

## Responses

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;
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

---

**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 revision.

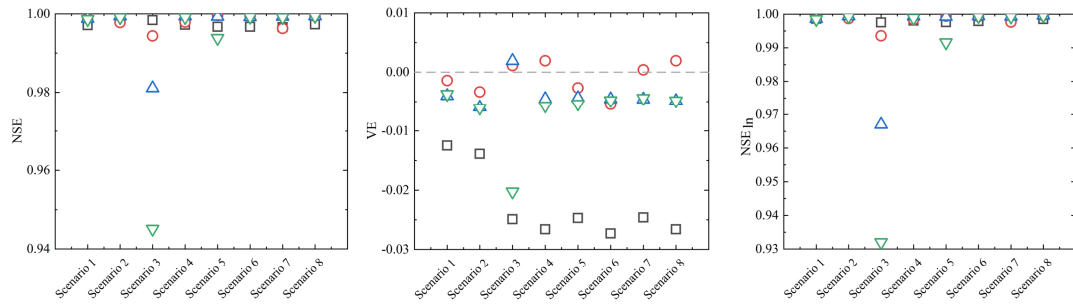
2. GENERAL COMMENTS

(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.

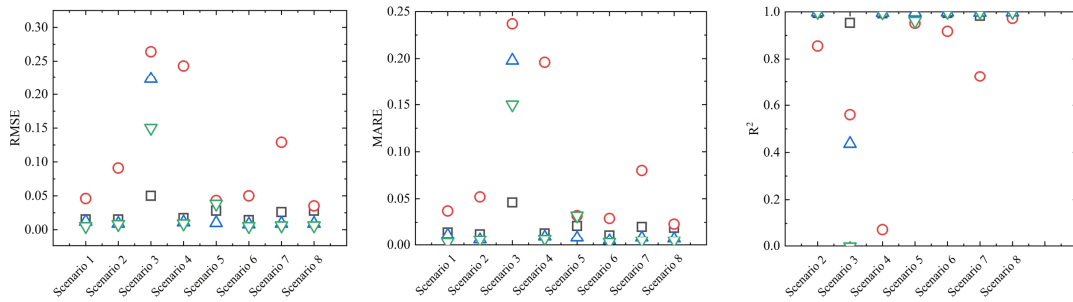
**Reply:**

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

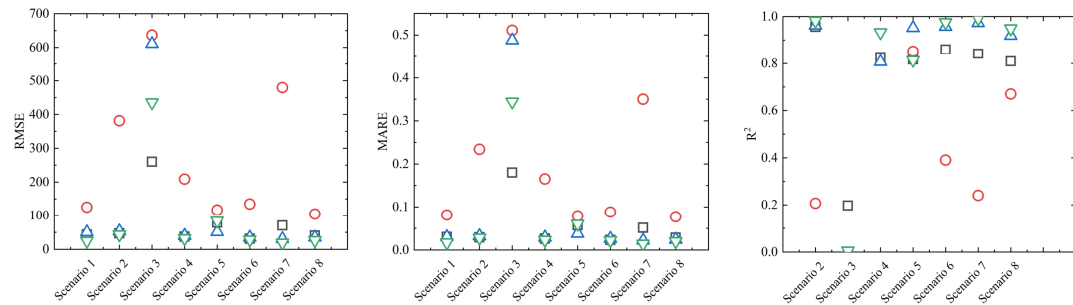
(a) Simulation performance for streamflow



(b) Estimation performance for parameter C



(c) Estimation performance for parameter SC



□ ENKF    ○ 3-SSC-DP    △ 6-SSC-DP    ▽ 12-SSC-DP

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:

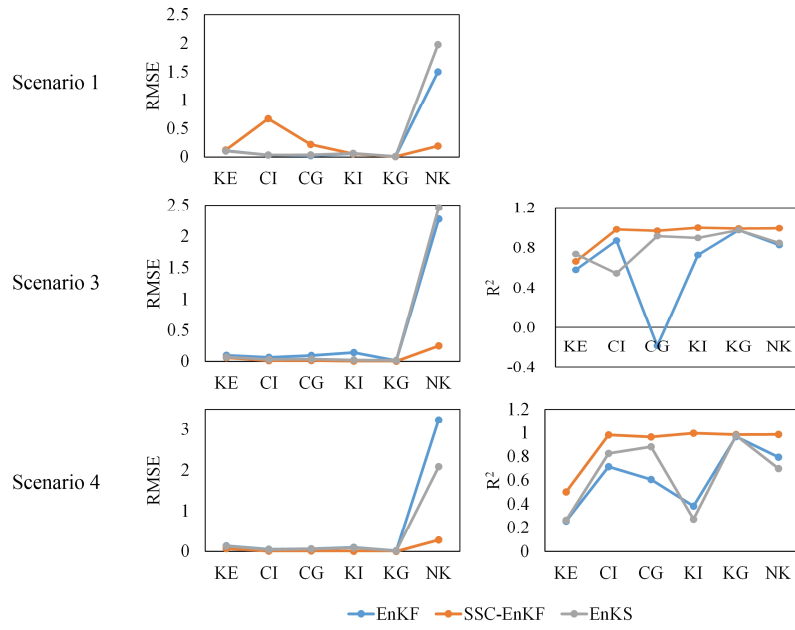


Figure 8 Comparison among EnKF, SSC-EnKF, and EnKS in the synthetic experiment with the Xinanjiang model

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

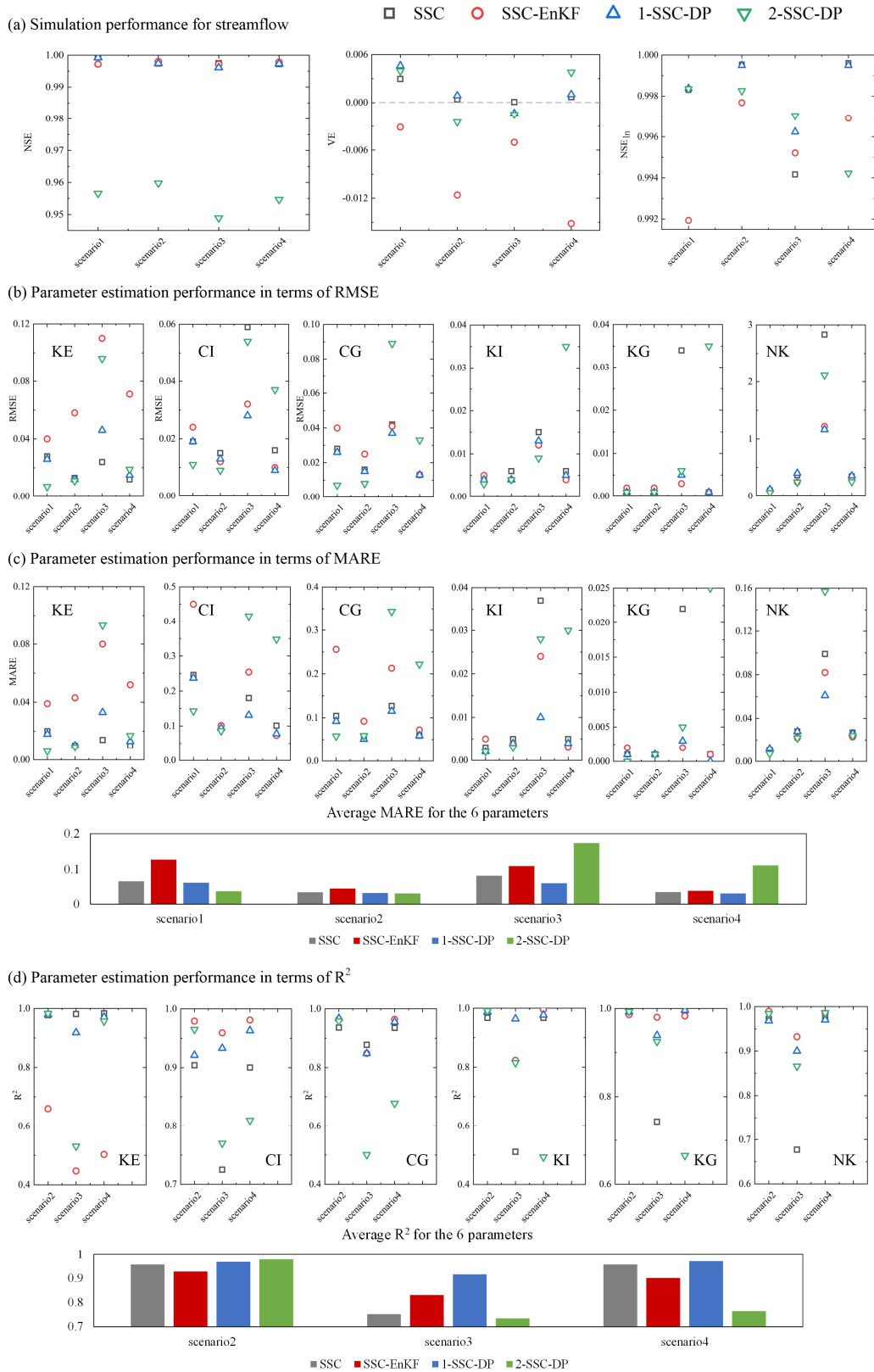


Figure 9 Comparison among the SSC, SSC-EnKF and SSC-DP methods for (a) streamflow simulation and parameter identification in terms of (b) RMSE, (c) MARE and (d)  $R^2$ .

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

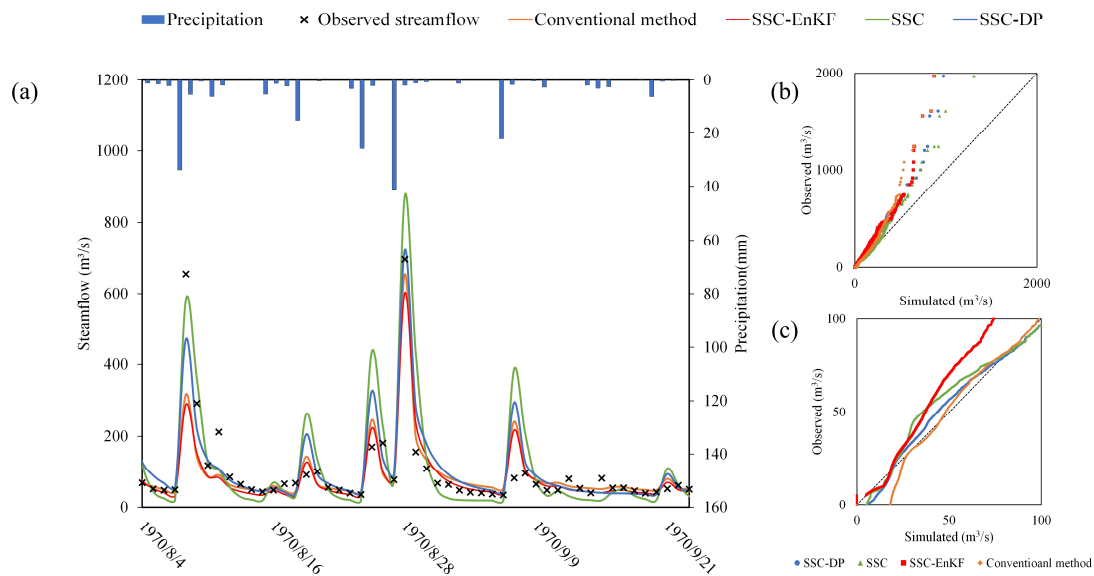


Figure 13 Streamflow simulation hydrograph (left panels) and quantile-quantile plots (right panels) using conventional method, SSC-EnKF, SSC, and SSC-DP for the Wuding River basin. (a) The quantile-quantile plot for all streamflow; (b) The quantile-quantile plot for streamflow lower than  $100 \text{ m}^3/\text{s}$ .

(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 at Line 137, line 150 and in Tables 1 and 2 as follows:

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**). **(Pages 8, Lines 164~166)**

Table 1 Parameters of the TMWB model

Parameter	Physical meaning	Range and units
<i>C</i>	Evapotranspiration parameter	0.2-2.0 (-)
<i>SC</i>	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. **(Pages 9, Lines 176~177)**

Table 2 Parameters of the Xinanjiang model

Category	Parameter	Physical meaning	Range and units
Evapotran spiration	WM	Tension water capacity	80-400 (mm)
	X	$WUM=X \times WM$ , WUM is the tension water capacity of lower layer	0.01-0.8 (-)
	Y	$WLM=Y \times WM$ , 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 Xinanjiang 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.

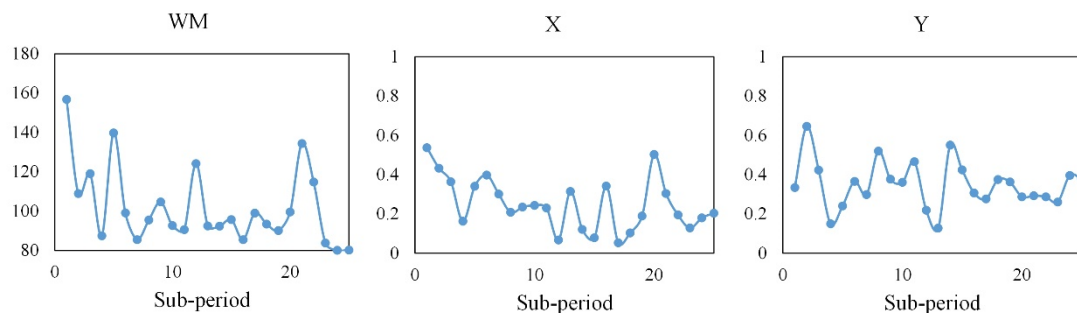


Figure R1 The parameters estimated without a sensitivity analysis

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. **(Pages 7, Lines 139-141)**

(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:

$$\mathcal{G}_{t+1}^{k-} = \mathcal{G}_t^{k+} + \delta_t^k, \delta_t^k \sim N(0, R_t) \quad (1)$$

where  $\mathcal{G}_{t+1}^{k-}$  is the forecasted parameter vector at the time step  $t+1$ , while  $\mathcal{G}_t^{k+}$  is the well-calibrated parameter vector.  $\delta_t^k$  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 the soil water storage capacity. [\(Pages 5, Lines 89-92\)](#)

(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<sub>ln</sub>).

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<sub>ln</sub>) (Hock, 1999). RE evaluates the error of the total volume of streamflow, while NSE and NSE<sub>ln</sub> evaluate the agreement between the hydrograph of observations and simulations. NSE is more sensitive to high flows, but NSE<sub>ln</sub> focuses more on low flows. Higher values of NSE, NSE<sub>ln</sub> and lower values of RE indicate better streamflow simulations. The NSE, RE and NSE<sub>ln</sub> are expressed as followed: [\(Pages 15~16, Lines 324-333\)](#)

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

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

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

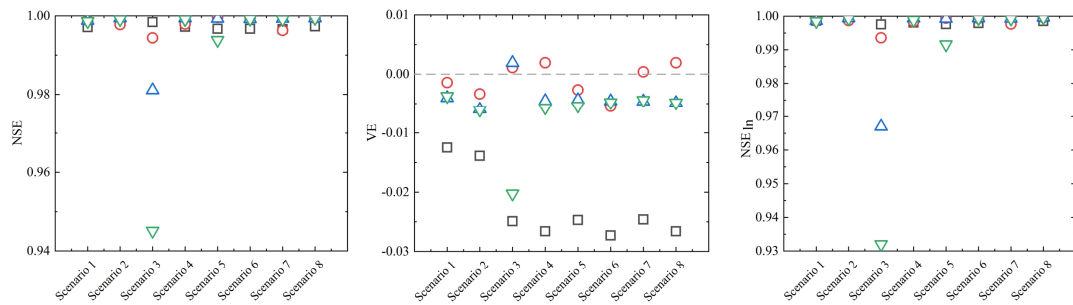
A description of the evaluation results has been added as follows:

- For results of the synthetic experiment with the TMWB model

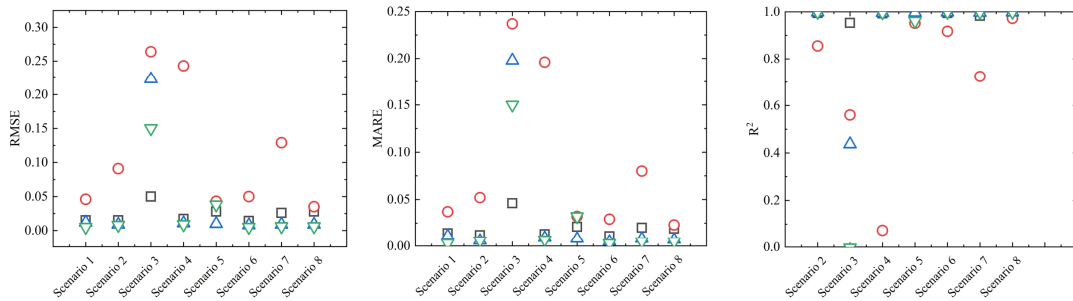
Figure 6(a) presents the runoff simulation performance for various scenarios. In scenario 1, the NSE values of the three SSC-DP methods are all higher than that of EnKF. The results of NSE<sub>ln</sub> show no significant differences among various methods.

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  $NSE_{In}$ . In scenario 3, where the true parameters have periodic variations and change every month, the NSE and  $NSE_{In}$  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  $NSE_{In}$ , 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  $NSE_{In}$ , 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)**

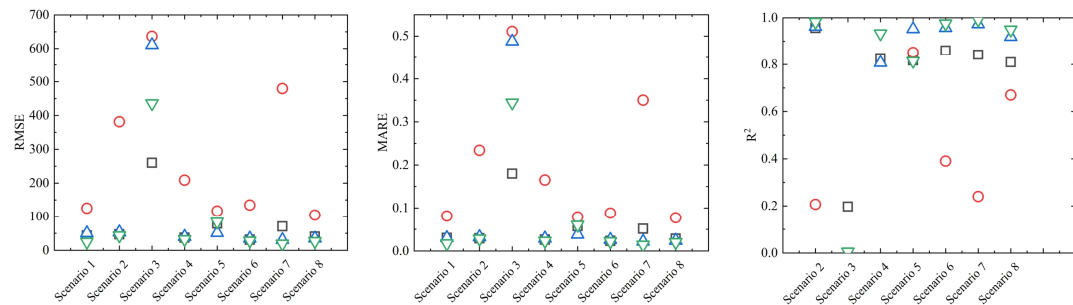
(a) Simulation performance for streamflow



(b) Estimation performance for parameter C



(c) Estimation performance for parameter SC



□ ENKF    ○ 3-SSC-DP    △ 6-SSC-DP    ▽ 12-SSC-DP

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  $NSE_{in}$ . 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  $NSE_{in}$ . **(Page 24, Lines 560~566)**

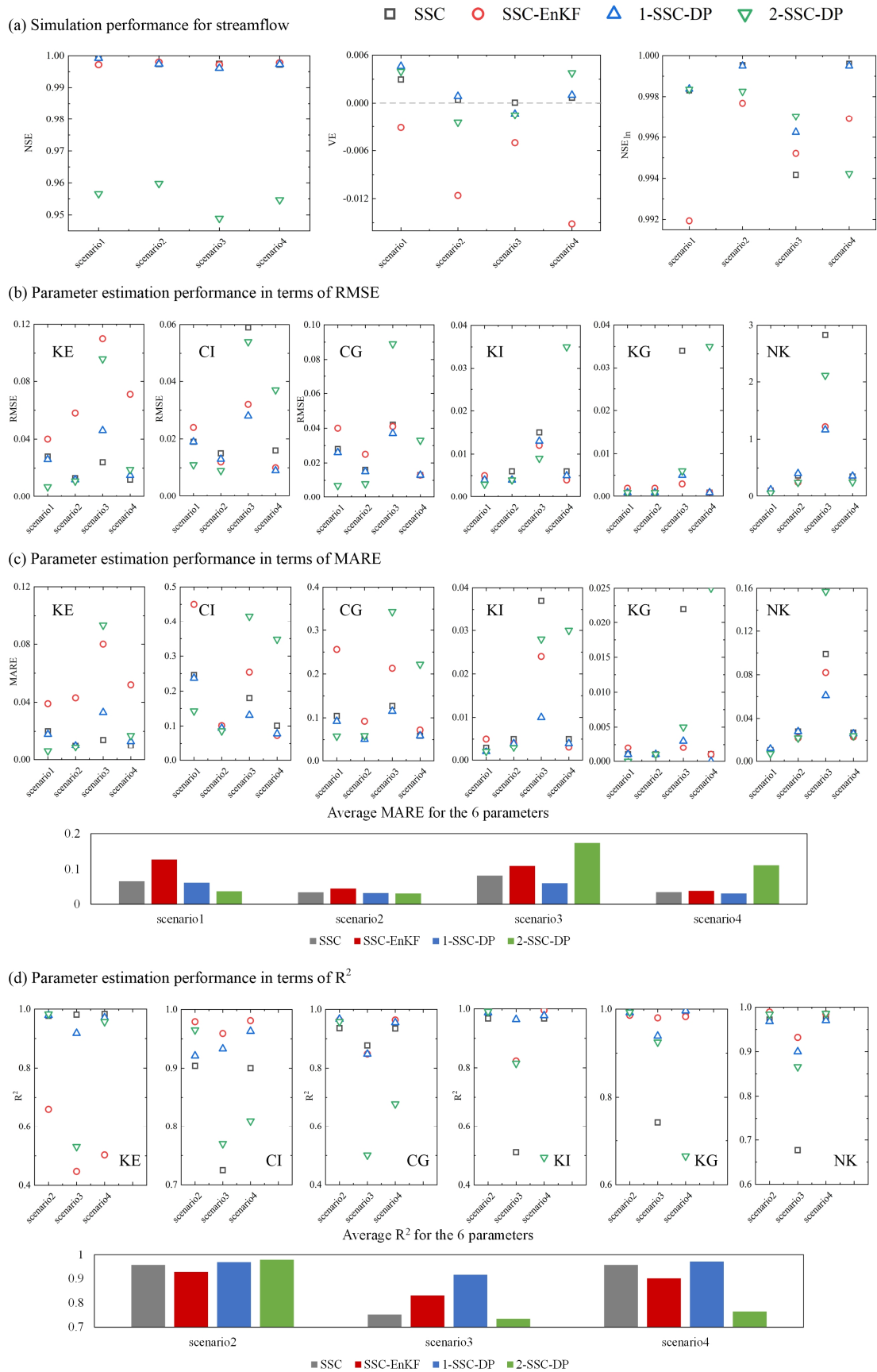


Figure 9 Comparison among the SSC, SSC-EnKF and SSC-DP methods for (a)

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, 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,  $NSE_{ln}$  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  $NSE_{ln}$ , 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)**

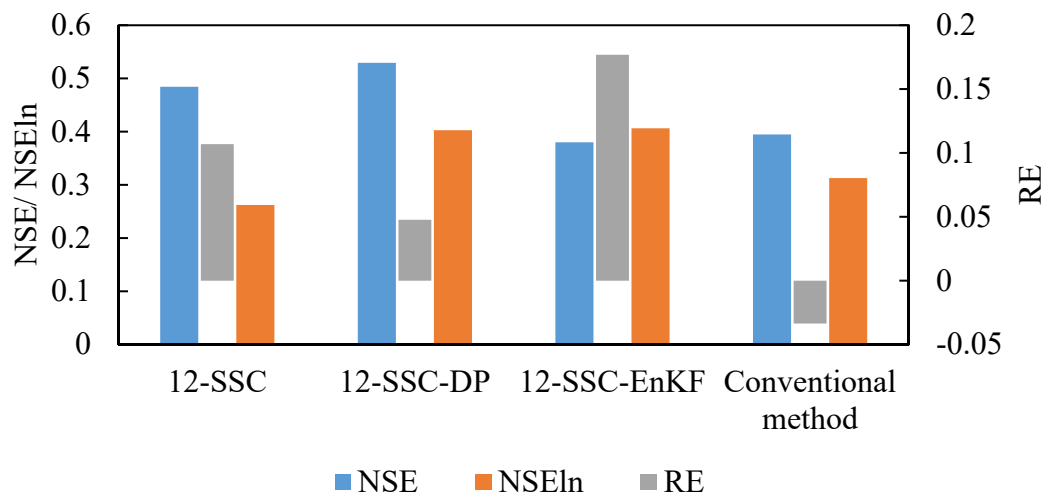


Figure 12 Simulation performance for streamflow in the Wuding River basin.

➤ 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 785~806)**

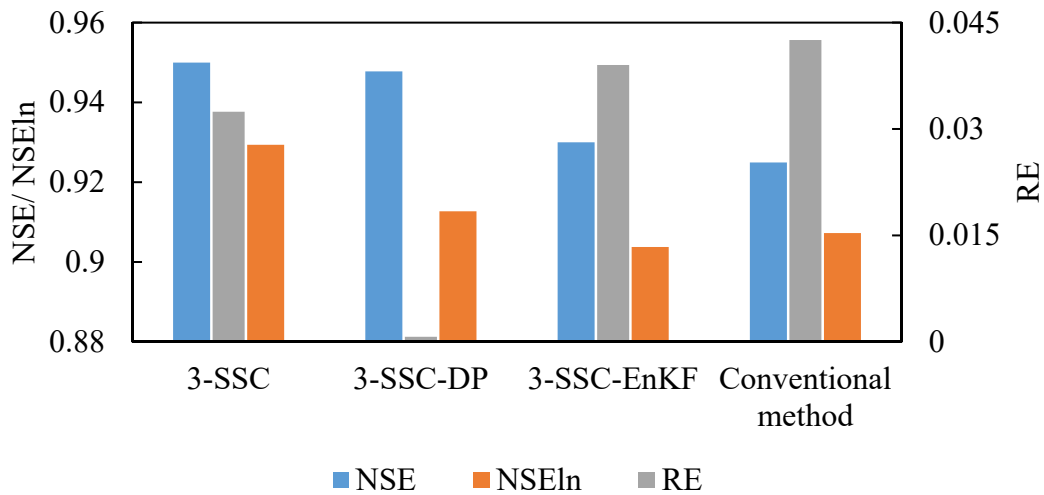


Figure 15 Simulation performance for streamflow in the Xun River basin.

(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:

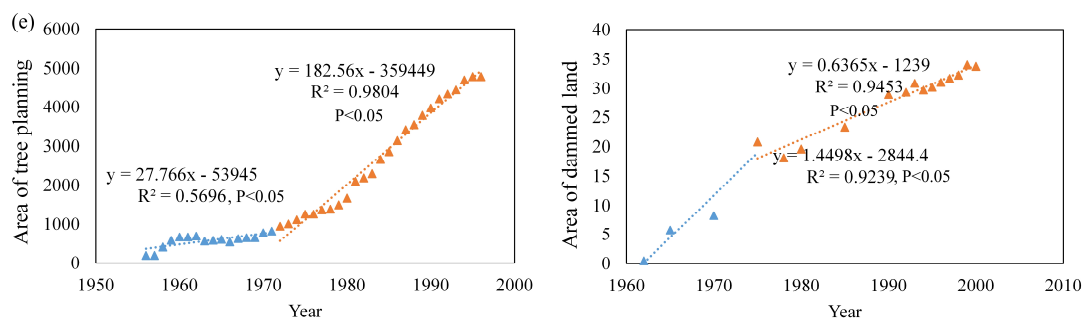


Figure 5(e) Temporal variations in the soil and water conservation measures.



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 planting 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 underlying surface of the watershed, and impacted the relationship between precipitation and runoff (Gao et al., 2017; Jiao et al., 2017). **(Pages 20, Lines 441-449)**

(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, Lines 433~436)**

(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 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)**

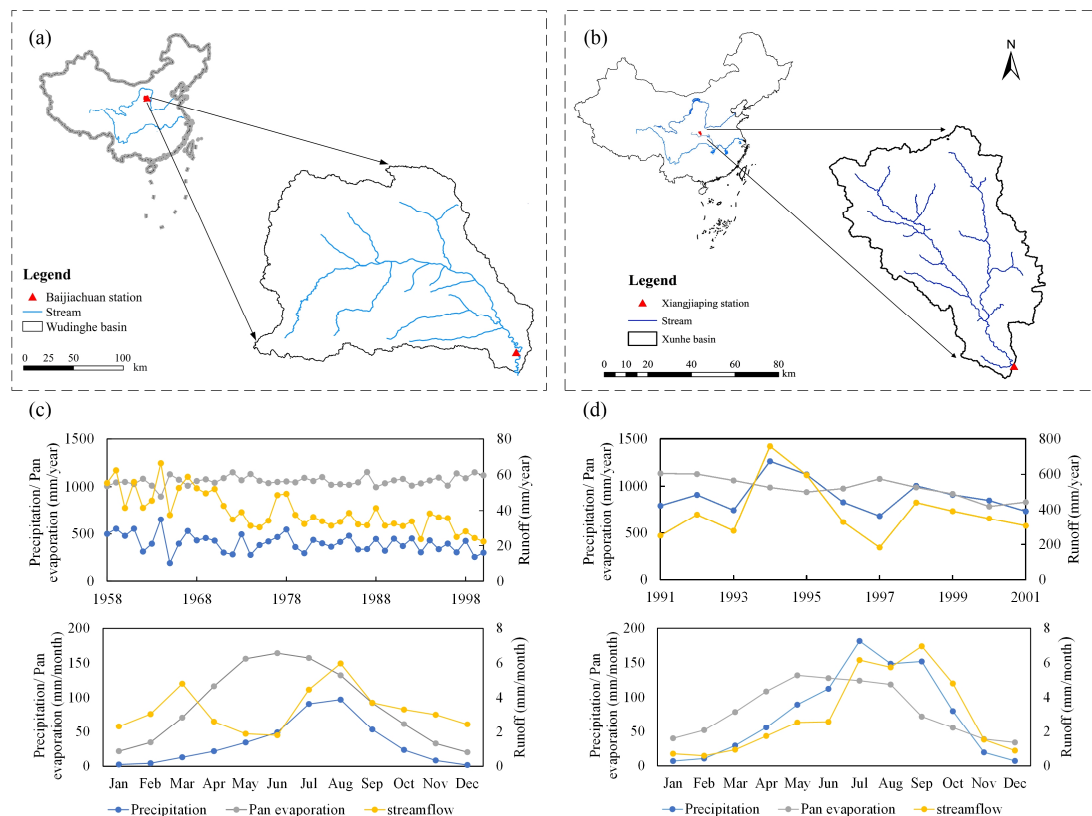


Figure 5 Location of (a) Wuding River basin and (b) Xun River basin. The plots (c) and (d) show the average yearly and monthly variations of precipitation, pan evaporation and streamflow in the Wuding River basin and Xun River basin, respectively.

(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. **(Pages 21, Lines 469-471)**

(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 534-546)**

(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 656-666)**

(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, Lines 665~666)**

(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 that 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:

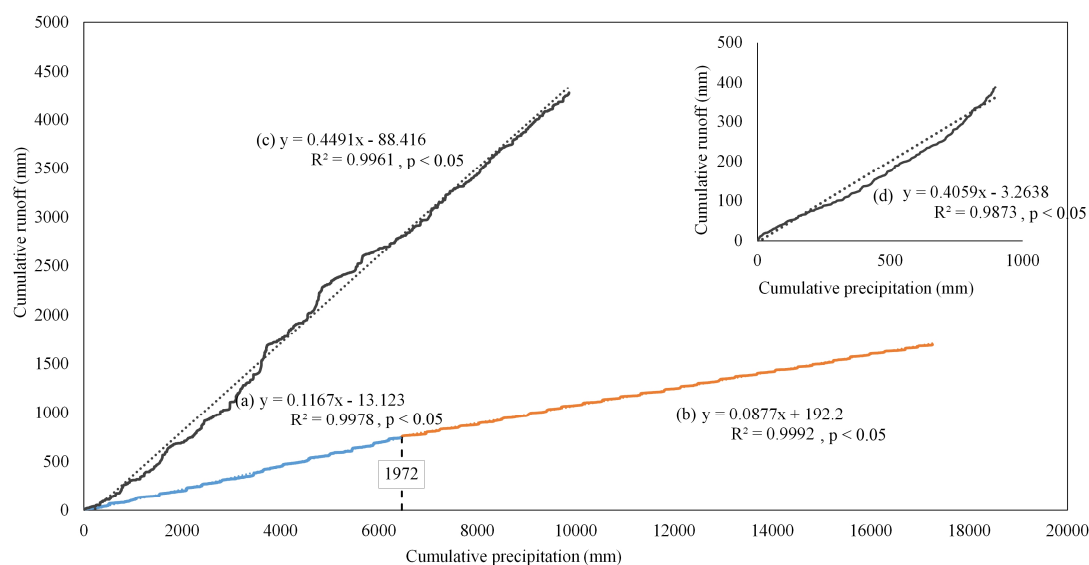


Fig. 11 Double mass curves between daily runoff and precipitation for (a) Wuding River basin from 1958–1972; (b) Wuding River basin from 1973–2000; (c) Xun River basin from 1991–2001. Subgraph (d) represents the double mass curve between the mean daily runoff and precipitation from 1991–2001.

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 669–672)**

(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 in the Revised Manuscript. **(Page 26, Lines 706–708; Pages 26–27, Lines 720–735)**

(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.

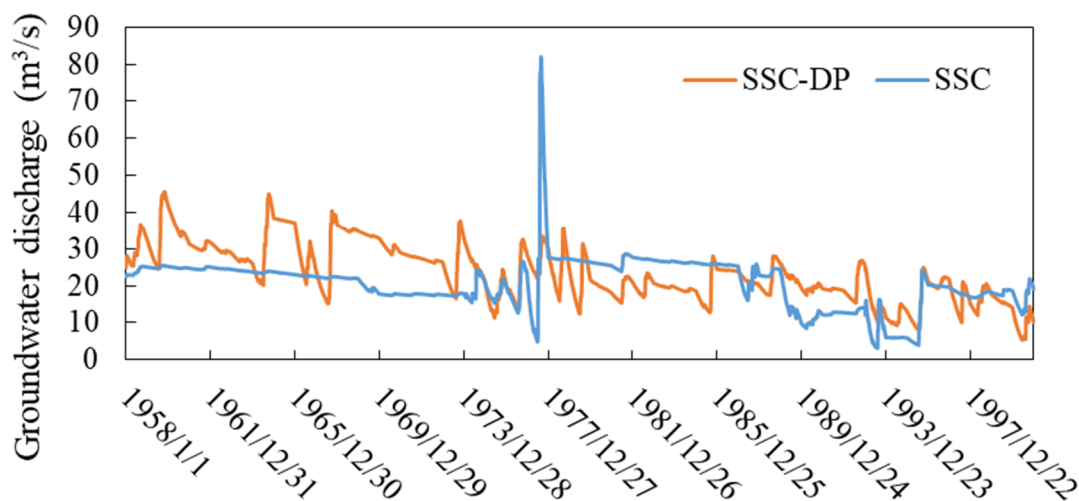


Fig. R2 The estimated groundwater discharge of the Wuding River basin

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 dramatic 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 the 12-SSC-DP is more applicable to the simulation of streamflow in the Wuding River basin. (Pages 27, Lines 739-740)

(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  $\rho=-0.38$ ,  $P<0.05$ ) between the areas of tree planning and WUM. While WLM has a nonsignificant positive correlation ( $\rho=0.26$ ,  $P>0.05$ ) 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 ( $\rho=-0.33$ ,  $P<0.05$ ). Except for afforestation, the areas of the dammed lands are significantly correlated with WLM ( $\rho=0.46$ ,  $P<0.05$ ), 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 741-770)**

(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 is higher than those of 3-SSC-DP, and 3-SSC-DP simulate better the flows ranging from 100 m<sup>3</sup>/s to 200 m<sup>3</sup>/s. **(Pages 29, Lines 806-809)**

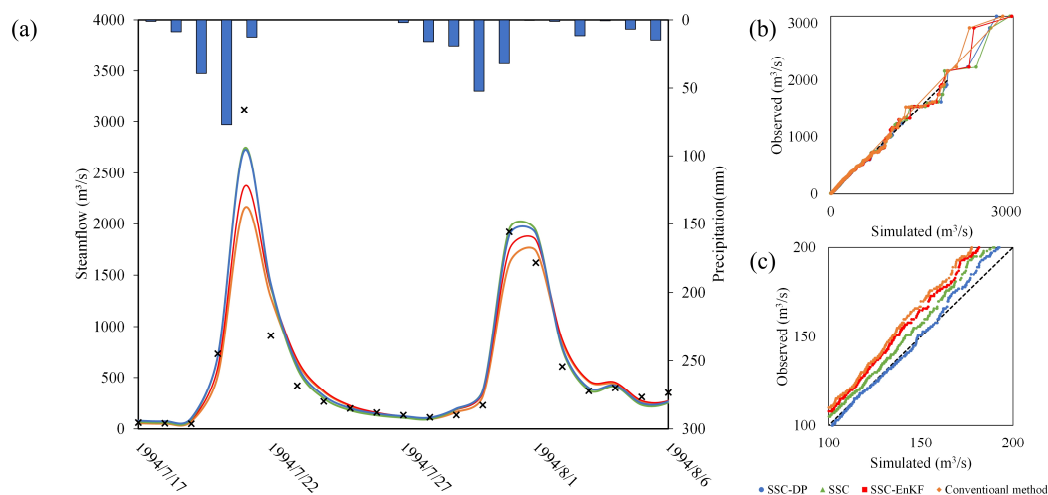


Figure 16 (a) Streamflow simulation hydrograph (left panels) and quantile-quantile plots (right panels) using conventional method, SSC-EnKF, SSC, and SSC-DP for the Xun River basin. (b) The quantile-quantile plot for all streamflow; (c) The quantile-quantile plot for streamflow ranging from 100 m<sup>3</sup>/s to 200 m<sup>3</sup>/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<sub>ln</sub>), 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 789-806)

(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.

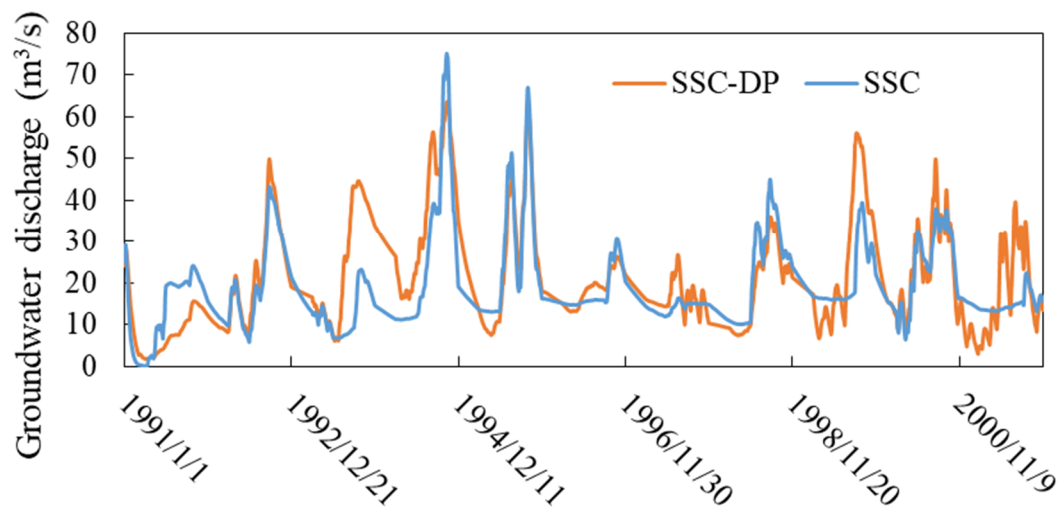


Fig. R3 The estimated groundwater discharge of the Xun River basin

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.