of afforestation and deforestation (Guzha et al., 2018; Siriwardena et al., 2006). These changes need to be taken into account through model parameters (Bronstert, 2004; Hundecha and Bardossy, 2004). Hence, temporal variations in parameters should reflect the changing environment.

One challenge here is the methodology used to identify time-varying parameters. In the literature, three approaches have been discussed. The first is split-sample calibration (SSC), whereby available data are split into a moderate number of sub-periods and the parameters are calibrated individually for each period (Thirel et al., 2015). The second method is data assimilation (Deng et al., 2016; Pathiraja et al., 2018). This method assimilates observational data to enable errors, states, and parameters to be updated (Li et al., 2013), making it possible to identify time-varying parameters. The third approach is to construct a functional form or empirical equation according to the correlation between parameters and some climatic variates such as precipitation and potential evapotranspiration (Deng et al., 2019; Jeremiah et al., 2013; Westra et al., 2014). Note that this study focuses on methods to identify time-varying parameters rather than modelling them; hence, only comparisons between SSC and data assimilation are discussed.

SSC is the most commonly used method (Coron et al., 2012; Fowler et al., 2018; Paik et al., 2005; Xie et al., 2018). Merz et al. (2011) investigated the time stability of parameters by estimating six parameter sets based on six consecutive five-year periods. Lan et al. (2018) clustered calibration data into 24 sub-annual periods to detect the seasonal hydrological dynamic behavior. Despite broad application, it remains

69 debatable whether a particular mathematical optimum gives the parameter value during
70 one period. Many equivalent optima can exist simultaneously for one dataset when
71 calibrating the model against observations (Poulin et al., 2011). Several studies
72 addressed this question by adding more constraints to the objective function over the
73 respective period. For example, Gharari et al. (2013) emphasized consistent
74 performance in different climatic conditions, while Xie et al. (2018) modified SSC by
75 selecting parameters with good simulation ability for both the current sub-period and
76 the whole period. However, few reports have considered the continuity of parameters
77 in the SSC method.

78 Continuity requires differences between the parameters in consecutive time steps
79 to be small. Some conceptual hydrological parameters reflect the catchment
80 characteristics. While climate change and human activities exert influence on these
81 catchment characteristics, they can hardly change dramatically in a very quick time,
82 such as the soil water storage capacity. This assumption is the basic idea behind data
83 assimilation methods. For example, the a priori parameters in ensemble Kalman filter
84 (EnKF) methods are commonly derived from updated values from the previous time
85 step (Moradkhani et al., 2005; Xiong et al., 2019). From this, a trade-off between
86 simulation accuracy and parameter continuity is established, and parameters that enable
87 greater continuity are more likely to be selected. Deng et al. (2016) validated the ability
88 of the EnKF to identify changes in two-parameter monthly water balance (TMWB)
89 model parameters. Pathiraja et al. (2016) proposed two-parameter evolution models for
90 improving conventional dual EnKF, and obtained superior results for diagnosing the

91 non-stationarity in a system. EnKF and its variants are relatively advanced approaches
92 for identifying time-varying parameters (Lu et al., 2013). However, for a hydrological
93 model, the states may change over every time step, whereas the parameters may not, in
94 particular for hourly time scales. This can be offset by SSC, which assumes that the
95 parameters retain stable for a pre-determined period (such as decades, years, or months).
96 Compared to EnKF, the simplicity of SSC is another advantage, as it has a less complex
97 mechanism and reduced redundancy (Chen and Zhang, 2006).

98 The aim of this study is to present a new method for time-varying parameter
99 estimation by combining the strengths of the basic concept of SSC and the continuity
100 assumption of data assimilation, which is a useful tool for diagnosing the non-
101 stationarity caused by a changing environment. Compared with data assimilation, the
102 proposed split-sample calibration based on dynamic programming (SSC-DP) avoids
103 overly frequent changes of parameters, such as hourly or daily variations. Compared
104 with SSC, the distinctive element is that SSC-DP considers the parameters to be related
105 over adjacent sub-periods, and selects parameter sets with good performance for each
106 period and small differences between adjacent time steps. In this study, three aspects of
107 the proposed method are evaluated: (1) The performance of SSC-DP is compared with
108 that of existing methods in terms of the estimation of time-varying parameters; (2) The
109 applicability of SSC-DP to more complex hydrological models with a considerable
110 number of parameters; (3) The ability of SSC-DP to provide additional insights on
111 parameter variations and their correlations with the properties of real catchments. To
112 investigate the above issues, the proposed method is compared with SSC and EnKF in

two synthetic experiments (one with a two-parameter monthly model, the other with a 15-parameter daily model). SSC-DP is also applied to two real catchments for parameter estimation under different environmental conditions.

The remainder of this paper is organized as follows. Section 2 describes the proposed method, reference methods, and performance evaluation indices. Section 3 describes two synthetic experiments and two real catchment case studies for comparison among different time-varying parameter estimation methods. Sections 4 and 5 present the results and discussion, respectively, before the conclusions to this study are drawn in Sect. 6.

2. Methodology

In this section, a SSC-DP method is proposed to identify the time-varying parameters of hydrological models. The two hydrological models considered in this study are the TMWB and Xinanjiang models. Their concepts and differences are presented in Sect. 2.1. A sensitivity analysis is employed to focus efforts on parameters important to calibration and avoid prohibitive computational cost, as outlined in Sect. 2.2. Three time-varying parameter estimation methods (SSC, SSC-DP, and data assimilation) are presented in Sect. 2.3. The SSC and data assimilation are provided for comparisons with the SSC-DP. Finally, to evaluate the performance of the time-varying parameter estimation methods, six evaluation criteria are selected and formulated in Sect. 2.4. The flowchart of the methodologies is shown in Fig. 1.

2.1 Hydrological models

2.1.1 Two-parameter monthly water balance model

The TMWB model developed by Xiong and Guo (1999) is efficient for monthly runoff simulations and forecasts (Dai et al., 2018; Guo et al., 2002; Kim et al., 2016; Yang et al., 2017). The model requires monthly precipitation and potential evapotranspiration as inputs. Its simplicity and efficiency of performance mean that TMWB can easily be used to investigate the impacts of climate change (Deng et al., 2016; Luo et al., 2019). Its outputs include monthly streamflow, actual evapotranspiration, and soil moisture content index. The model has only two parameters (Table 1), *C* and *SC*. The parameter *C* takes account of the effect of the change of time scale when simulating actual evapotranspiration. The parameter *SC* represents the field capacity (mm).

2.1.2 Xinanjiang model

The Xinanjiang model (Zhao, 1992) is widely used in China (Li and Zhang, 2017; Si et al., 2015; Yin et al., 2018). It takes precipitation and pan-evaporation data as inputs and estimates the actual evapotranspiration, soil moisture storage, surface runoff, interflow, and groundwater runoff from the watershed. The simulated streamflow is calculated by summing the routing results of the surface, interflow, and groundwater runoff (Sun et al., 2018). In this study, the surface runoff is routed by the instantaneous unit hydrograph (Lin et al., 2014), while the interflow and groundwater runoff are routed by the linear reservoir method (Jayawardena and Zhou, 2000). A schematic

overview of the model is presented in Fig. 2. The meaning, range and units of all the parameters in the Xinanjiang model are listed in Table 2.

There are two important differences between the TMWB and Xinanjiang models: (1) the TMWB model has two parameters, while the Xinanjiang model has fifteen parameters; (2) TMWB is a monthly rainfall-runoff model, whereas the Xinanjiang model can run on hourly or daily step sizes.

2.2 Parameter sensitivity analysis method

Sensitivity analysis is used to identify which parameters significantly affect the performance of the Xinanjiang model and reduce the number of parameters to be calibrated. Numerous sensitivity analysis methods are available, such as the Morris method (Morris, 1991) and Sobol analysis (Sobol, 1993). The Morris method provides similar results to Sobol analysis with a reduced computational burden (Rebolho et al., 2018; Teweldebrhan et al., 2018; Yang et al., 2018).

The Morris method assumes that if parameters change by the same relative amount, the parameter that causes the larger elementary effect is the more sensitive (King and Perera, 2013). The elementary effect is calculated as follows:

$$EE_p(\theta_1,\theta_2,...,\theta_{Np},\Delta)=\frac{y(\theta_1,\theta_2,...,\theta_{p-1},\theta_p+\Delta,\theta_{p+1},...,\theta_{Np})-y(\theta_1,\theta_2,...,\theta_{Np})}{\Delta} \quad (1)$$

where θ_p represents the p-th parameter; Δ is the relative amount; Np is the total number of parameters, and y is the model output based on a particular parameter set.

Each parameter is changed in turn and every parameter set produces an elementary effect. The parameter sensitivity is evaluated using the mean value μ of the

elementary effects. If a parameter has a higher value of μ, it is more sensitive. In fact, interactions between parameters should be taken into account (Jie et al., 2018). Hence, the standard deviation σ can be calculated. A higher value of σ indicates a stronger nonlinear correlation between parameters (Pappenberger et al., 2008).

2.3 Time-varying parameter estimation method

2.3.1 Split-sample calibration

SSC provides a simple way of diagnosing parameter non-stationarity under a changing environment (Merz et al., 2011). As illustrated in Fig. 3(a), the method usually has two steps (Hughes, 2015; Kim et al., 2015): (1) Available data are divided into several consecutive periods, which can be arbitrarily chosen as hours, days, months, seasons, or years; (2) Parameters are calibrated separately for the respective period. This procedure gives better simulation performance than using constant parameters, but leads to the estimated parameters fluctuating strongly over adjacent sub-periods, producing false temporal variants.

2.3.2 Split-sample calibration based on dynamic programming

To overcome this problem, the SSC-DP method identifies time-varying parameters with consideration of temporal continuity. SSC-DP has five steps (Fig. 3(b)):

(1) Split-sample periods. This process is the same as the first step of the SSC.

(2) Feasible parameter space generation. An ensemble of nearly optimal parameter sets for each sub-period is obtained using Markov chain Monte Carlo (MCMC) sampling (Chib and Greenberg, 1995). The likelihood measure of the *i*-th sub-period

links the parameter to observations using the Nash–Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) as follows:

$$L_i(\theta) = 1 - \frac{\sum_{t=(i-1)\times l+1}^{i\times l} (Q_t - \hat{Q}_t)^2}{\sum_{t=(i-1)\times l+1}^{i\times l} (Q_t - \overline{Q_t})^2} \tag{2}$$

where Q_t and $\hat{Q}_t$ are the observed and simulated runoff at time step t, respectively, and l is the length of the sub-period.

(3) Dynamic programming optimization. The goal is to find parameters that provide both good model performance and continuity. The continuity condition aims to minimize the difference between the estimated parameters for sub-periods *i* and *i*+1. For *N* sub-periods, the objective function can be expressed as follows:

$$\text{Max } F = \sum_{i=1}^{N} [(\text{NSE}_i + \text{NSE}_{ln,i} + \text{NSE}_{abs,i}) - \alpha \times \sum_{p=1}^{N_P} \frac{|\theta_{i+1,p} - \theta_{i,p}|}{\theta_{max,p} - \theta_{min,p}}] \tag{3}$$

$$NSE_{\ln,i} = 1 - \frac{\sum_{t=(i-1)\times l+1}^{i\times l} (\ln(Q_t) - \ln(\hat{Q}_t))^2}{\sum_{t=(i-1)\times l+1}^{i\times l} (\ln(Q_t) - \ln(\overline{Q_t}))^2} \tag{4}$$

$$NSE_{abs,i} = 1 - \frac{\sum_{t=(i-1)\times l+1}^{i\times l} \left|Q_t - \hat{Q}_t\right|}{\sum_{t=(i-1)\times l+1}^{i\times l} \left|Q_t - \overline{Q_t}\right|} \tag{5}$$

where $\theta_{i,p}$ is the *p*-th estimated parameter over the *i*-th sub-period; $\theta_{max,p}$ and $\theta_{min,p}$ are its maximum and minimum values, respectively; N_P is the number of the parameters; and α is the weight, reflecting parameter continuity. The weights of NSE_i, $\text{NSE}_{ln,i}$, and $\text{NSE}_{abs,i}$ are set to 1 following the work of Merz et al. (2011), who used equal weights for the NSE and its variants.

As the decision-making process during the current sub-period is related to that of the previous sub-period, the parameter estimation over N periods becomes a multi-stage optimization problem. To solve this, a dynamic programming technique (Bellman, 1957) is employed to decompose the optimization into a number of single-stage problems and determine the optimal trajectory of the time-varying parameters. Dynamic programming is a useful method for handling sequential operation decisions. It allows the problem to be solved using a backward recursive procedure, whereby the decision-making for each sub-period maximizes the sum of current and future benefits (Li et al., 2018; Ming et al., 2017). In this study, the objective function is formulated as the following recursive equation:

$$\begin{cases} F_i^* = max\{f_i[\vartheta_{i,1}, \vartheta_{i,2}, \vartheta_{i,3}, \cdots, \vartheta_{i,p}] + F_{i+1}^*\} \\ F_N^* = 0 \end{cases} \quad (6)$$

where F_i^* is the evaluation index using the optimal time-varying parameters from the N-th to the i-th sub-periods, and Eq. (6) calculates the objective function from the N-th sub-period to the first sub-period.

(4) Update initial states. The initial states, such as that of the soil water content, are important in model simulation and calibration. As the final states for sub-period i are not used as the initial states for sub-period i+1 during steps (1)–(3), the time-varying parameter set obtained from step (3) is applied to the hydrological model to update the initial states of each sub-period for the next iteration.

(5) Steps (1)–(4) are repeated until the initial states of each sub-period are generally stable.

2.3.3 Data assimilation

Another approach for diagnosing variations in parameters is data assimilation, using methods such as the EnKF and ensemble Kalman smoother (EnKS). These are used here as reference methods. The EnKF has been widely applied to conceptual models, including TMWB (Deng et al., 2016). Li et al. (2013) noted that the EnKF struggles to handle the time-lag in routing processes. However, the routing component is vital to the Xinanjiang model. EnKS can efficiently determine the states of the Xinanjiang model (Meng et al., 2017), but the estimation of routing parameters deserves discussion. Most previous studies have used a fixed distribution of the routing hydrograph in data assimilation (Lu et al., 2013), i.e., the parameters are constant for routing processes. With respect to these issues, a modified EnKF (named SSC-EnKF) is established as a third data assimilation reference method in the synthetic experiment with the Xinanjiang model (described in Sect. 3.1).

The EnKF includes two main steps: model prediction and assimilation. The state vector is augmented with parameter variables so that time-varying parameters can be estimated simultaneously with model states. For model prediction, the augmented vector is derived by adding noise on that from the previous time step through the following equation:

$$\begin{pmatrix} \vartheta_{t+1}^{k-} \\ x_{t+1}^{k-} \end{pmatrix} = \begin{pmatrix} \vartheta_{t}^{k+} \\ f\left(x_t^{k+}, \theta_{t+1}^{k-}, u_{t+1}\right) \end{pmatrix} + \begin{pmatrix} \delta_t^k \\ \varepsilon_t^k \end{pmatrix},\ \delta_t^k \sim N\left(0, R_t\right), \varepsilon_t^k \sim N\left(0, G_t\right) \tag{7}$$

where ϑ_t is the parameter vector at time step t, represented as $\left(\theta_{t,1}, \theta_{t,2}, ..., \theta_{t,Np}\right)$; x_t is the state vector; ϑ_{t+1}^{k-} and x_{t+1}^{k-} are the *k*-th ensemble member forecasts at time step

t+1; ϑ_t^{k+} and x_t^{k+} are the updated values of the k-th ensemble member forecasts at time step t; u_{t+1} denotes the forcing data (e.g., precipitation) at time step t+1; δ_t^k and ε_t^k are the white noise for the k-th ensemble member, which follow a Gaussian distribution with zero mean and specified covariance of R_t and G_t, respectively.

In the assimilation process, the augmented vector is updated using the following equations if suitable observations are available:

$$\begin{pmatrix} x_{t+1}^{k+} \\ \vartheta_{t+1}^{k+} \end{pmatrix} = \begin{pmatrix} x_{t+1}^{k-} \\ \vartheta_{t+1}^{k-} \end{pmatrix} + \begin{pmatrix} K_{t+1}^{x} \left[y_{t+1}^{k} - \hat{y}_{t+1}^{k} \right] \\ K_{t+1}^{\vartheta} \left[y_{t+1}^{k} - \hat{y}_{t+1}^{k} \right] \end{pmatrix} \tag{8}$$

$$y_{t+1}^{k} = y_{t+1} + \xi_{t+1}^{k},\ \xi_{t+1}^{k} \sim N\left(0, W_t\right), \tag{9}$$

$$\hat{y}_{t+1}^{k} = h(x_{t+1}^{k-}, \vartheta_{t+1}^{k-}) \tag{10}$$

where y_{t+1} is the observation vector at time t+1; y_{t+1}^{k} is the k-th observation ensemble member at time step t+1; $\hat{y}_{t+1}$ is the simulation vector at time t+1; h is the observational operator that converts the model states to observations; ξ_{t+1}^{k} is the measurement error, which follows a Gaussian distribution with a covariance of W_t; and K_{t+1}^{k} is the Kalman gain matrix (for details, see (Feng et al., 2017).

The EnKS is based on the EnKF. Whereas the EnKF updates the model states and parameters at the current time step, the EnKS takes account of those values over the past time steps. The main steps of the EnKS are identical to those of the EnKF, but the equation of the assimilation process is formulated as follows:

$$\begin{pmatrix} x_{t+1\to t-n+2}^{k+} \\ \vartheta_{t+1\to t-n+2}^{k+} \end{pmatrix} = \begin{pmatrix} x_{t+1\to t-n+2}^{k-} \\ \vartheta_{t+1\to t-n+2}^{k-} \end{pmatrix} + \begin{pmatrix} K_{t+1}^{x*} \left[y_{t+1}^{k} - \hat{y}_{t+1}^{k} \right] \\ K_{t+1}^{\vartheta*} \left[y_{t+1}^{k} - \hat{y}_{t+1}^{k} \right] \end{pmatrix} \tag{11}$$

$$\hat{y}_{t+1}^{k} = h(x_{t+1\to t-n+2}^{k-}, \vartheta_{t+1\to t-n+2}^{k-}) \tag{12}$$

where K_{t+1}^{*} is the Kalman gain matrix of EnKS. The fixed time window n of EnKS

is pre-determined based on the response function or unit hydrograph. Meng et al. (2017) suggested that the time window should be set as half of the recession time of a flood.

A third data assimilation approach is constructed based on the SSC. Instead of assimilating one observed variable, it assimilates the observed variables during a given period in one assimilation process. Assuming that the parameters are constant in the given period, the equation of the assimilation process for the i-th sub-period is expressed as follows:

$$\begin{pmatrix} x_{i+1}^{k+} \\ \vartheta_{i+1}^{k+} \end{pmatrix} = \begin{pmatrix} x_{i+1}^{k-} \\ \vartheta_{i+1}^{k-} \end{pmatrix} + \begin{pmatrix} K_{i+1}^{x^*} \left[y_{i\times l+1\to(i+1)\times l,}^{k} - \hat{y}_{i\times l+1\to(i+1)\times l}^{k} \right] \\ K_{i+1}^{\vartheta^*} \left[y_{i\times l+1\to(i+1)\times l}^{k} - \hat{y}_{i\times l+1\to(i+1)\times l}^{k} \right] \end{pmatrix} \tag{13}$$

$$\hat{y}_{i\times l+1\to(i+1)\times l}^{k} = h(x_{i+1}^{k-}, \vartheta_{i+1}^{k-}) \tag{14}$$

where ϑ_i is the parameter vector for sub-period i, represented as $(\theta_{i,1}, \theta_{i,2}, ..., \theta_{i,Np})$; x_i is the initial state vector for sub-period i; and l is the length of the sub-period.

This approach addresses the routing-lag issue by allowing parameters of the routing processes, such as the instantaneous unit hydrograph, to remain constant for each sub-period and to be time-varying over the whole period.

2.4 Model evaluation criteria

The streamflow simulations given by the proposed method are verified using the NSE, relative error (RE) and NSE on logarithm of streamflow (NSE_{ln}) (Hock, 1999). RE evaluates the error of the total volume of streamflow, while NSE and NSE_{ln} evaluate the agreement between the hydrograph of observations and simulations. NSE is more sensitive to high flows, but NSE_{ln} focuses more on low flows. Higher values

297 of NSE, NSE_{ln} and lower absolute values of RE indicate better streamflow simulations.

298 The NSE, RE and NSE_{ln} are expressed as followed:

299
$$NSE = 1 - \frac{\sum_{t=1}^{m}(Q_t - \hat{Q}_t)^2}{\sum_{t=1}^{m}(Q_t - \overline{Q}_t)^2} \tag{15}$$

300
$$RE = \frac{\sum_{t=1}^{m}(Q_t - \hat{Q}_t)}{\sum_{t=1}^{m} Q_t} \tag{16}$$

301
$$NSE_{\ln} = 1 - \frac{\sum_{t=1}^{m}(\ln(Q_t) - \ln(\hat{Q}_t))^2}{\sum_{t=1}^{m}(\ln(Q_t) - \ln(\overline{Q}_t))^2}$$

302 The estimated parameters are evaluated by the RMSE (Alvisi et al., 2006), MARE

303 (Khalil et al., 2001) and R^2 (Kim et al., 2007). RMSE and MARE quantify the accuracy

304 of the estimated parameters, but RMSE is more sensitive to high values than MARE.

305 R^2 records the overall agreement between the true and estimated parameters. Smaller

306 values of RMSE, MARE and higher values of R^2 indicate stronger parameter

307 identification ability. For the *p*-th parameter, the formulations are as follows:

308
$$RMSE_p = \sqrt{\frac{1}{m}\sum_{t=1}^{m}(\theta_{t,p} - \hat{\theta}_{t,p})^2} \tag{18}$$

309
$$MARE_p = \frac{1}{m}\sum_{t=1}^{m}\frac{\left|\theta_{t,p} - \hat{\theta}_{t,p}\right|}{\theta_{t,p}} \tag{19}$$

310
$$R^2{}_p = \frac{\sum_{t=1}^{m}\left(\hat{\theta}_{t,p} - \overline{\hat{\theta}}_p\right)\left(\theta_{t,p} - \overline{\theta}_p\right)}{\sqrt{\sum_{t=1}^{m}\left(\hat{\theta}_{t,p} - \overline{\hat{\theta}}_p\right)^2\left(\theta_{t,p} - \overline{\theta}_p\right)^2}} \tag{20}$$

where θ_t and $\hat{\theta}_t$ are the true parameter and its estimated value at the *t*-th time step, respectively; $\bar{\theta}_p$ and $\bar{\hat{\theta}}_p$ are the mean value of the true parameters and its estimated values, respectively; and m is the length of the data during the whole period.

3. Synthetic experiment and real catchment case study

Two synthetic experiments and two real catchment case studies were designed to assess the performance of SSC-DP. The experiments are described in Table 3.

3.1 Synthetic experiments

The two synthetic experiments examine the ability of SSC-DP to identify the time-varying parameters of the TMWB and Xinanjiang hydrological models. The merit of synthetic experiments is that the parameters can be synthetically generated to be either constant or time varying. Hence, it is convenient to compare the estimated values with the a priori known parameters to evaluate different parameter estimation methods. Note that synthetic experiments have been successfully used in several time-varying parameter identification studies (Deng et al., 2016; Pathiraja et al., 2016; Xiong et al., 2019).

3.1.1 Synthetic experiment with the TMWB model

Synthetic data of monthly precipitation and potential evapotranspiration were collected from the 03451500 catchment of the Model Parameter Estimation Experiment (MOPEX) (Duan et al., 2006). The data cover 252 months. Runoff was derived by the

TMWB model using synthetic precipitation, potential evapotranspiration, and the known parameters. Gaussian noise was added to the simulated runoff to represent uncertainties. The mean of the noise was set to zero, and the standard deviation was assumed to be 3 % of the magnitude of the values (Deng et al., 2016).

Eight scenarios with different known parameters were investigated (Table 4). The first scenario considered constant parameters. Scenarios 2 and 3 considered month-by-month variations in TMWB model parameters, i.e., the parameters remain constant during each month, but change from month to month. Scenarios 4 and 5 considered parameters that change every six months. Scenarios 6–8 considered year-by-year variations. The changes in both *C* and *SC* were considered to be linear in scenarios 2, 4, and 6 (Trend) and sinusoidal in scenarios 3, 5 and 7 (periodicity), reflecting the impacts of climate change and human activities (Pathiraja et al., 2016). Scenario 8 considered a periodic variation with an increasing trend for parameter *C* and only the linear variation in *SC*.

3.1.2 Synthetic experiment with the Xinanjiang model

Hourly precipitation and pan evaporation data were collected from the Baiyunshan Reservoir basin in China. The data cover a period of 18000 h. The Xinanjiang model has 15 parameters, which can lead to a significant computational burden. To reduce the total number of model runs, only the sensitive parameters were considered to be free. The Morris method was used to detect the free parameters (Fig. 4), with the results showing that *KE*, *CI*, *CG*, *KI*, *KG*, and *NK* are sensitive parameters. Thus, the other

parameters were held constant for the whole period.

Similar to the experiment with the TMWB model, synthetic runoff was derived from the Xinanjiang model with added Gaussian noise. The mean of the noise was set to zero, and the standard deviation was assumed to be 5 % of the magnitude of the values. As presented in Table 5, all 15 parameters were set to be constant in the first scenario. The known sensitive parameters were considered to vary with a certain trend and periodicity in scenarios 2 and 3, respectively. Scenario 4 considered a combined variation of trend and periodicity for the parameter *KE*, with the other free parameters set to vary linearly. The parameter variations in scenarios 2–4 were assumed to occur once a month.

3.2 Study area: Wuding River basin

The Wuding River basin (Fig. 5(a)) examined in the first case study is a large sub-basin of the Yellow River basin located on the Loess Plateau (Xu, 2011). The Wuding River has a drainage area of 30261 km^2 and a total length of 491 km. The average slope is 0.2 %, and the elevation varies from 600–1800 m above sea level. The area is a semi-arid region with mean annual precipitation of ~401 mm. The annual potential evapotranspiration is 1077 mm, and the mean annual runoff is 39 mm. The data for this basin were collected over the period 1958–2000. The daily precipitation was obtained from Thiessen polygons using records from 122 rain gauges. Based on meteorological data from the China Meteorological Data Sharing Service System (http://data.cma.cn), areal pan evaporation data were obtained. As illustrated in Fig. 5(a), the station furthest

downstream, Baijiachuan, drains an area of 29,662 km^2 (98 % of the total basin) and records the daily runoff data. The data of the daily precipitation and streamflow in the Wuding River basin were obtained from the local Hydrology and Water Resources Bureau of China, the quality of which has been checked by the official authorities, and there are no gaps among these data for all the hydrological stations. It can be seen from Fig. 5(c) that the annual streamflow in the Wudinghe River basin has a distinct decreasing trend, while seasonal variations are not significant, but the annual precipitation and pan evaporation generally have no trend, suggesting the impacts of human activities on rainfall–runoff relationships.

Soil and water conservation measures, such as construction of the check dams and afforestation, have been undertaken since the 1960s. The areas of two soil and water conservation measures are plotted in Fig. 5(e), the data of which were collected from Zhang et al. (2002). The areas of tree planning have an increasing trend, but the slope gets much larger after 1972. It indicates that the greater efforts have been made for afforestation since the turning point. Similarly, the areas of dammed lands also increase, but the rate gets slower after 1972. These two soil and water conservation measures had changed the underlying surface of the watershed, and impacted the relationship between precipitation and runoff (Gao et al., 2017; Jiao et al., 2017).

3.3 Study area: Xun River basin

The proposed method was also applied to the Xun River basin, China (Fig. 5(b)). Located between 108°24’–109°26’ E and 32°52’–33°55’ N, the study area covers

394 approximately 6448 km^2. The Xun River is ~218 km long and has an average annual
395 flow of 73 m^3/s (Li et al., 2016). The basin has a subtropical monsoon climate. The
396 weather is wet and moderate with an annual average temperature of 15–17 °C. The daily
397 hydrological data from 1991–2001 include precipitation from 28 rainfall stations, pan
398 evaporation from three hydrological gauged stations, and discharge at the outlet of the
399 Xun River basin. Areal precipitation was obtained using the Thiessen polygon method,
400 and areal pan evaporation was computed using the average value of the data from
401 gauged stations. The data in the Xun River basin were also obtained from the local
402 Hydrology and Water Resources Bureau of China, and there are no gaps among these
403 data for all the hydrological stations.

404 It can be observed from Fig. 5(d) that no trend is found in annual precipitation,
405 pan evaporation and streamflow, suggesting that the relationship between precipitation
406 and runoff of the Xun River basin is rarely affected by human activities during 1991-
407 2001. However, there exhibit strong seasonal patterns in these three climatic and
408 hydrological variables, suggesting that seasonal variations in hydrological parameters
409 should be considered.

410 # 4. Results

411 ## 4.1 Synthetic experiment

412 ### 4.1.1 Results of synthetic experiment with the TMWB model

413 When using SSC-DP, the first task is to define how the hydrological data series
414 should be split into the k sub-periods within which the parameters are assumed to be

415 constant. As climate change can induce seasonal or half-annual variations while human
416 activities usually influence the watershed annually, lengths of three months, six months,
417 and 12 months were arbitrarily chosen. Thus, this experiment compared the following
418 four methods: (1) EnKF; (2) 3-SSC-DP; (3) 6-SSC-DP, and (4) 12-SSC-DP.

419 Figure 6(a) presents the runoff simulation performance for various scenarios. In
420 scenario 1, the NSE values of the three SSC-DP methods are all higher than that of
421 EnKF. The results of NSE_{ln} show no significant differences among various methods.
422 For scenarios 2, 4, and 6, where true parameters have linear trends, the 6-SSC-DP and
423 12-SSC-DP are superior to the EnKF and 3-SSC-DP in terms of NSE and NSE_{ln}. In
424 scenario3, where the true parameters have periodic variations and change every month,
425 the NSE and NSE_{ln} values of 6-SSC-DP and 12-SSC-DP decrease significantly,
426 because the assumed sub-period length is longer than the time-scale of actual variations.
427 Similarly, in scenario 5, 12-SSC-DP performs worst for NSE and NSE_{ln}, but 6-SSC-
428 DP performs best. In scenario 7 and 8, both 6-SSC-DP and 12-SSC-DP perform better
429 than EnKF. According to the evaluations of NSE and NSE_{ln}, the SSC-DP offers
430 improved accuracy than the EnKF if the proper length is chosen. Another advantage of
431 the SSC-DP is the small RE. For all scenarios, the SSC-DP methods significantly
432 outperform for RE compared with EnKF. Among the SSC-DP methods, the RE of 3-
433 SSC-DP is the smallest.

434 Figures 6 (b) and (c) focuses on the ability of the four methods to identify time-
435 varying parameters. It can be seen that the RMSE and MARE values of the 3-SSC-DP
436 are larger than those of other methods in most cases. That is because the sub-period

437 length that serves as a calibration period for MCMC is too short (i.e., three months) that
438 the estimated parameters are associated with higher uncertainties.
439 Regarding the synthetic true parameters are constant values (scenario 1), 12-SSC-
440 DP gives the best performance with the lowest RMSE, MARE and highest R^2. The
441 observations and estimated parameters are presented in Figure 7 (b). It shows that the
442 estimated parameters obtained by EnKF vary at every time step, resulting in larger
443 deviations from the observations than 6-SSC-DP and 12-SSC-DP.
444 When the synthetic true parameters vary linearly (scenarios 2, 4, and 6), 12-SSC-
445 DP produces best estimations in comparison with EnKF, 3-SSC-DP, and 6-SSC-DP.
446 The performances of 6-SSC-DP and EnKF are similar.
447 When the synthetic true parameters vary sinusoidally from month to month, EnKF
448 gives the best estimations in scenario 3. The poor performances of 6-SSC-DP and 12-
449 SSC-DP can be explained by the sub-period length being much longer than the actual
450 one. When the parameters vary periodically at six-month intervals (scenario 5), 6-SSC-
451 DP yields the best performance with the lowest RMSE, MARE and highest R^2. The
452 differences of estimation performances among 3-SSC-DP, 12-SSC-DP and EnKF are
453 small. The estimated parameters for scenario 5 have been plotted in Fig. 7(a). Although
454 3-SSC-DP and 12-SSC-DP have different lengths of sub-periods, they can also detect
455 the correct seasonal signal of the parameters. For the annual variation in parameters
456 (scenario 7), 12-SSC-DP and 6-SSC-DP produce better results than EnKF. Similar
457 results can be seen in scenario 8 where C has a combined variation from year to year.
458 In summary, the results indicate that the SSC-DP with a suitable length can estimate

459 more accurate parameters than EnKF.

460 **4.1.2 Results of synthetic experiment with the Xinanjiang model**

461 The Xinanjiang model is more complex than TMWB, and so some sensitivity
462 analysis is necessary. As stated in Sect. 3.1.2, the sensitive parameters are *KE*, *CI*, *CG*,
463 *KI*, *KG*, and *NK*. The 18000 hourly hydrological data points were divided into 25 sub-
464 periods (monthly time scale) and 12 sub-periods (bimonthly time scale). It is considered
465 that a monthly time scale helps diagnose seasonal variations, whereas a two-monthly
466 time scale provides data for longer calibration lengths.

467 Three data assimilation methods (see Sect. 2.3.2 for details) were applied to the
468 synthetic data: (1) EnKF; (2) EnKS, and (3) SSC-EnKF. The results in Fig. 8 indicate
469 that EnKS is superior to EnKF, as previously observed (Li et al., 2013), although SSC-
470 EnKF gives the best results. This is probably because SSC-EnKF is based on the
471 assumption that the parameters remain constant during each sub-period.

472 The simulated streamflow and identification of time-varying parameters was
473 compared across four methods: 1-SSC, SSC-EnKF, 1-SSC-DP, and 2-SSC-DP. The
474 simulation performance is summarized in Figure 9(a). For all scenarios, the NSE of 2-
475 SSC-DP is the lowest, but it performs better for low flows. The SSC-EnKF produces
476 the highest RE in scenarios 2, 3 and 4, indicating the problem of simulating water
477 balance. The SSC and 1-SSC-DP perform well for all scenarios in terms of NSE, RE
478 and NSE_{ln}. Wherein, the SSC performs better than the 1-SSC-DP with regard to RE,
479 while 1-SSC-DP is slightly superior to SSC in scenario 3 with higher NSE_{ln}.

480 Figures 9(b) and (c) compare the time-varying parameter estimation performance
481 among the four methods. In scenarios 1 and 2, 2-SSC-DP produces the lowest RMSE,
482 MARE and R^2, followed by the 1-SSC-DP. The 1-SSC-DP is slightly superior to the 1-
483 SSC and significantly outperforms the SSC-EnKF for the two scenarios.
484 When the synthetic true parameters vary sinusoidally from month to month
485 (scenario 3), the estimated parameters are plotted in Fig. 10. It can be seen that 1-SSC-
486 DP successfully detects seasonal signal in every parameter. The SSC-EnKF performs
487 well for R^2, but it has high MARE. Although the average MARE of the SSC and 2-
488 SSC-DP are lower than that of SSC-EnKF, the R^2 of them are relatively low. Therein,
489 from Fig. 10, the estimated parameters by the 1-SSC fluctuate generally periodically,
490 but the variations are dramatic, resulting in lowest R^2 for CI, KI, KG and NK. The
491 estimated parameters of the 2-SSC-DP fluctuate more slowly, but the sub-period length
492 is too long. In scenario 4, 1-SSC performs better than the SSC-EnKF and 2-SSC-DP,
493 but is still slightly inferior to the 1-SSC-DP. Overall, the 1-SSC-DP achieves higher-
494 quality and more robust parameter estimations performances than the other methods.

495 **4.2 Case study: Wuding River basin**

496 Figures 11(a) and (b) show the double mass curves between daily runoff and
497 precipitation for the Wuding River basin. Similar to the work of Deng et al. (2016), the
498 two linear slopes (p-value < 0.05) of the curves are different before and after 1972,
499 demonstrating the relationship between precipitation and runoff changes under the soil
500 and water conservation measures. This suggests that there are annual variations in the

501 watershed characteristics. Hence, the length of each sub-period was set to 12 months,
502 and the time-varying parameters were identified using 12-SSC-DP. Based on daily
503 Wuding data from 1958–2000, sensitivity analysis showed that nine parameters of the
504 Xinanjiang model are relatively sensitive: *WM*, *WUM*, *WLM*, *KE*, *IMP*, *KI*, *KG*, *N*, and
505 *NK*.

506 The simulation results given by 12-SSC-DP were benchmarked against those from
507 12-SSC, data assimilation, and the conventional method in which all Xinanjiang model
508 parameters remain constant. The simulation performance is presented in Figure 12. The
509 values of the NSEs are relatively low, because the streamflow in dry regions is difficult
510 to simulate. It can be seen that the 12-SSC-DP gives the best simulation results among
511 different methods with the highest NSE, NSE_{ln} and small RE. Although the 12-SSC
512 produces relatively high NSE, it performs worst simulations for low flows. The SSC-
513 EnKF has relatively high NSE_{ln}, but the RE of it is the largest. Overall, the 12-SSC-DP
514 significantly improves the simulation performance of the Xinanjiang model in the
515 Wuding River basin.

516 Although the objective function of 12-SSC-DP considers the trade-off between
517 simulation accuracy and parameter continuity, 12-SSC-DP gives a higher NSE value.
518 This may be because 12-SSC locates a local peak over one sub-period, resulting in
519 unreasonable model states for the beginning of the next sub-period, whereas 12-SSC-
520 DP uses dynamic programming to explore more reasonable parameter values and model
521 states. Figure 13 shows the quantile-quantile plots, from which it can be seen that if the
522 parameters are assumed to be constant, streamflow is highly underestimated. The

underestimation mainly derives from the deficiencies of the model structure. Methods 12-SSC and 12-SSC-DP reduce this underestimation by using time-varying parameters. Additionally, 12-SSC-DP is slightly inferior to 12-SSC in terms of peak flows, but is superior in terms of simulating streamflow lower than 100 m^3/s, which accounts for 80 % of the whole streamflow time series. It can be inferred that the 12-SSC-DP is more applicable to the simulation of streamflow in the Wuding River basin.

The estimated time-varying parameters estimated by 12-SSC-DP are plotted in Fig.14. The results show that *WM* remains constant before and after 1972, but *WUM* varies significantly over this period, indicating that the distribution of soil water capacity may change, i.e., *WUM* decreases but *WLM* increases. A Person correlation analysis is applied to investigate the relationship between the areas of tree planning and *WUM* as well as *WLM*. It is found that there is a significant negative correlation (Pearson correlation efficient ρ=-0.38, $P<0.05$) between the areas of tree planning and *WUM*. While *WLM* has a nonsignificant positive correlation (ρ=0.26, $P>0.05$) with the areas of tree planning. It can be inferred that less severe soil erosion occurred, because the upper layers became thinner while the lower layer, where vegetation roots dominate, became thicker (Jayawardena and Zhou, 2000). Additionally, *IMP* is significantly correlated with the areas of tree planning (ρ=-0.33, $P<0.05$). Except for afforestation, the areas of the dammed lands are significantly correlated with *WLM* (ρ=0.46, $P<0.05$), suggesting that the construction of the check dams also has influence on the soil water capacity of the Wuding river basin. Other parameters, *KE*, *KI*, *KG*, *N* and *NK* have little differences before and after 1972. The variations in *WLM* and *IMP* slowed down after

545 the turning point, similar to the results of Deng et al. (2016).

546 **4.3 Case study: Xun River basin**

547 Figures 11(c) and (d) show the double mass curves between runoff and
548 precipitation for the Xun River basin. The linear slope of the curve is generally
549 stationary for the whole ten-year period shown in Fig. 11(c), with a correlation
550 coefficient of 99.6 %. In contrast, the linear slope for an intra-annual timescale is non-
551 stationary (Fig. 11(d)). Based on these results, it can be inferred that the relationship
552 between precipitation and runoff is stable from 1990–2000, but varies over the intra-
553 annual timescale. Hence, sub-periods of three and 12 months were examined in the Xun
554 River basin using models 3-SSC-DP and 12-SSC-DP. From the Xun River basin data
555 from 1991–2000, sensitivity analysis suggested that five parameters of the Xinanjiang
556 model are relatively sensitive, namely *KE*, *B*, *KI*, *KG*, and *NK*.

557 Similar to the case study of the Wuding River basin, the simulation performance
558 of 3-SSC-DP was benchmarked against that of 3-SSC, data assimilation, and the
559 conventional calibration method. Among the data assimilation methods described in
560 Sect. 2.3.2, 3-SSC-EnKF gives the highest simulation accuracy. The simulation
561 performance is presented in Figure 15. All methods performed well, with NSE values
562 of 92.5 %, 93.0 %, 95.0 %, and 94.8 % for the conventional method, 3-SSC-EnKF, 3-
563 SSC, and 3-SSC-DP, respectively. 3-SSC and 3-SSC-DP also perform well for NSE_{ln}
564 compared with 3-SSC-EnKF and the conventional method. However, as regards to RE,
565 the values are 0.0007 and 0.0324 for 3-SSC-DP and 3-SSC-DP, respectively. It

indicated that the 3-SSC-DP can better simulate water balance than the 3-SSC in the Xun River basin. Figure 16 illustrates the hydrograph and quantile-quantile plots for the simulations in the Xun river basin. It is evident that the peak flows estimated by the 3-SSC are higher than those of 3-SSC-DP, and 3-SSC-DP simulate better the flows ranging from 100 m^3/s to 200 m^3/s.

The estimated parameters using 3-SSC-DP are presented in Fig. 17(a). Some parameters vary significantly over an intra-annual time scale. Among them, the parameter *KE*, representing the ratio of potential evapotranspiration to pan evaporation, exhibits the most distinct seasonal variations. A fast Fourier transform was used to calculate the spectral power of the *KE* time series to explore its periodic characteristics. As can be observed from Fig. 17(b), 3-SSC-DP had the greatest spectral power, for a period of 4.0 cycles per year, somewhat higher than the power obtained by 3-SSC and 3-SSC-EnKF. This means a stronger periodic pattern is captured by 12-SSC-DP. Given that the estimated *KE* varies at three-monthly intervals, it has a one-year periodicity. The other parameters do not exhibit significant one-year periodic patterns. This may be because only *KE*, linking potential evapotranspiration and pan evaporation, is directly impacted by seasonal climate variations, such as temperature.

5. Discussion

As noted in the methodology section, the performance of the proposed method is influenced by several factors, such as the weights in the objective function and the choice of lengths. Some suggestions regarding the improvement of the proposed approach are now discussed in detail.

588 **5.1 Objective function of dynamic programming in SSC-DP**

589 In the conventional method, a parameter set is identified as optimal for providing
590 the best simulation over the calibration period. However, other parameter sets with
591 slightly worse (but still good) performance can also be candidates. Allowing for input
592 data uncertainty and local optima, SSC-DP identifies parameter sets that perform near-
593 optimally and display less fluctuations over sub-periods. This can be adjusted by
594 weights in the objective function of the dynamic programming approach (see Eq. (3)).
595 As the weighting for accuracy increases, parameters providing more accurate
596 simulations are chosen, but parameter continuity is less important. If too much
597 importance is given to continuity, the variations in real world processes may be
598 underestimated. Here, the influence of different weights has been assessed for
599 simulation accuracy and parameter continuity based on synthetic experiments with the
600 TMWB and Xinanjiang models, respectively. Specifically, the weight for simulation
601 accuracy was set to 1, and the weight for parameter continuity α varied from zero to a
602 small positive value (e.g., 1). When $\alpha = 0$, only simulation accuracy was considered.

603 Figure 18(a) shows the R^2 value of 12-SSC-DP with various continuity weights for
604 scenario 4 in the synthetic experiment with the TMWB model. It can be seen that R^2 is
605 lowest when $\alpha = 0$ for both *C* and *SC*. There is some improvement when a nonzero
606 weight is applied. As α increases, the performance of 12-SSC-DP improves, and then
607 worsens; the differences among schemes with nonzero weights are not distinct. Similar
608 results can be observed in Fig. 18(b), which presents the R^2 value of 12-SSC-DP with
609 various α for scenario 2 in the synthetic experiment with the Xinanjiang model.

610 Therefore, nonzero continuity weights can significantly improve the parameter
611 estimation performance compared with the zero-weight case. It is suggested that
612 weights of 1 (accuracy) and 0.005 (continuity) be used with the TMWB model and
613 weights of 1 (accuracy) and 0.2 (continuity) be applied with the Xinanjiang model, as
614 in this study.

615 **5.2 Choice of sub-period length in SSC-DP**

616 As mentioned by Gharari et al. (2013), there are different ways of determining the
617 sub-period lengths. The sub-periods can be non-continuous hydrological years (Seiller
618 et al., 2012), months or seasons (Deng et al., 2018; Paik et al., 2005), and discharge or
619 precipitation events (Singh and Bardossy, 2012). This introduces a controversial issue
620 whereby parameters are impacted by the length of the calibration period. Merz et al.
621 (2009) suggested that 3–5 years is an acceptable calibration period, whereas Singh and
622 Bardossy (2012) indicated that a small number of events may be sufficient for
623 parameter identification. It is suggested that the determination of the sub-period length
624 considers three factors:

625 (1) The temporal scale of climate change or human activities. For example, the
626 Wudinghe River basin is taken as a case study. The soil and water conservation
627 measures have led to a durative and long-term change in the catchment characteristic
628 since 1960s. Due to this, the yearly sub-period is preferred.

629 (2) The seasonality. Contrary to the Wudinghe River basin, the relationship
630 between precipitation and runoff of the Xun River basin is rarely affected by human

activities during 1991-2001. However, its significant seasonal dynamics can be observed and has been studied in literature (Lan et al., 2020; Lan et al., 2018). In order to diagnose the seasonality, the stable period of 3-month is adopted.

(3) The simulation accuracy. The length should be neither too long nor too short so as to increase the reliability of the calibration while guaranteeing that variations in real processes are captured. Thus, given that the time scale of the variations is unknown, the proposed SSC-DP can be used with different split-sample lengths. It is suggested that the length should be as long as possible without degrading the simulation performance significantly. For example, in the synthetic experiment with the TMWB model, if the difference between the NSE values of 6-SSC-DP and 3-SSC-DP is small, the preferred length is 6-month.

However, many studies are based on the conventional assumption that parameters of different sub-periods are independent. Hence, the sub-period lengths should be long enough to reduce the degree of uncertainty. In this study, the assumption of parameter continuity is introduced to give another constraint that considers correlations between parameters of adjacent sub-periods. It appears that the determination of sub-period lengths deserves further investigation.

6. Conclusions

This paper has described a time-varying parameter estimation approach based on dynamic programming. The proposed SSC-DP combines the basic concept of SSC and the continuity assumption of data assimilation to estimate more continuous parameters while providing comparably good streamflow simulations. Two synthetic experiments

were designed to evaluate its applicability and efficiency for time-varying parameter identification. Furthermore, two case studies were conducted to explore the advantages of SSC-DP in real catchments. From the results, the following conclusions can be drawn:

1. The proposed method with a suitable length not only produces better simulation performance, but also ensures more accurate parameter estimates than SSC and EnKF in the synthetic experiment using the TMWB model with two parameters. The impact of sub-period lengths on the performance of SSC-DP is significant when the known parameters vary sinusoidally.

2. The proposed method can be used to deal with complex hydrological models involving a large number of parameters, demonstrated by the synthetic experiment using the Xinanjiang model with 15 parameters. A sensitivity analysis was performed to reduce the probable computational cost and improve the efficiency of identifying the time-varying parameters.

3. The proposed method has the potential to detect the relationship between the time-varying parameters and dynamic catchment characteristics. For example, SSC-DP produces the best simulation performance in the case study of the Wuding River basin and detects that parameters representing soil water capacity and impervious areas changed significantly after 1972, reflecting the soil and water conservation projects carried out from 1958–2000. Additionally, SSC-DP detects the strongest seasonal signal in the case study of Xun River basin, indicating the distinct impacts of seasonal climate variability.

This study has demonstrated that the proposed method is an effective approach for

identifying time-varying parameters under changing environments. Further work is still needed, such as to determine an objective method for choosing the sub-period lengths.

Acknowledgements

This study was supported by the Natural Science Foundation of Hubei Province (2017CFA015), the National Natural Science Foundation of China (51861125102), and Innovation Team in Key Field of the Ministry of Science and Technology (2018RA4014). The authors would like to thank the editor and anonymous reviewers for their comments that helped improve the quality of the paper.

Code/Data availability

The data and codes that support the findings of this study are available from the corresponding author upon request.

Author contribution

All of the authors helped to develop the method, designed the experiments, analyzed the results and wrote the paper.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Alvisi, S., Mascellani, G., Franchini, M., Bardossy, A., 2006. Water level forecasting through fuzzy logic and artificial neural network approaches. Hydrology and Earth System Sciences 10(1), 1-17.
Bellman, R., 1957. Dynamic programming. Princeton University Press, Princeton.
Broderick, C., Matthews, T., Wilby, R.L., Bastola, S., Murphy, C., 2016. Transferability of hydrological models and ensemble averaging methods between contrasting climatic periods. Water Resources Research 52(10), 8343-8373.
Bronstert, A., 2004. Rainfall-runoff modelling for assessing impacts of climate and land-use change. Hydrological Processes 18(3), 567-570.
Chen, Y., Zhang, D., 2006. Data assimilation for transient flow in geologic formations via ensemble kalman filter. Advances in Water Resources 29(8), 1107-1122.
Chib, S., Greenberg, E., 1995. Understanding the metropolis-hastings algorithm. American Statistician 49(4), 327-335.
Coron, L. et al., 2012. Crash testing hydrological models in contrasted climate conditions: An experiment on 216 australian catchments. Water Resources Research 48.
Dai, C., Qin, X.S., Chen, Y., Guo, H.C., 2018. Dealing with equality and benefit for water allocation in a lake watershed: A gini-coefficient based stochastic optimization approach. Journal of Hydrology 561, 322-334.
Deng, C., Liu, P., Guo, S., Li, Z., Wang, D., 2016. Identification of hydrological model parameter variation using ensemble kalman filter. Hydrology and Earth System Sciences 20(12), 4949-4961.
Deng, C., Liu, P., Wang, D., Wang, W., 2018. Temporal variation and scaling of parameters for a monthly hydrologic model. Journal of Hydrology 558, 290-300.
Deng, C., Liu, P., Wang, W., Shao, Q., Wang, D., 2019. Modelling time-variant parameters of a two-parameter monthly water balance model. Journal of Hydrology 573, 918-936.
Duan, Q. et al., 2006. Model parameter estimation experiment (mopex): An overview of science strategy and major results from the second and third workshops. Journal of Hydrology 320(1-2), 3-17.
Feng, M. et al., 2017. Deriving adaptive operating rules of hydropower reservoirs using time-varying parameters generated by the enkf. Water Resources Research 53(8), 6885-6907.
Fowler, K., Peel, M., Western, A., Zhang, L., 2018. Improved rainfall-runoff calibration for drying climate: Choice of objective function. Water Resources Research 54(5), 3392-3408.
Fowler, K.J.A., Peel, M.C., Western, A.W., Zhang, L., Peterson, T.J., 2016. Simulating runoff under changing climatic conditions: Revisiting an apparent deficiency of conceptual rainfall-runoff models. Water Resources Research 52(3), 1820-1846.
Gao, S. et al., 2017. Derivation of low flow frequency distributions under human activities and its implications. Journal of Hydrology 549, 294-300.
Gharari, S., Hrachowitz, M., Fenicia, F., Savenije, H.H.G., 2013. An approach to identify time consistent model parameters: Sub-period calibration. Hydrology and Earth System Sciences 17(1), 149-161.
Guo, S.L., Wang, J.X., Xiong, L.H., Ying, A.W., Li, D.F., 2002. A macro-scale and semi-distributed monthly water balance model to predict climate change impacts in china. Journal of Hydrology 268(1-4), 1-15.
Guzha, A.C., Rufino, M.C., Okoth, S., Jacobs, S., Nobrega, R.L.B., 2018. Impacts of land use and land cover change on surface runoff, discharge and low flows: Evidence from east africa. Journal of Hydrology-Regional Studies 15, 49-67.
Hock, R., 1999. A distributed temperature-index ice- and snowmelt model including potential direct

solar radiation. Journal of Glaciology 45(149), 101-111.
Hughes, D.A., 2015. Simulating temporal variability in catchment response using a monthly rainfall-runoff model. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques 60(7-8), 1286-1298.
Hundecha, Y., Bardossy, A., 2004. Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. Journal of Hydrology 292(1-4), 281-295.
Jayawardena, A.W., Zhou, M.C., 2000. A modified spatial soil moisture storage capacity distribution curve for the xinanjiang model. Journal of Hydrology 227(1-4), 93-113.
Jeremiah, E., Marshall, L., Sisson, S.A., Sharma, A., 2013. Specifying a hierarchical mixture of experts for hydrologic modeling: Gating function variable selection. Water Resources Research 49(5), 2926-2939.
Jiao, Y. et al., 2017. Impact of vegetation dynamics on hydrological processes in a semi-arid basin by using a land surface-hydrology coupled model. Journal of Hydrology 551, 116-131.
Jie, M.X. et al., 2018. Transferability of conceptual hydrological models across temporal resolutions: Approach and application. Water Resources Management 32(4), 1367-1381.
Khalil, M., Panu, U.S., Lennox, W.C., 2001. Groups and neural networks based streamflow data infilling procedures. Journal of Hydrology 241(3-4), 153-176.
Kim, S., Hong, S.J., Kang, N., Noh, H.S., Kim, H.S., 2016. A comparative study on a simple two-parameter monthly water balance model and the kajiyama formula for monthly runoff estimation. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques 61(7), 1244-1252.
Kim, S.M., Benham, B.L., Brannan, K.M., Zeckoski, R.W., Doherty, J., 2007. Comparison of hydrologic calibration of hspf using automatic and manual methods. Water Resources Research 43(1).
Kim, S.S.H., Hughes, J.D., Chen, J., Dutta, D., Vaze, J., 2015. Determining probability distributions of parameter performances for time-series model calibration: A river system trial. Journal of Hydrology 530, 361-371.
King, D.M., Perera, B.J.C., 2013. Morris method of sensitivity analysis applied to assess the importance of input variables on urban water supply yield - a case study. Journal of Hydrology 477, 17-32.
Klemes, V., 1986. Operational testing of hydrological simulation-models. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques 31(1), 13-24.
Lan, T., Lin, K., Xu, C.-Y., Tan, X., Chen, X., 2020. Dynamics of hydrological-model parameters: Mechanisms, problems and solutions. Hydrology and Earth System Sciences 24(3), 1347-1366.
Lan, T. et al., 2018. A clustering preprocessing framework for the subannual calibration of a hydrological model considering climate-land surface variations. Water Resources Research 54(0).
Li, H. et al., 2018. Hybrid two-stage stochastic methods using scenario-based forecasts for reservoir refill operations. Journal of Water Resources Planning and Management 144(12).
Li, H., Zhang, Y., 2017. Regionalising' rainfall-runoff modelling for predicting daily runoff: Comparing gridded spatial proximity and gridded integrated similarity approaches against their lumped counterparts. Journal of Hydrology 550, 279-293.
Li, Y., Ryu, D., Western, A.W., Wang, Q.J., 2013. Assimilation of stream discharge for flood forecasting: The benefits of accounting for routing time lags. Water Resources Research 49(4), 1887-1900.
Li, Z. et al., 2016. Evaluation of estimation of distribution algorithm to calibrate computationally intensive hydrologic model. Journal of Hydrologic Engineering 21(6).
Lin, K. et al., 2014. Xinanjiang model combined with curve number to simulate the effect of land use

change on environmental flow. Journal of Hydrology 519, 3142-3152.
Lu, H. et al., 2013. The streamflow estimation using the xinanjiang rainfall runoff model and dual state-parameter estimation method. Journal of Hydrology 480, 102-114.
Luo, M., Pan, C., Zhan, C., 2019. Diagnosis of change in structural characteristics of streamflow series based on selection of complexity measurement methods: Fenhe river basin, china. Journal of Hydrologic Engineering 24(2).
Meng, S., Xie, X., Liang, S., 2017. Assimilation of soil moisture and streamflow observations to improve flood forecasting with considering runoff routing lags. Journal of Hydrology 550, 568-579.
Merz, R., Parajka, J., Bloeschl, G., 2009. Scale effects in conceptual hydrological modeling. Water Resources Research 45.
Merz, R., Parajka, J., Bloeschl, G., 2011. Time stability of catchment model parameters: Implications for climate impact analyses. Water resources research 47(W02531).
Ming, B., Liu, P., Bai, T., Tang, R., Feng, M., 2017. Improving optimization efficiency for reservoir operation using a search space reduction method. Water Resources Management 31(4), 1173-1190.
Moradkhani, H., Sorooshian, S., Gupta, H.V., Houser, P.R., 2005. Dual state-parameter estimation of hydrological models using ensemble kalman filter. Advances in Water Resources 28(2), 135-147.
Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics 33(2), 161-174.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part i — a discussion of principles. Journal of Hydrology 10(3), 282-290.
Paik, K., Kim, J.H., Kim, H.S., Lee, D.R., 2005. A conceptual rainfall-runoff model considering seasonal variation. Hydrological Processes 19(19), 3837-3850.
Pappenberger, F., Beven, K.J., Ratto, M., Matgen, P., 2008. Multi-method global sensitivity analysis of flood inundation models. Advances in Water Resources 31(1), 1-14.
Pathiraja, S. et al., 2018. Time-varying parameter models for catchments with land use change: The importance of model structure. Hydrology and Earth System Sciences 22(5), 2903-2919.
Pathiraja, S., Marshall, L., Sharma, A., Moradkhani, H., 2016. Hydrologic modeling in dynamic catchments: A data assimilation approach. Water resources research 52(5), 3350-3372.
Poulin, A., Brissette, F., Leconte, R., Arsenault, R., Malo, J.-S., 2011. Uncertainty of hydrological modelling in climate change impact studies in a canadian, snow-dominated river basin. Journal of Hydrology 409(3-4), 626-636.
Quoc Quan, T., De Niel, J., Willems, P., 2018. Spatially distributed conceptual hydrological model building: A genetic top-down approach starting from lumped models. Water Resources Research 54(10), 8064-8085.
Rebolho, C., Andreassian, V., Le Moine, N., 2018. Inundation mapping based on reach-scale effective geometry. Hydrology and Earth System Sciences 22(11), 5967-5985.
Refsgaard, J.C., Knudsen, J., 1996. Operational validation and intercomparison of different types of hydrological models. Water Resources Research 32(7), 2189-2202.
Seiller, G., Anctil, F., Perrin, C., 2012. Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions. Hydrology and Earth System Sciences 16(4), 1171-1189.
Si, W., Bao, W., Gupta, H.V., 2015. Updating real-time flood forecasts via the dynamic system response curve method. Water Resources Research 51(7), 5128-5144.

828 Singh, S.K., Bardossy, A., 2012. Calibration of hydrological models on hydrologically unusual events.
829 Advances in Water Resources 38, 81-91.
830 Siriwardena, L., Finlayson, B.L., McMahon, T.A., 2006. The impact of land use change on catchment
831 hydrology in large catchments: The comet river, central queensland, australia. Journal of
832 Hydrology 326(1-4), 199-214.
833 Sobol, I.M., 1993. Sensitivity estimates for nonlinear mathematical models. Mathematical modelling
834 and computational experiments 1(4), 407-414.
835 Stephens, C.M., Marshall, L.A., Johnson, F.M., 2019. Investigating strategies to improve hydrologic
836 model performance in a changing climate. Journal of Hydrology 579.
837 Sun, Y. et al., 2018. Development of multivariable dynamic system response curve method for real-time
838 flood forecasting correction. Water Resources Research 54(7), 4730-4749.
839 Teweldebrhan, A.T., Burkhart, J.F., Schuler, T.V., 2018. Parameter uncertainty analysis for an operational
840 hydrological model using residual-based and limits of acceptability approaches. Hydrology and
841 Earth System Sciences 22(9), 5021-5039.
842 Thirel, G. et al., 2015. Hydrology under change: An evaluation protocol to investigate how hydrological
843 models deal with changing catchments. Hydrological Sciences Journal-Journal Des Sciences
844 Hydrologiques 60(7-8), 1184-1199.
845 Toth, E., Brath, A., 2007. Multistep ahead streamflow forecasting: Role of calibration data in conceptual
846 and neural network modeling. Water Resources Research 43(11).
847 Westra, S., Thyer, M., Leonard, M., Kavetski, D., Lambert, M., 2014. A strategy for diagnosing and
848 interpreting hydrological model nonstationarity. Water Resources Research 50(6), 5090-5113.
849 Xie, S. et al., 2018. A progressive segmented optimization algorithm for calibrating time-variant
850 parameters of the snowmelt runoff model (srm). Journal of Hydrology 566, 470-483.
851 Xiong, L.H., Guo, S.L., 1999. A two-parameter monthly water balance model and its application. Journal
852 of Hydrology 216(1-2), 111-123.
853 Xiong, M. et al., 2019. Identifying time-varying hydrological model parameters to improve simulation
854 efficiency by the ensemble kalman filter: A joint assimilation of streamflow and actual
855 evapotranspiration. Journal of Hydrology 568, 758-768.
856 Xu, J., 2011. Variation in annual runoff of the wudinghe river as influenced by climate change and human
857 activity. Quaternary International 244(2), 230-237.
858 Yang, N. et al., 2017. Evaluation of the trmm multisatellite precipitation analysis and its applicability in
859 supporting reservoir operation and water resources management in hanjiang basin, china.
860 Journal of Hydrology 549, 313-325.
861 Yang, X. et al., 2018. A new fully distributed model of nitrate transport and removal at catchment scale.
862 Water Resources Research 54(8), 5856-5877.
863 Yin, J. et al., 2018. A copula-based analysis of projected climate changes to bivariate flood quantiles.
864 Journal of Hydrology 566, 23-42.
865 Zhang, J., Ji, W., Feng, X., 2002. Water and sediment changes in the wudinghe river: Present state,
866 formative cause and tendency in the future. A study of water and sediment changes in the
867 Yellow River 2, 393-429.
868 Zhao, R.J., 1992. The xinanjiang model applied in china. Journal of Hydrology 135(1-4), 371-381.
869

Table 1 Parameters of the TMWB model

Parameter	Physical meaning	Range and units
C	Evapotranspiration parameter	0.2-2.0 (–)
SC	Catchment water storage capacity	100-2000 (mm)

Table 2 Parameters of the Xinanjiang model

Category	Parameter	Physical meaning	Range and units
Evapotranspiration	WM	Tension water capacity	80-400 (mm)
	X	WUM=X×WM, WUM is the tension water capacity of lower layer	0.01-0.8 (–)
	Y	WLM=Y×WM, WLM is the tension water capacity of deeper layer	0.01-0.8 (–)
	K	Ratio of potential evapotranspiration to pan evaporation	0.4-1.5 (–)
	C	The coefficient of deep evapotranspiration	0.01-0.4 (–)
Runoff production	B	The exponent of the tension water capacity curve	0.1-10 (–)
	IMP	The ratio of the impervious to the total area of the basin	0.01-0.15 (–)
Runoff separation	SM	The areal mean of the free water capacity of the surface soil layer	10-80 (mm)
	EX	The exponent of the free water capacity curve	0.6-6 (–)
	CG	The outflow coefficients of the free water storage to groundwater	0.01-0.45 (–)
	CI	The outflow coefficients of the free water storage to interflow	0.01-0.45 (–)
Flow concentration	N	Number of reservoirs in the instantaneous unit hydrograph	0.5-10 (–)
	NK	Common storage coefficient in the instantaneous unit hydrograph	1-20 (–)
	KG	The recession constant of groundwater storage	0.6-1 (–)
	KI	The recession constant of the lower interflow storage	0.9-1 (–)

Table 3 Different cases of synthetic experiments and real catchment case studies for comparison and evaluation

	Data	Hydrological model	Time-varying parameter estimation methods		
			SSC	SSC-DP	Data assimilation
Synthetic experiment	Monthly synthetic data	TMWB model		✓	✓
	Hourly synthetic data	Xinanjiang model	✓	✓	✓
Real catchment case study	Daily data from Wuding River basin	Xinanjiang model	✓	✓	✓
	Daily data from Xun River basin	Xinanjiang model	✓	✓	✓

Table 4 True parameters of different scenarios in the synthetic experiment with the TMWB model

Scenario	Description
1	Both *C* and *SC* are constant
2	Both *C* and *SC* have increasing linear trends and change every month
3	Both *C* and *SC* have periodic variations and change every month
4	Both *C* and *SC* have increasing linear trends and change every six months
5	Both *C* and *SC* have periodic variations and change every six months
6	Both *C* and *SC* have increasing linear trends and change every year
7	Both *C* and *SC* have periodic variations and change every year
8	*C* has a periodic variation with an increasing linear trend, whereas *SC* only has an increasing linear trend. The parameters change every year

Table 5 True parameters of different scenarios in the synthetic experiment with the Xinanjiang model

Scenario	Description
1	*KE*, *CI*, *CG*, *KI*, *KG*, and *NK* remain constant
2	*KE*, *CI*, *CG*, *KI*, *KG*, and *NK* have linear trends and change every month
3	*KE*, *CI*, *CG*, *KI*, *KG*, and *NK* have periodic variations and change every month
4	*KE* has a periodic variation with an increasing linear trend, whereas *CI*, *CG*, *KI*, *KG*, and *NK* only have periodic variations. The parameters change every month

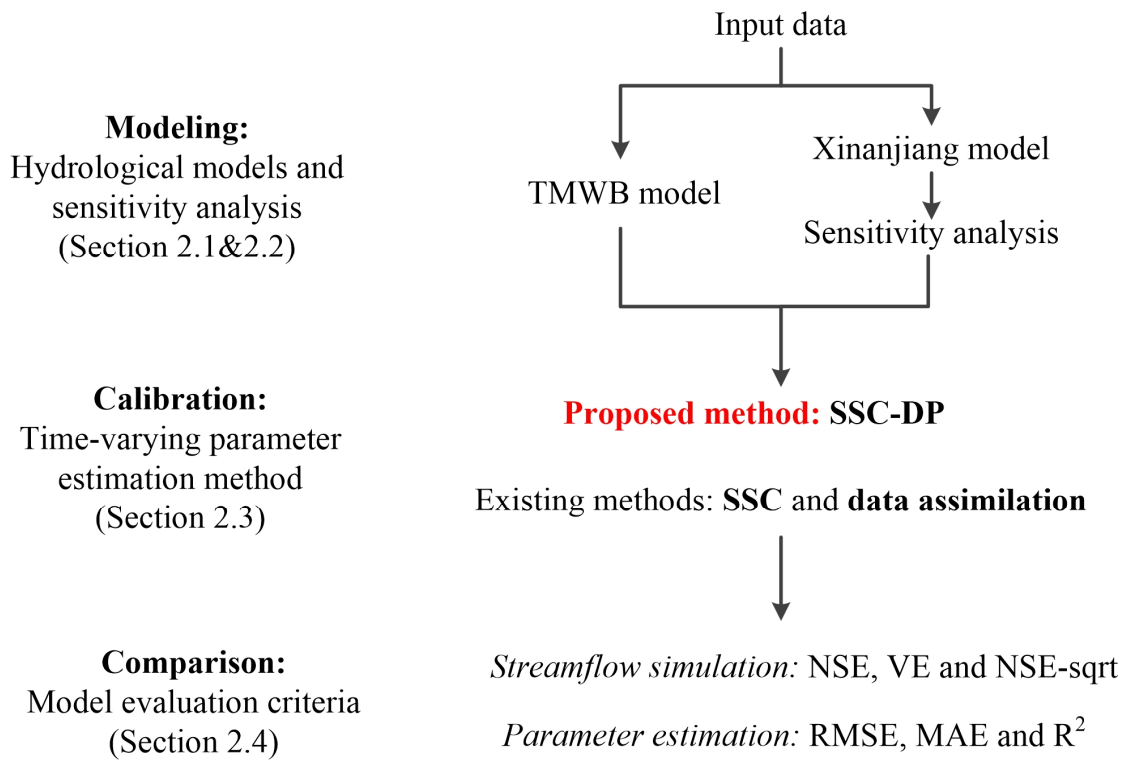


Figure 1 Flowchart of the methodologies

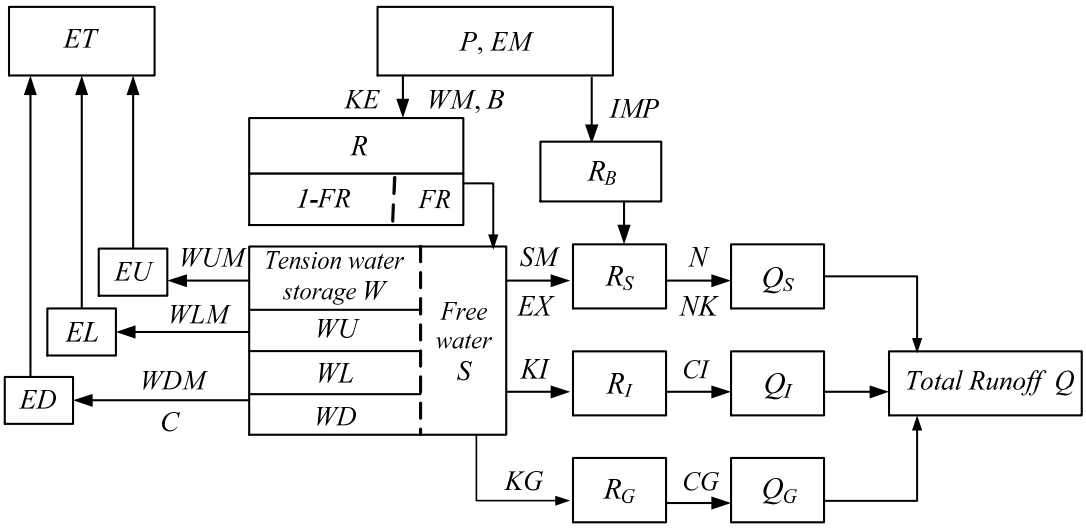


Figure 2 Flowchart of the Xinanjiang model.

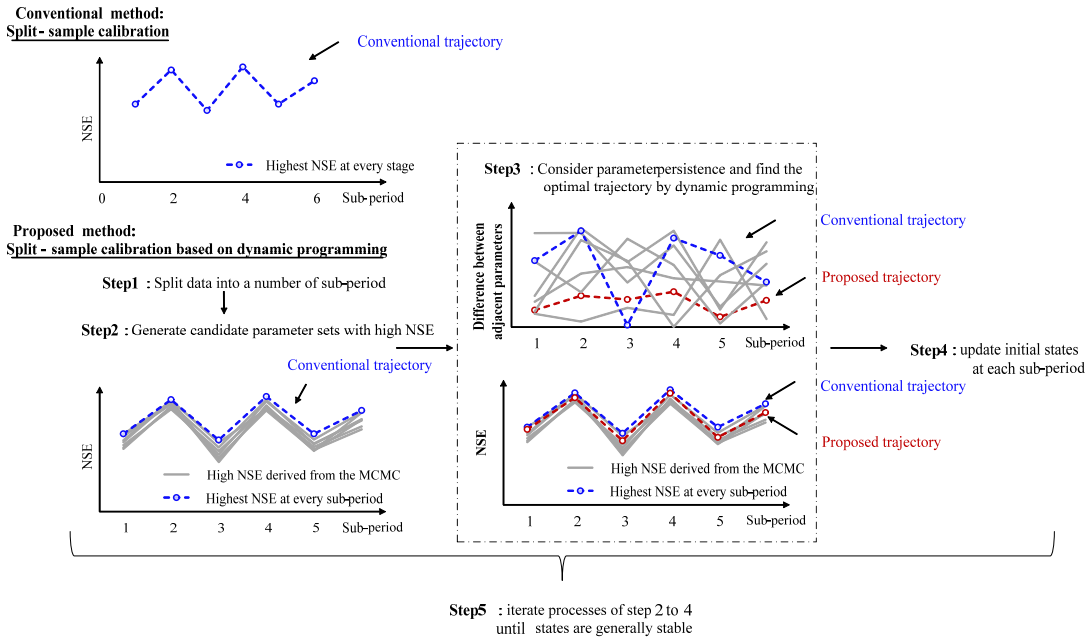


Figure 3 Flowchart of SSC-DP.

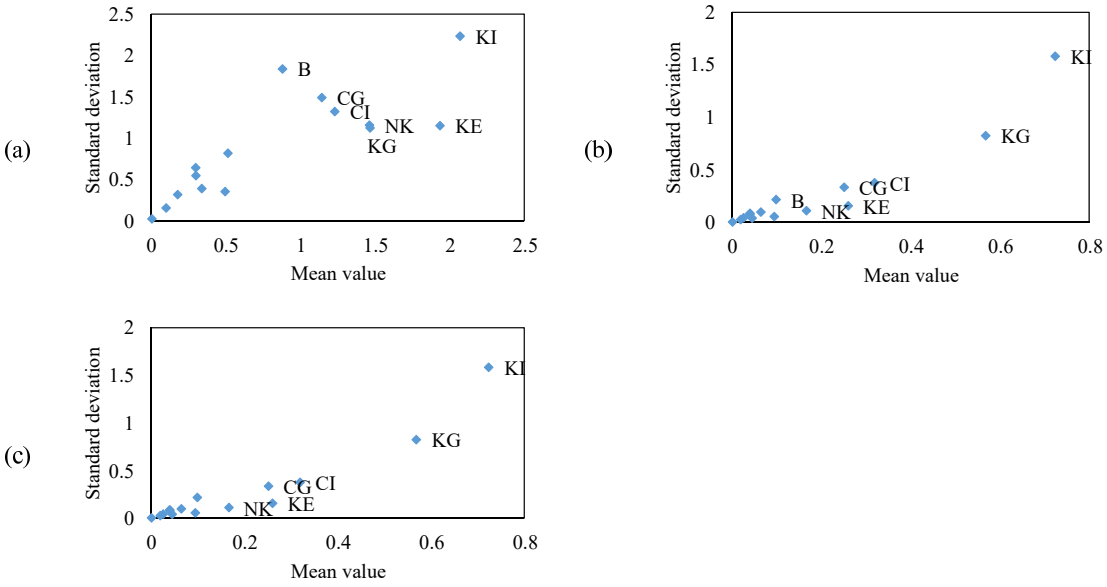


Figure 4 Results of the Morris method for the synthetic experiment with the Xinanjiang model. The sensitivity analysis is based on three different kinds of model responses: (a) NSE; (b) NSE_{abs}; (c) NSE_{ln}. Only the most sensitive parameters are labeled.

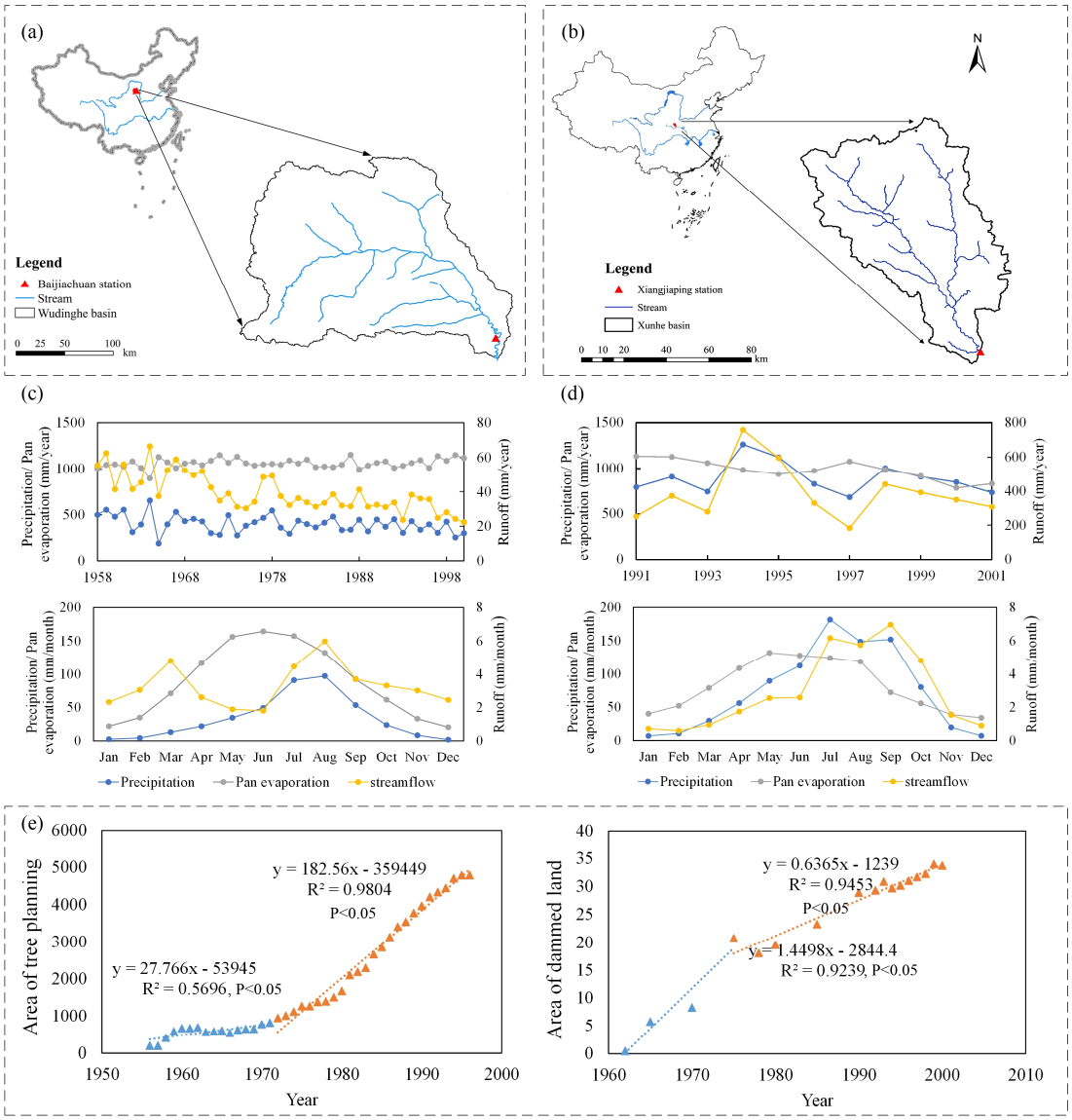


Figure 5 Location of (a) Wuding River basin and (b) Xun River basin. The plots (c) and (d) show the average yearly and monthly variations of precipitation, pan evaporation and streamflow in the Wuding River basin and Xun River basin, respectively. The plot (e) shows the temporal variations in the soil and water conservation measures.

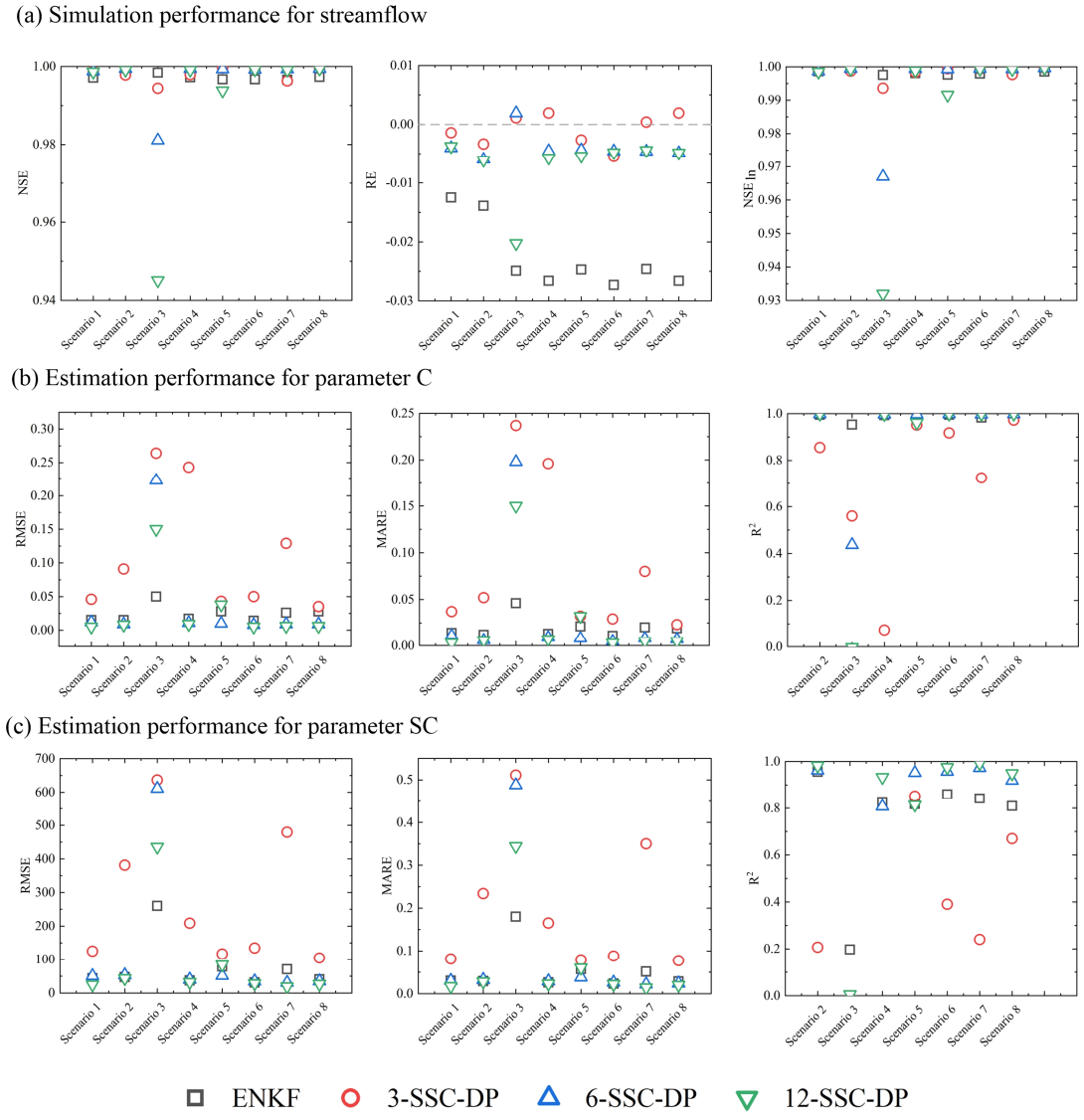


Figure 6 Comparison between the EnKF and SSC-DP methods for (a) streamflow simulation and identification of (b) parameter C and (c) parameter SC.

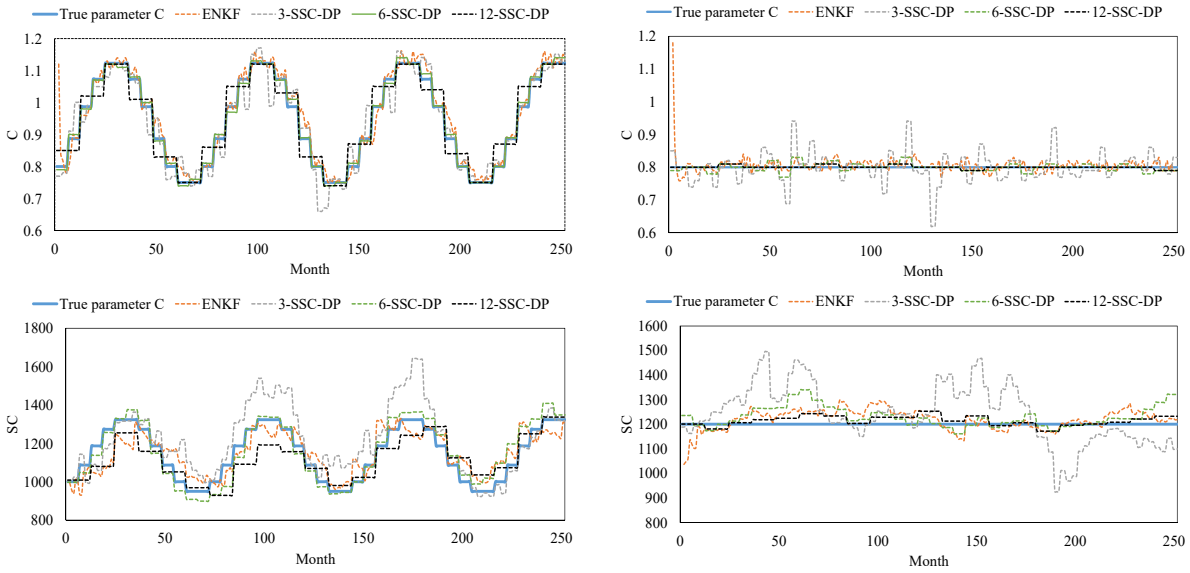


Figure 7 Comparison among different methods for (a) scenario 5 and (b) scenario 1 of the synthetic experiment with the TMWB model.

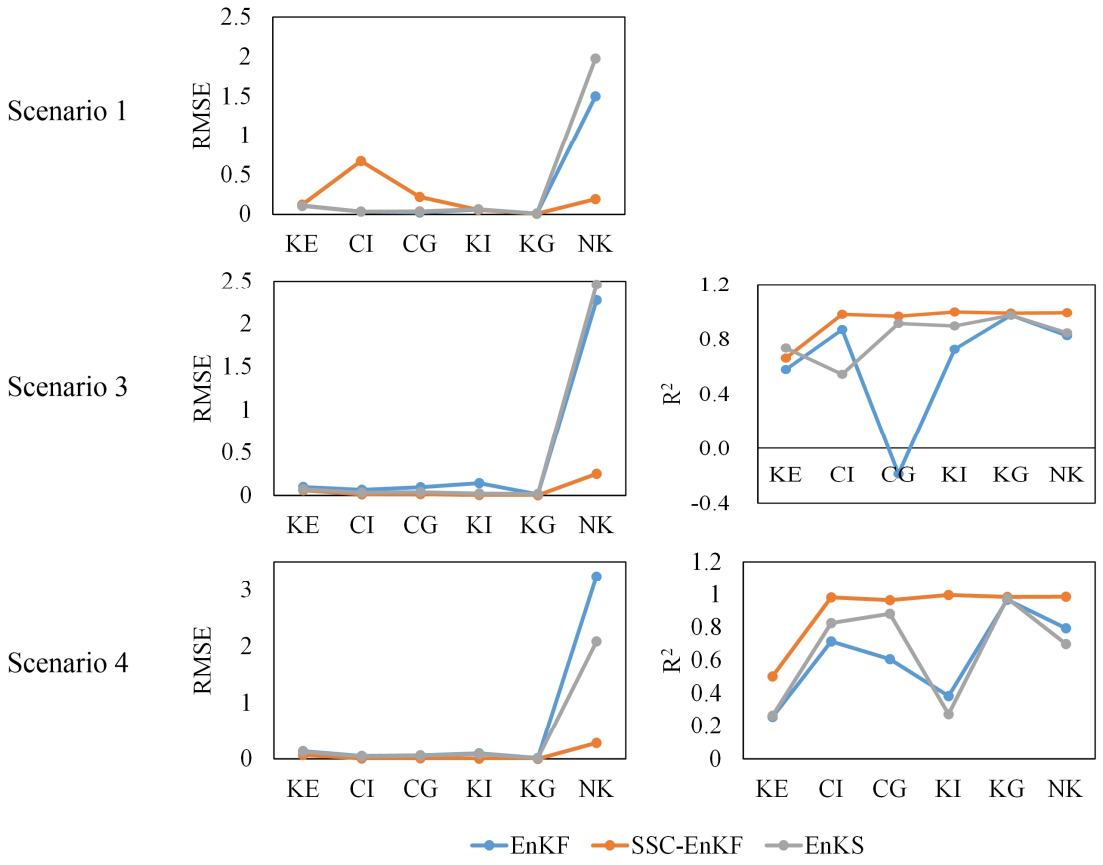


Figure 8 Comparison among EnKF, SSC-EnKF, and EnKS in the synthetic experiment with the Xinanjiang model.

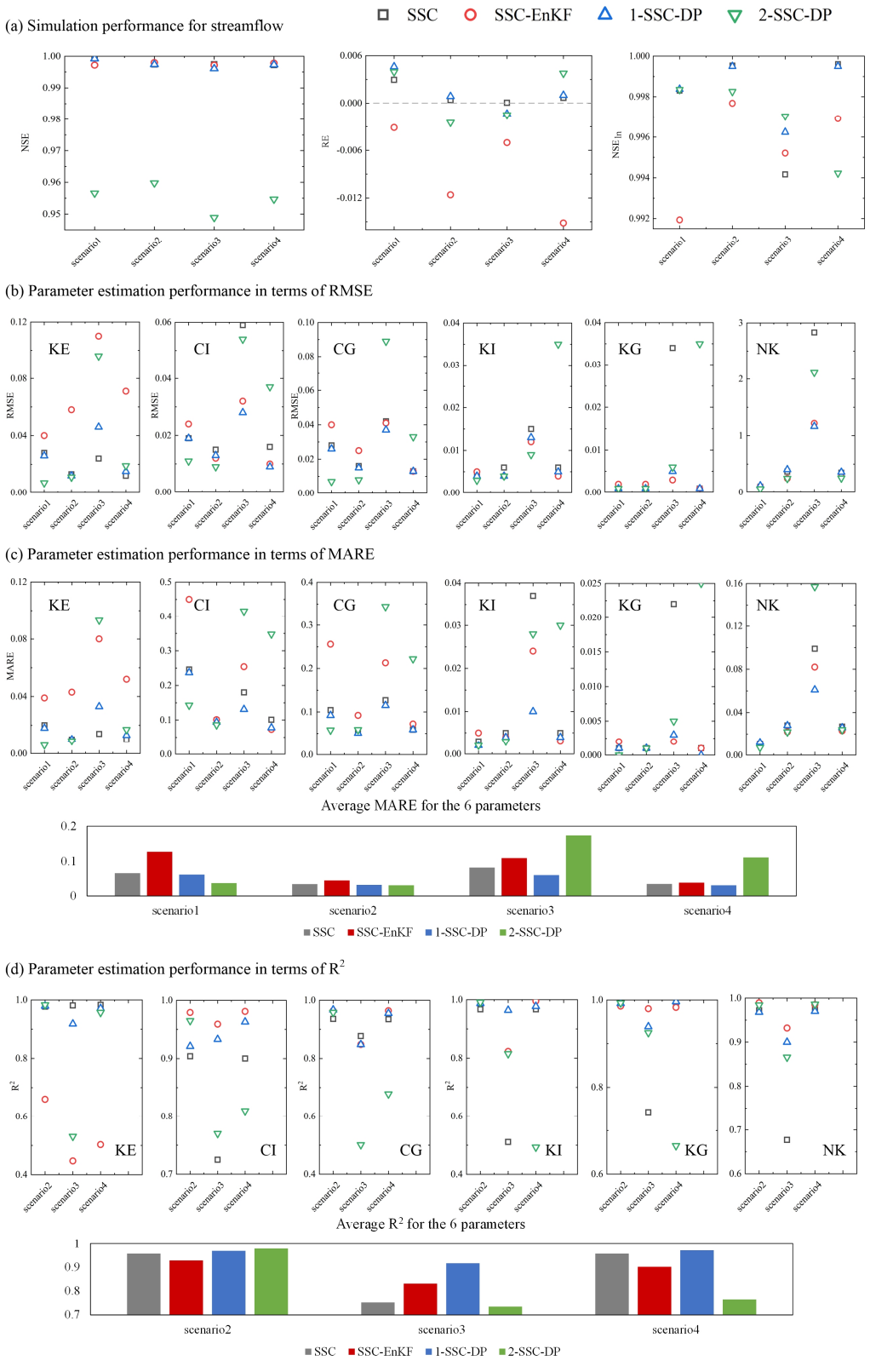


Figure 9 Comparison among the SSC, SSC-EnKF and SSC-DP methods for (a) streamflow simulation and parameter identification in terms of (b) RMSE, (c) MARE and (d) R^2.

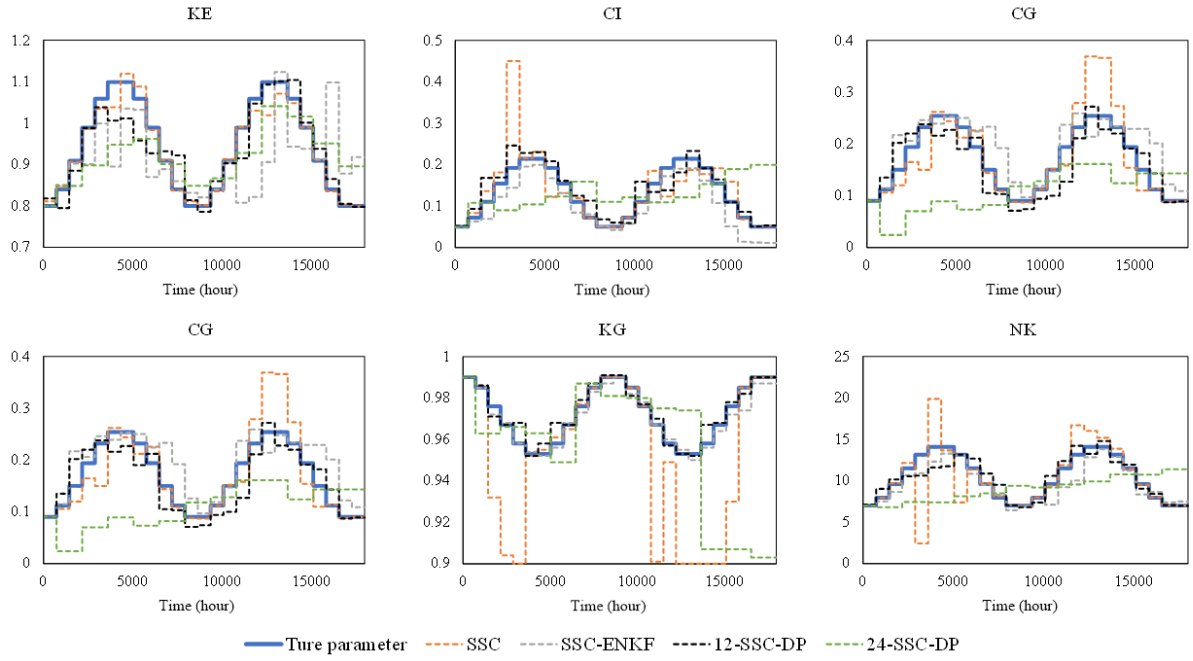


Figure 10 Comparison between estimated parameters and their true values for scenario 3 of the synthetic experiment with the Xinanjiang model.

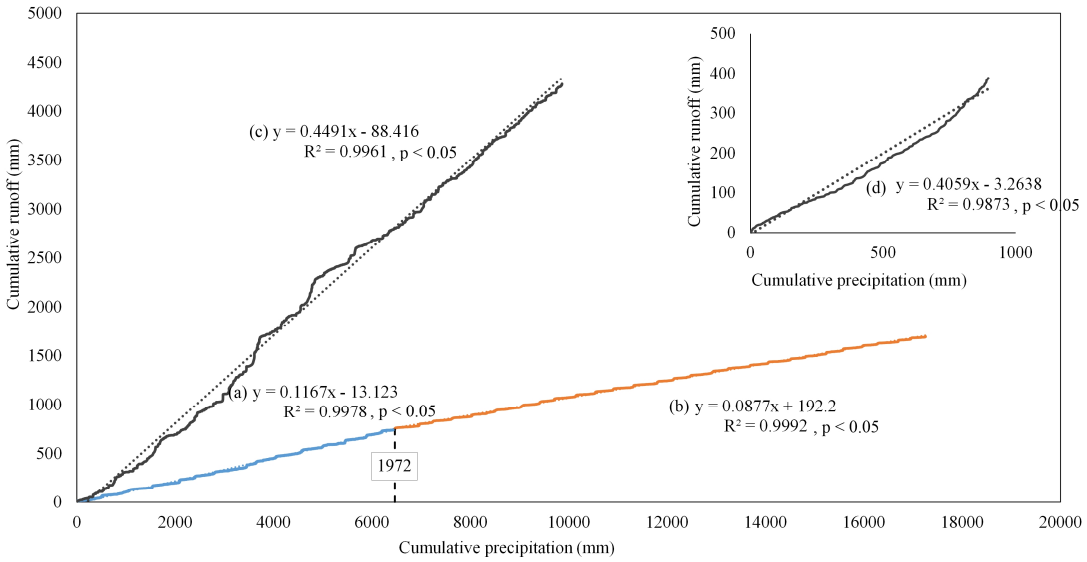


Figure 11 Double mass curves between daily runoff and precipitation for (a) Wuding River basin from 1958–1972; (b) Wuding River basin from 1973–2000; (c) Xun River basin from 1991–2001. Subgraph (d) represents the double mass curve between the mean daily runoff and precipitation from 1991–2001.

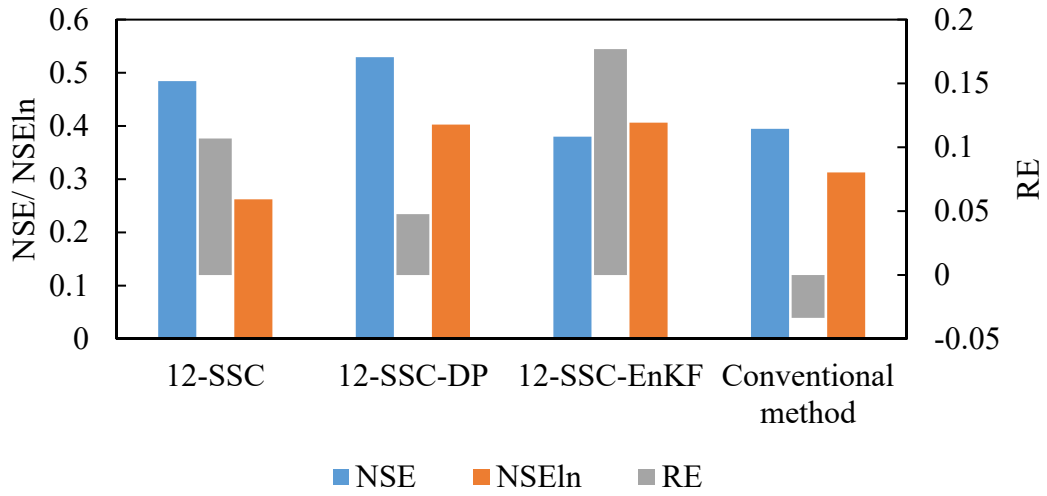


Figure 12 Simulation performance for streamflow in the Wuding River basin.

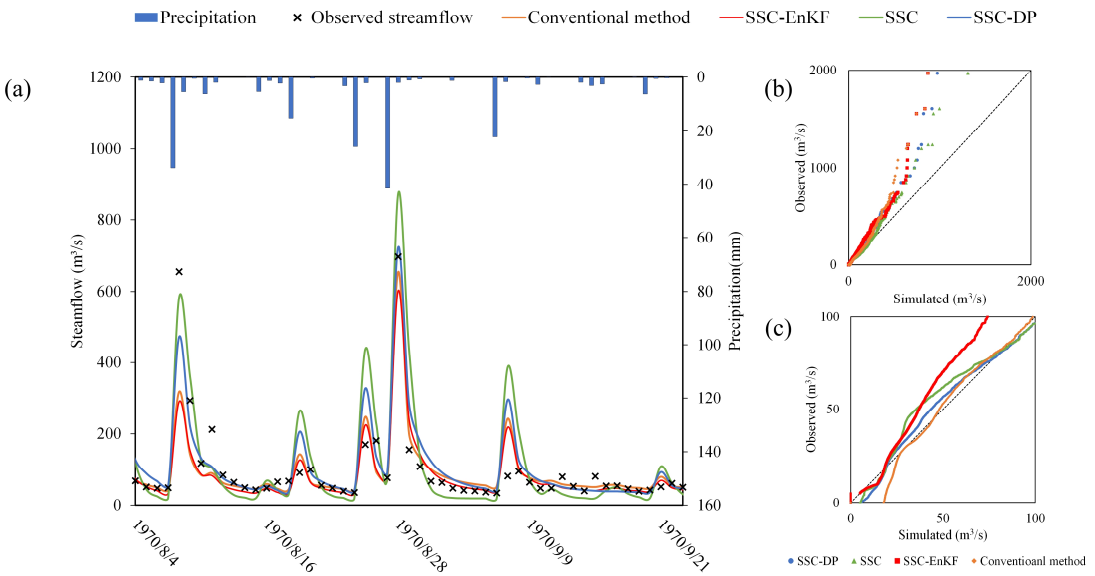


Figure 13 The simulated and observed streamflow using the conventional method, SSC-EnKF, SSC, and SSC-DP for the Wuding River basin. (a) Streamflow simulation hydrograph; (b) The quantile-quantile plot for all streamflow; (c) The quantile-quantile plot for streamflow lower than 100 m^3/s.

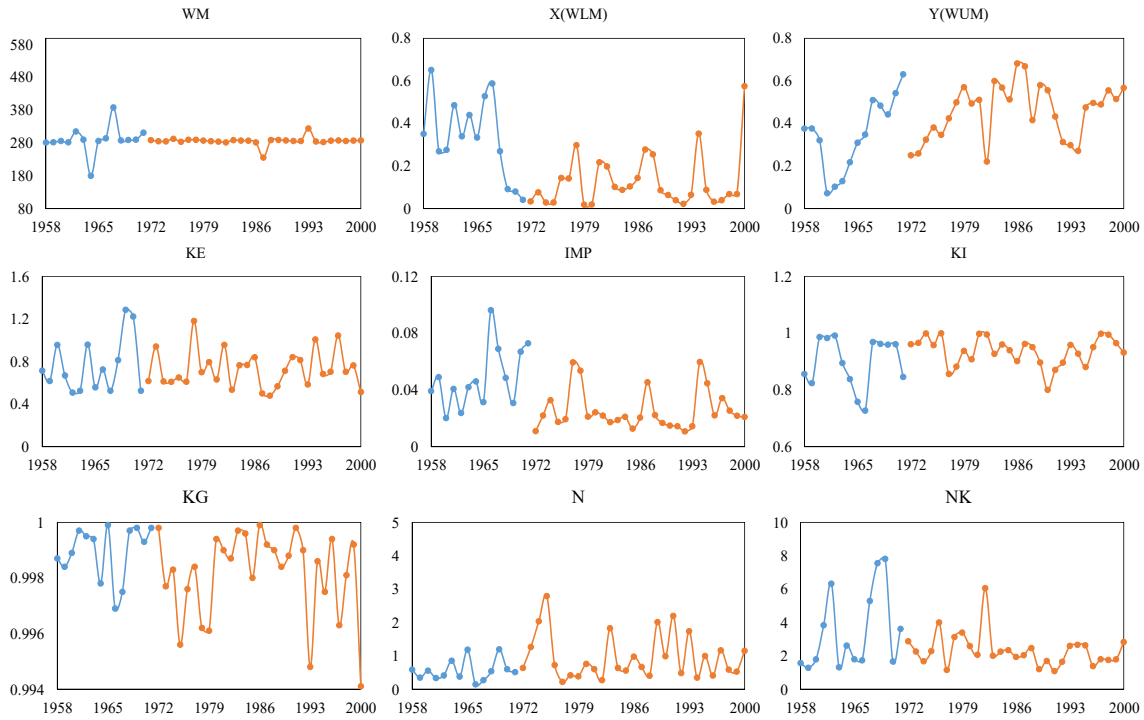


Figure 14 Estimated sensitive parameters of the Xinanjiang model for the Wuding River basin. The blue and orange solid lines represent the estimated parameters pre- and post-1972, respectively.

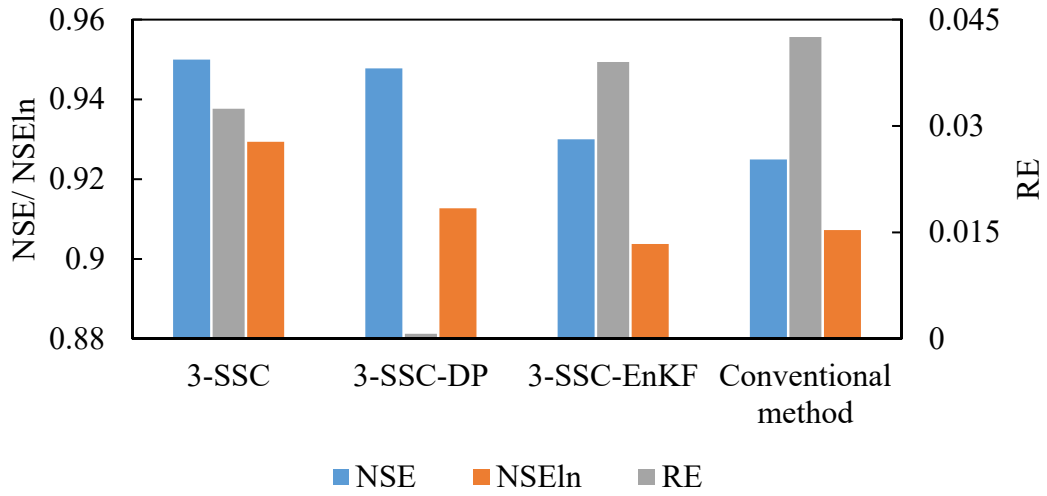


Figure 15 Simulation performance for streamflow in the Xun River basin.

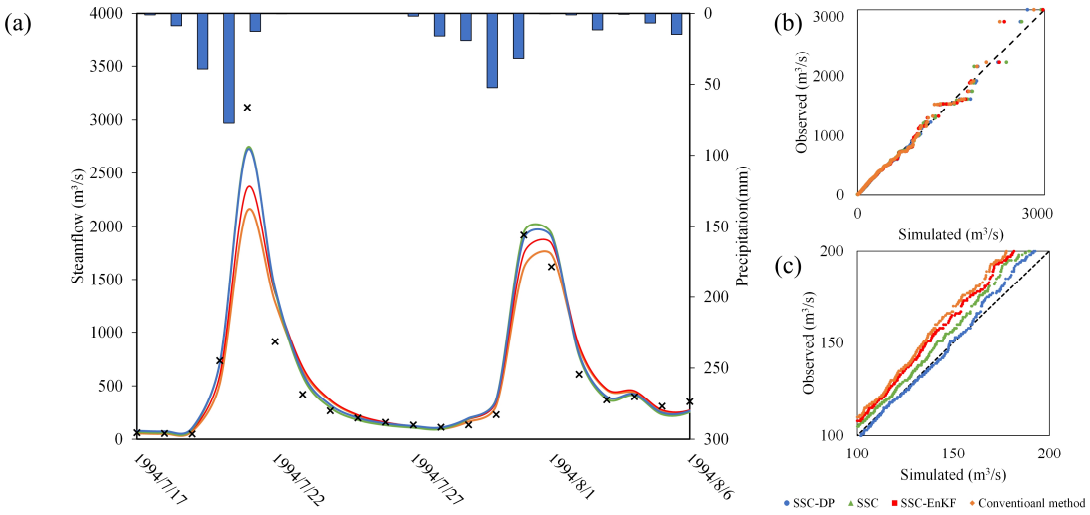


Figure 16 The simulated and observed streamflow using the conventional method, SSC-EnKF, SSC, and SSC-DP for the Xun River basin. (a) Streamflow simulation hydrograph; (b) The quantile-quantile plot for all streamflow; (c) The quantile-quantile plot for streamflow ranging from 100 m^3/s to 200 m^3/s.

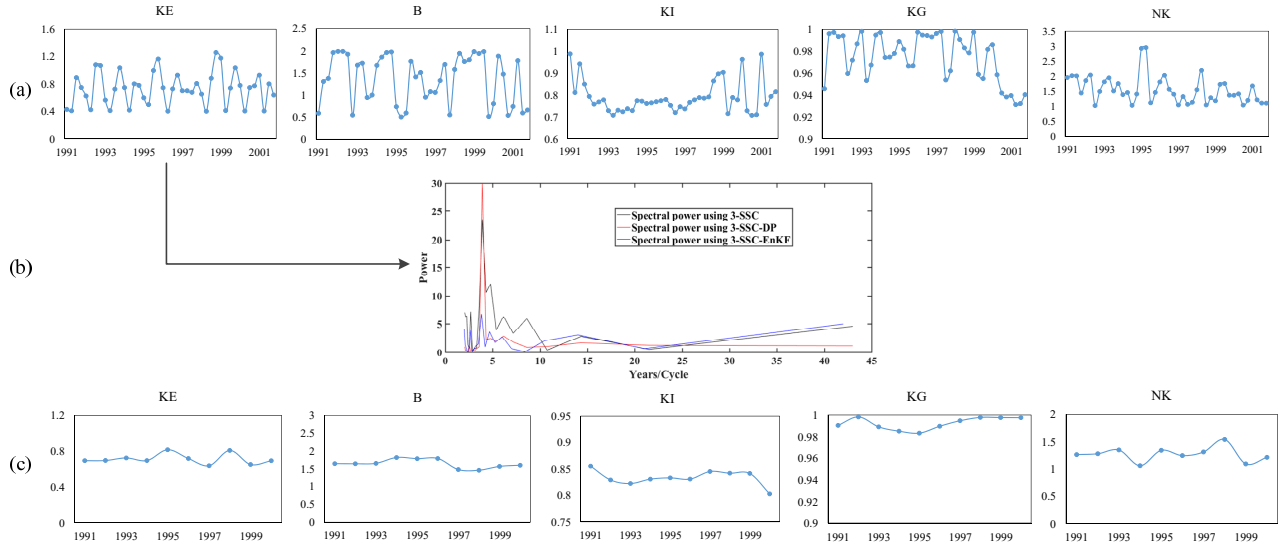
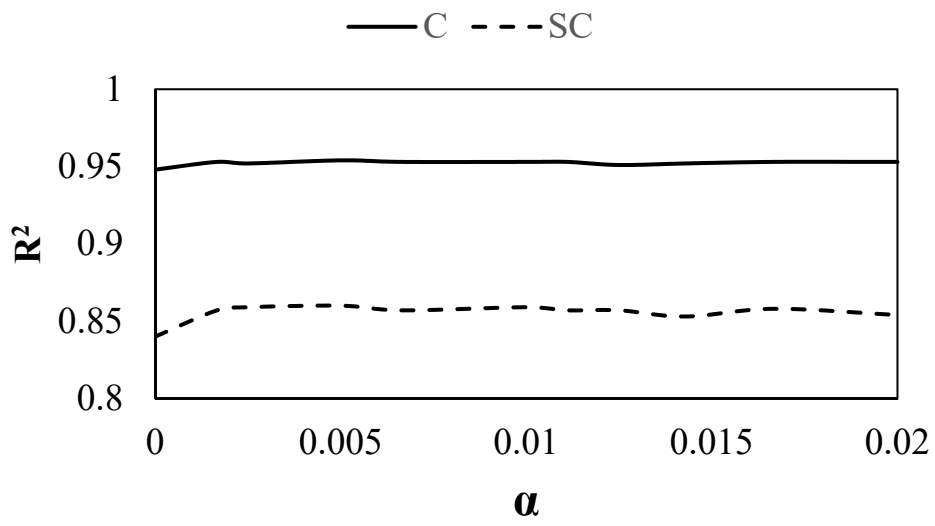
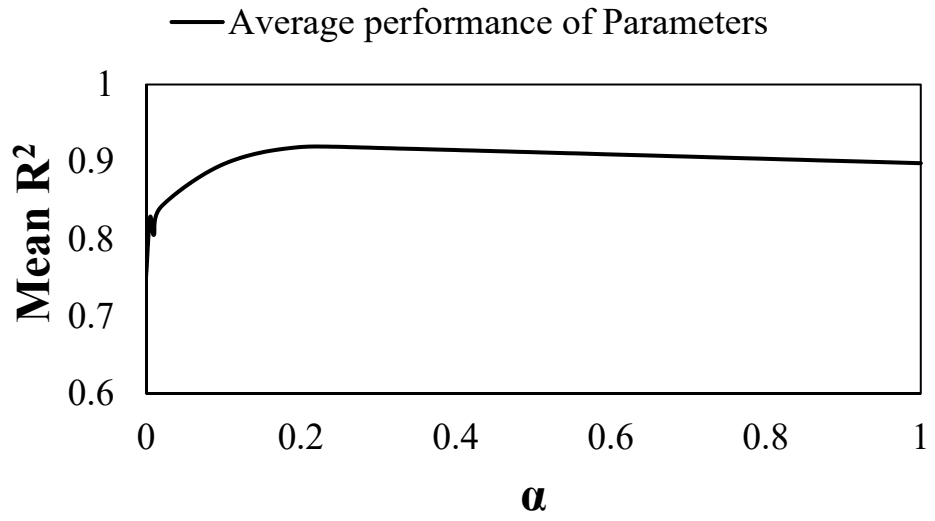


Figure 17 Estimated sensitive parameters of the Xinanjiang model for the Xun River basin over (a) seasonal time scale and (c) annual time scale. Plot (b) illustrates the spectral power of parameter KE using different methods.



(a)



(b)

Figure 18 Correlation efficiency results of SSC-DP using different weights of parameter continuity for synthetic experiments with (a) TMWB model and (b) Xinanjiang model. The mean R^2 is the average value of the R^2 such that the identification results for parameters with different ranges can be summarized.