

Interactive comment on “A time-varying parameter estimation approach using split-sample calibration based on dynamic programming” by Xiaojing Zhang and Pan Liu

Xiaojing Zhang and Pan Liu

liupan@whu.edu.cn

Received and published: 26 September 2020

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 not 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

C1

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 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, include 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 reflecting soil water capacity and impervious areas changed significantly after 1972, reflecting the soil and water

C2

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. (Pages 33, Lines 925-941)

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 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

C3

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 lead to a durative and long-term change in the catchment characteristic since 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 perfor-

C4

mance 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, Lines 881~905)

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 benefit 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, Lines 178~181)

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 the 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 in Fig.1. 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

C5

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 136-145)

7. The sensitive hydrograph phases of model performance criteria, i.e., RMSE, R² 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 values of RE indicate better streamflow simulations. To see the equations of NSE, RE and NSE_{ln}, please refer to the supplement.(Pages 15~16, Lines 324-333) Description of the evaluation results has been added in Revised Manuscript as follows: īČÝ 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

C6

worst for NSE and NSEIn, 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 NSEIn, the SSC-DP offers improved accuracy than the EnKF if the proper length is chosen. Another advantage of the SSC-DP is the low 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 492~506) 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 NSEIn. 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 NSEIn. (Page 24, Lines 560~566) 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, it is 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, NSEIn and low RE. Although the 12-SSC produces relatively high NSE, but it performs worst simulations for low flows. The SSC-EnKF has relative high NSEIn, but the RE of it is the largest. Overall, the 12-SSC-DP significantly improve the simulation performance of the Xinanjiang model in the Wuding River basin. (Page 26, Lines 706~713) 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 NSEIn 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

C7

simulate water balance than the 3-SSC in the Xun River basin. (Page 28~29, Lines 785~806) (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 R2 (Kim et al., 2007). RMSE and MARE quantifies the accuracy of the estimated parameters, but RMSE is more sensitive to high values than MARE. R2 records the overall agreement between the true and estimated parameters. Smaller values of RMSE, MARE and higher values of R2 indicate stronger parameter identification ability. To see the equations of RMSE, MARE and R2, please refer to the supplement. (Pages 16-17, Lines 334-345) 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 a constant value (scenario 1), 12-SSC-DP gives the best performance with the lowest RMSE, MARE and highest R2. 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 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).

C8

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 22~24, Lines 507-546) For results of the synthetic experiment with the Xinanjiang model Figures 9(b) and (c) compares 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 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 24, Lines 652-666) (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. 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 km² (98 % of the total basin) and records the

C9

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. (Pages 19~20, Lines 429-440)

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 km² (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. (Pages 19~20, Lines 429-440) 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 exhibits strong seasonal patterns in these three climatic and hydrological variables, suggesting that seasonal variations in hydrological parameters should be considered. (Pages 21, Lines 472-477)

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

C10

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 . To see the equation, please refer to the supplement. In the equation, there is a 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. (Pages 5, Lines 89-92)

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.

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

C11

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.

Please also note the supplement to this comment:

<https://hess.copernicus.org/preprints/hess-2019-639/hess-2019-639-AC1-supplement.pdf>

Interactive comment on *Hydrol. Earth Syst. Sci. Discuss.*, <https://doi.org/10.5194/hess-2019-639>, 2019.

C12

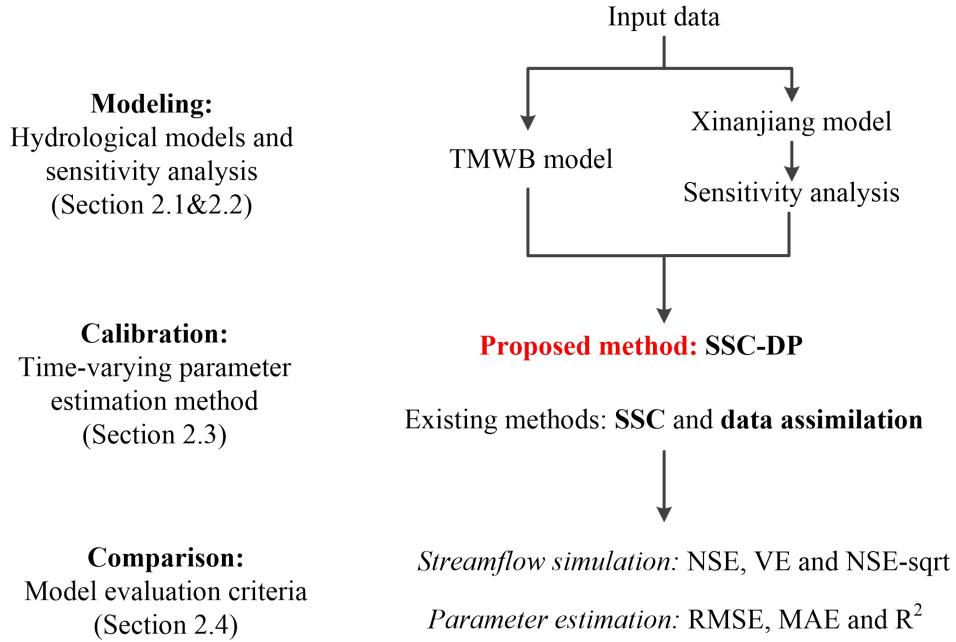


Fig. 1. Figure 1 The flowchart of the methodologies

C13

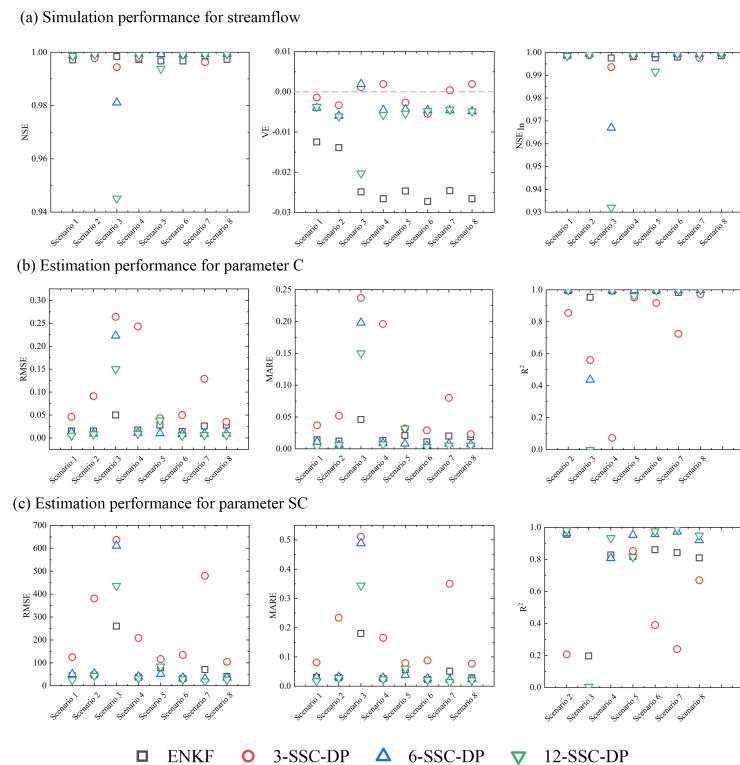


Fig. 2. Figure 6 Comparison between the EnKF and SSC-DP methods for (a) streamflow simulation and identification of (b) parameter C and (c) parameter SC.

C14

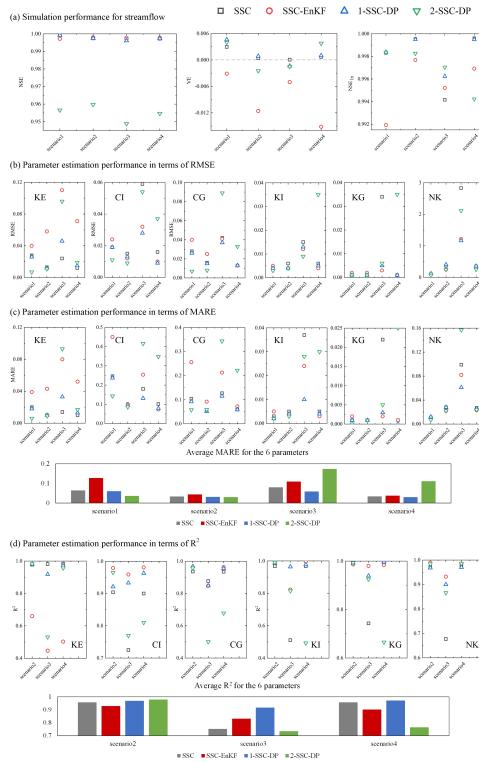


Fig. 3. Figure 9 Comparison among the SSC, SSC-EnKF and SSC-DP methods for (a) streamflow simulation and parameter identification in terms of (b) RMSE, (c) MARE and (d) R2.

C15

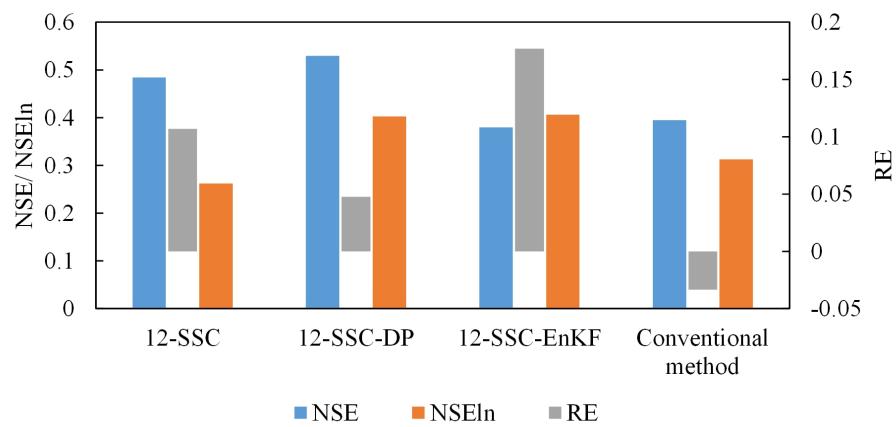


Fig. 4. Figure 12 Simulation performance for streamflow in the Wuding River basin.

C16

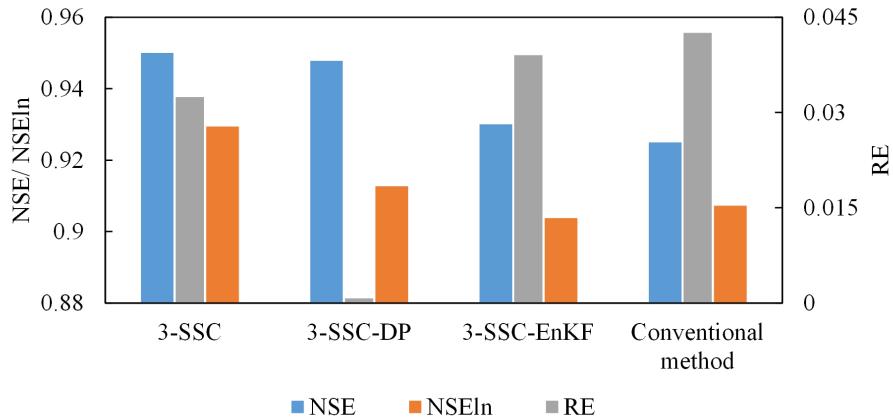


Fig. 5. Figure 15 Simulation performance for streamflow in the Xun River basin.

C17

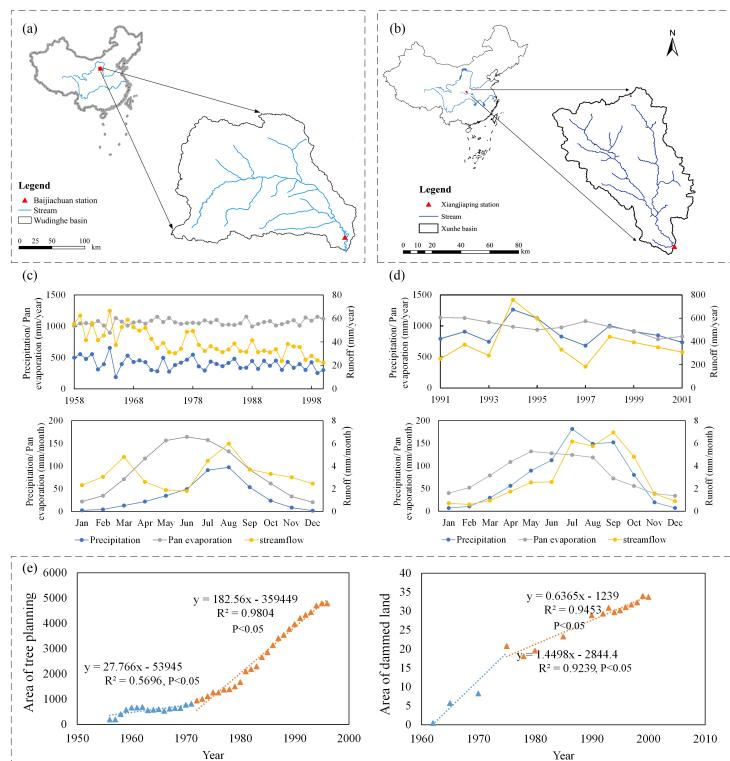


Fig. 6. Figure 5 Location of (a) Wuding River basin and (b) Xun River basin. The plots (c) and (d) show the average yearly and monthly variations of precipitation, pan evaporation and streamflow in the Wuding

C18