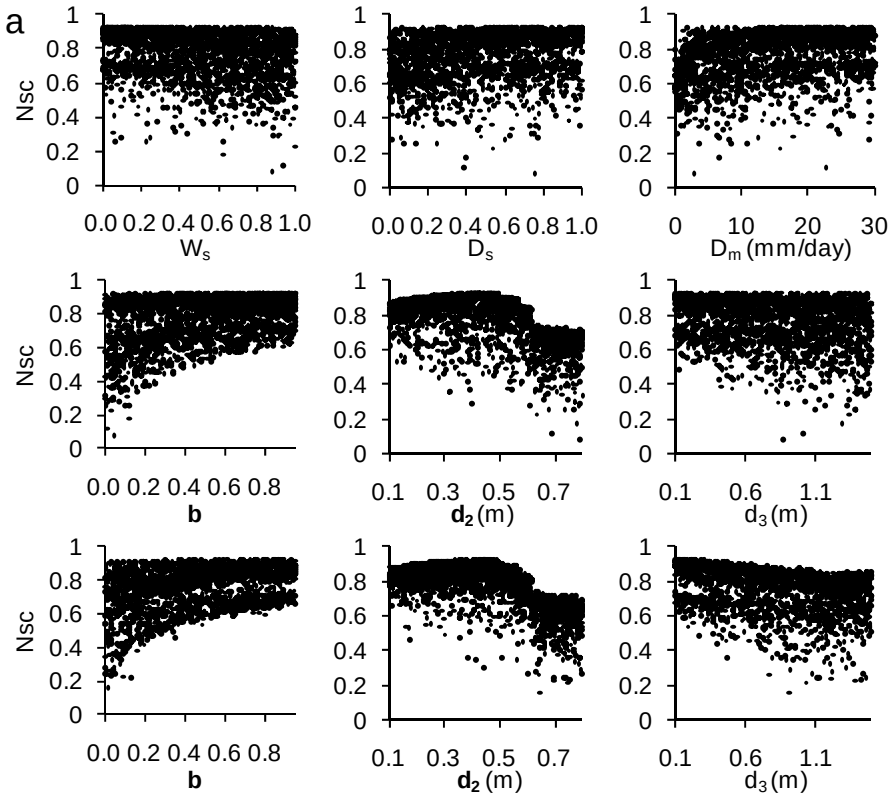
Responses to E. Demaria

First of all, we thanks for your thoughtful comments. The responses to the comments are listed point by point as following:

Major comments

1. My main concern is related to the number of Monte Carlo simulations used to perform the sensitivity analysis. The Authors do not specify neither the number of simulations nor the sampling method used to perform the analysis. I think the small number simulation is masking the results (based on the plots the number of MC does not seem be to be larger than 100). This can be solved by reducing the number of catchments to 3-4 based on the number of hydroclimatic environments found in China and increasing the number of MC simulations only for those selected basins.

We agreed with your comments. The following part will be added to the revised paper. The Latin Hypercube methodology was used for the Monte Carlo simulations. The previous numbers of the Monte Carlo simulations was 600, 400, and 400 for Gaoqitou, Taolinkou, and Minhe catchment, respectively. In order to more accurately capture the parameters spaces, the number of Monte Carlo simulations was increased to 2000 for the three catchments. The relationships between *Nsc* value and each parameter were shown in Fig. R1.



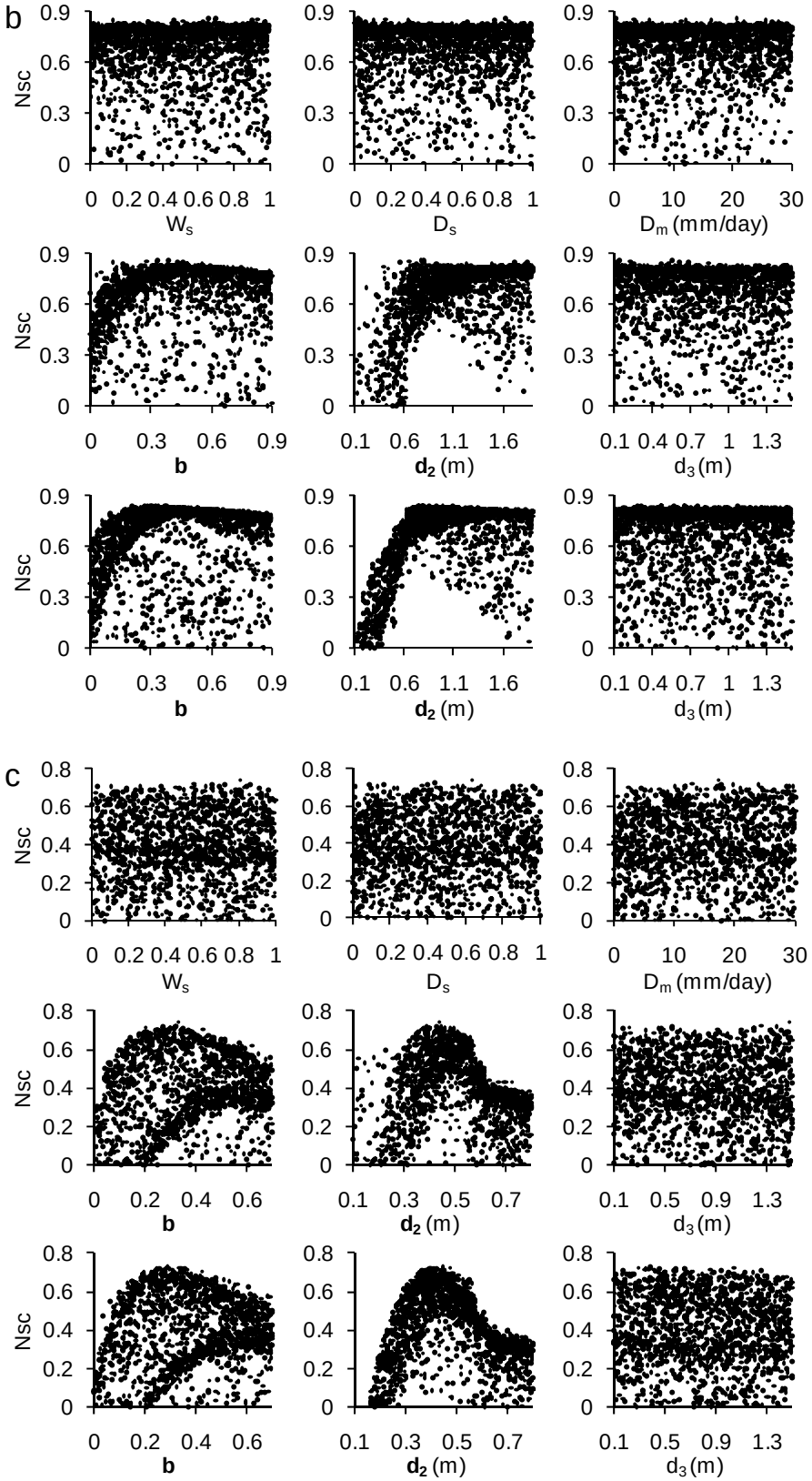


Fig. R1. Scatterplots between model parameters and Nash-Sutcliffe coefficient (*Nsc*) in three catchments. (a: Gaoqitou catchment (humid), b: Taolinkou catchment (semi-humid), c: Minhe catchment (arid); The first 6 figures were under 6-parameter methodology, and the last 3 figures were under 3-parameter methodology

2. The Authors do not show the parameter values obtained using physically-based equations. How the values (Ws, Ds and Dm) compared to the values from the sensitivity analysis? Where the soil and topographic information was obtained? What was the spatial scale of the soil maps and DEM used?

We agreed with you, and the information was important. The description of the parameter values obtained using physically-based equations would be added to section 4.2. Using the soil properties, the three baseflow parameters (D_m, W_s, and D_s) were estimated in the 24 catchments. Compared to sensitivity analysis, the values of Ws and Ds belonged to the original parameter space. But most Dm values were higher than 30 mm/day. Maybe the previous parameter space was not reasonable. The estimated values were illustrated in three catchments: Gaoqitou, Taolinkou, and Minhe catchment (Fig. R2). For the D_m-parameter, the average values were 43.6, 29.1, and 33.9 mm/day, in Gaoqitou, Taolinkou, and Minhe catchment, respectively. The variation was highest in Minhe catchment (25.6-59.5 mm/day), followed by Gaoqitou catchment (32.5-45.6 mm/day), and then Taolinkou catchment (Fig. R2 (a)). The average values of W_s-parameter (D_s-parameter), were 0.76 (0.50), 0.74 (0.27), and 0.72 (0.30), in Gaoqitou, Taolinkou, and Minhe catchment, respectively. As same as the D_m-parameter, the highest variations for W_s-parameter and D_s-parameter were in Minhe catchment, followed by Gaoqitou catchment, and then Taolinkou catchment (Fig. R2 (b and c)).

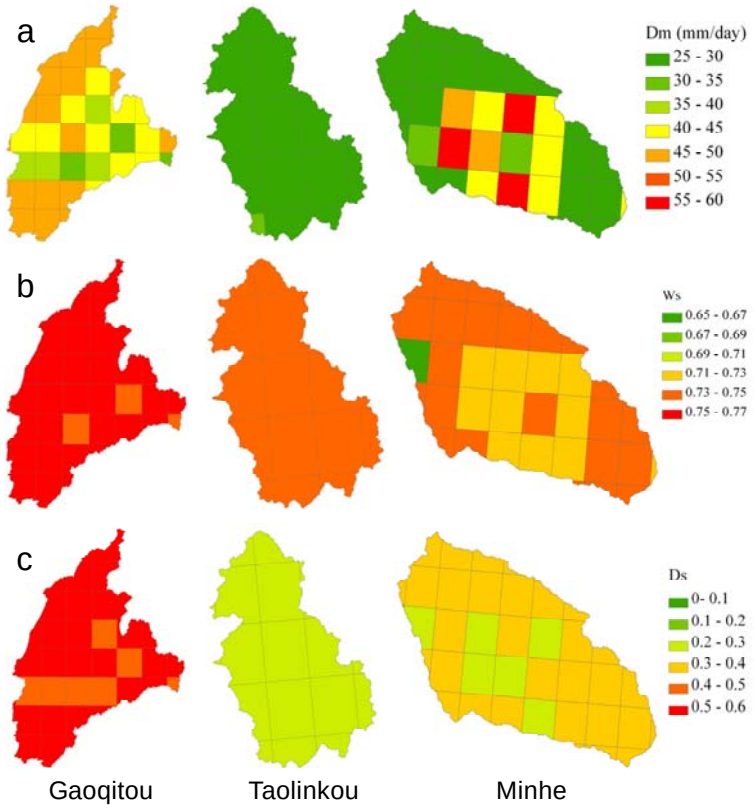


Fig. R2. The three baseflow parameters values obtained using physically-based equations in Gaoqitou, Taolinkou, and Minhe catchment, respectively. (a: D_m parameter; b: W_s parameter; c: D_s parameter)

The description of data sources would be added to section 4.1 study area and dataset.

1) The soil data was extracted from the FAO two-layer 5-minute 16-category global soil texture maps. In this datasets, the soil was classified into 16 categories, the first 12 kinds of which were used in this study. The resolution of the soil data was 5-minute, but it was 30-second in USA.

2) The relationships between soil characteristics (Ksat, etc.) value and each soil type were referenced to Rawls et al. (1998). Page 7025, line 16. The resolution of the grid cell in this study was $0.25^{o}\times0.25^{o}$. Thereby, the Ksat value in each grid cell could be averaged by the Ksat value with the resolution of 5-minute.

3) The DEM data was obtained from SRTM 90m Digital Elevation Data.

3. The Authors do not point out how efficient the 3-parameter method is compared to the traditional 6-parameter method in terms of computation and time efficiencies.

Calculating the three baseflow parameters using Eq. 8, 9, and 11 was very quick. For example, in Gaoqitou catchment, it only cost 12 seconds (The CPU of the computer was Intel (R) Core (TM) 2 Duo E6550 @ 2.33 GHz). For one catchment, it needed to be calculated only once. Compared to Monte Carlo simulation, the time for calculating the three baseflow parameters could be ignored. During the Monte Carlo simulation, the model would be run with the same times, using the two parameter setting methodology. The 3-parameter method had the same computation and time efficiencies compared to the 6-parameter method.

4. How realistic is to obtain all the physical information needed to calculate parameters Ws, Ds and Dm, especially in poorly-instrumented basins?

Because the soil and DEM data were global data, this framework for baseflow parameters estimation could be used widely.

5- The Authors fail to convey the advantages, if any, of the 3-parameter methodology over the 6-parameter (conventional) procedure.

The advantages of the 3-parameter methodology over the 6-parameter procedure were:

1) The other three parameters became more sensitive when the 3-parameter method was used.

2) Parameters and streamflow uncertainty was reduced with the 3-parameter method compared to the original 6-parameter approach.

The framework presented in this study was for predictions in ungauged basins. Under the 6-parameter method, the highest value of Nsc could be got no matter nearly whatever values of the three baseflow parameters were set, when calibration in gauged basins (Fig. 5, page 7045-7047). That was because of the equifinality. Therefore, it would be inaccurate for transferring baseflow parameters from gauged basins to ungauged basins. But, under 3-parameter method, the three baseflow parameters would be directly estimated by Eq. 8, 9, and 11. On the other hand, accompanied by the reduction of model parameters uncertainties, the streamflow uncertainties in ungauged basins would be reduced.

This part would be added to the discussion part of the paper.

6- Figure 6 needs to be explained in detailed or eliminated from the paper. How the authors obtained it?

The previous Fig. 6 was replaced by a simulation result, i.e., comparison of extents of d_2 parameter spaces under two parameters setting methodologies in Minhe catchment when *Nsc* was higher than 0.6 (Fig. R3). When a threshold of goodness-of-fit was given (e.g., *Nsc*>0.6), the extent of parameters space meeting the conditions in the 3-parameter methodology would be smaller than that in 6-parameter methodology (extent 2<extent 1).

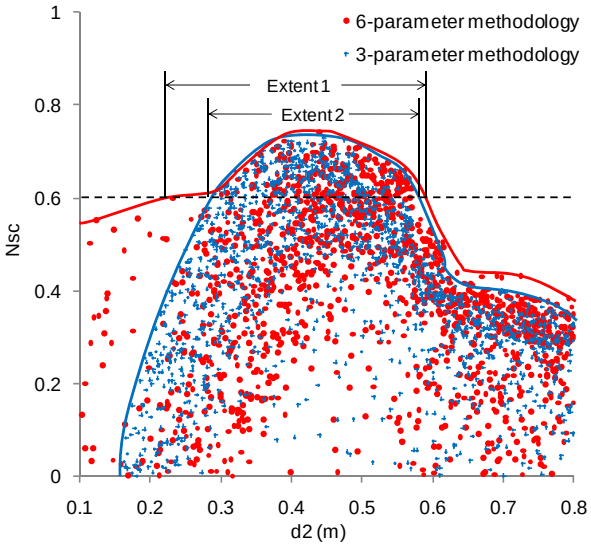


Fig. R3. The comparison of extents of d_2 parameter spaces under two parameters setting methodologies in Minhe catchment when *Nsc* was higher than a given threshold (e.g., 0.6). (extent 1: 6-parameter methodology; extent 2: 3-parameter methodology)

7- The improvement in parameter sensitivity is minimum, in cases negligible, for the 3-parameter method.

As shown in Fig. R3, when *Nsc*>0.6, the extent of parameters space of d_2 in the 3-parameter methodology was smaller than that in the 6-parameter methodology, i.e., extent 2<extent 1. That meant the parameter d_2 would be more sensitive in the 3-parameter methodology compared to it in the 6-parameter methodology.

Minor comments

1. The English needs some improvement but it is overall easy to read. -The paper is clear and flows well. - Summary Section should be Summary and Conclusions - I think Figure1 and 2 could be merged into one figure with two subplots. - I would considerer replacing the 3 parameters methodology by 3-parameter methodology.

The Summary will be revised as conclusion; Fig.2 and another added figure will be merged into one figure; “3 parameters methodology” will be replaced by “3-parameter methodology”.

2. Abstract, lines 1-2: Repetitive sentence. “Equifinality is unavoidable when transferring

model parameters from gauged catchments to ungauged catchments for predictions in ungauged basins (PUB).” Perhaps replace with: Equifinality is unavoidable when transferring model parameters from gauged catchments to ungauged catchments for hydrologic predictions.

The sentence has been revised.

3. Line 3: it should read: the Variable Infiltration Capacity (VIC) model Line 10: replace “Using the new parameters estimation approach, model parameters become more sensitive and the extent of parameters space will be smaller when a threshold of goodness-of-fit is given” with “Using the new parameter estimation approach, model parameters become more sensitive and the extent of parameters space is smaller when a threshold of goodness-of-fit is given.”

The sentence has been revised.

4. Line 15: it should read: ⋯ compared to the uncertainty given by the original calibration method”

The sentence has been revised.

5. Introduction Line 19: it should read: ⋯ is a macro-scale land surface model.

The sentence has been revised.

6. Page 7019: Line 4: remove, vice versa

The sentence has been revised.

7. Line 6: replace “by simulated” with “ with simulated”

The sentence has been revised.

8. Line 19: replace “with regression equations” with “using regression equations”. Line 19: remove: “However”

The sentence has been revised.

9. Line 22: remove: “Meanwhile”

The sentence has been revised.

10. Line 26: replace “Due to above” with “Due to the above”

The sentence has been revised.

11. Line 27: replace “when they are verified” with “when optimized parameters are applied”

The sentence has been revised.

12. Page 7020 Line 1: replace “have been more and more popularly” with “ have been widely”

The sentence has been revised.

13. Page 7021: replace " compare" with "compared"

The sentence has been revised.

14. Page 7022 Line 3: it should read: … is a macro-scale land surface model.

The sentence has been revised.

15. Line 8: replace "With refined describing of" with "Due to its refined description of"

The sentence has been revised.

16. Line 12: The authors may want to add an explanation of why those two objective functions were selected. Why Nsc and Re, what behavior in the model can they capture?

Nsc indicates how well the simulations fit the observations; and *Re* is a water balance criterion which indicates the relative error of simulations versus observations.

17. Page 7023 Line 8: This is no quite right; the VIC model is highly parameterized. There are twenty-one soil related parameters. The six model parameters you mentioned are the most widely parameters prompted to calibration. You need to rephrase this sentence.

The sentence has been revised.

18. Line 11: replace "The six parameters are calibrated by two objectives:" with "Two objective functions are used to measure the goodness of the fitting:"

The sentence has been revised.

19. Line 23: replace ",and has been applied in many researches" with "that has been applied in numerous studies"

The sentence has been revised.

20. Page 7023 Line 5: replace ",using Monte Carlo (MC) method." With ",using a Monte Carlo (MC) approach."

The sentence has been revised.

21. Line 10: Why to use the average of the Nsc and Re? It is a reason behind this? It seems to me that you are losing information by doing this average. Please explain.

There is not a widely used criteria considering both *Nsc* and *Re*. Despite in Eq 5 *Re* is multiplied by 100%, but in Eq 6, it is still used as decimal fraction and will not become to dominating. For example, if in one result, the *Nsc*=0.9, *Re*=15%, in another result, *Nsc*=0.89, *Re*=6%, it may regarded as the second result is better than the first result. The *Mnc* will be 0.875 and 0.915 in the first and second result, respectively. Therefore, in this section, the *Mnc* may be better than *Nsc*.

22. Line 15: Do you mean small a given threshold?

The threshold of GLUE estimation in this study is *Mnc*>0.6.

23. Line 16: replace "as "nonbehavioral" and is rejected" with " "nonbehavioral" and it is

rejected”

The sentence has been revised.

24. Line 18: Not clear the sentence: “That is defined as likelihood weight looked like probability, and is regarded as the posterior parameters probability distribution.” Do you mean: “ The likelihood weight is defined as probability, and it is regarded as the posterior parameter probability distribution.” ? Please clarify?

Due to the cumulative sum of 1, the rescaled likelihoods of the remaining parameters sets looks like probability; and it is regarded as the posterior parameters probability distribution.

25. Line 20: replace: “with likelihood” with “ with the likelihood”

The sentence has been revised.

26. Line 23: replace” “In addition of confidence interval, a quantitative estimator is used for uncertainty analysis” with “In addition to the confidence intervals, a quantitative estimator is used for the uncertainty analysis”

The sentence has been revised.

27. Page 7024 My main comment about this section is the lack of description of the methodology used to sample the parameters used in the analysis. Was it stratified sampling, was it Latin Hypercube, was it uniform sampling? How many Monte Carlo simulations were done? Based on Figure 5 the number of model simulations is quite small (perhaps 100). Nijssen and Lettenmaier (2004) used 1000 MC simulations of the VIC model at a large scale (4500 km2) spanning 6 years at the daily time step (4500 time steps). Demaria et al., 2007 used _50,000 simulations for a lumped model in a small-scale basin at a daily time step. Line 14: Equation 6. What was the reason to get an average of the Nsc and Re? The advantage of using two objective functions instead of one is to be able to capture the different model responses, i.e., mass balance versus peaks or low flows. Please explain.

Some detailed information about the Monte Carlo simulation is expressed in major comments 1. When compared the streamflow simulations, the two objective functions (*Nsc* and *Re*) are both used. But only one criterion could be used in GLUE methodology. Hence, *Mnc* was used in uncertainty estimation.

28. Page 7025 Line 5: Your statement: “In VIC model, the three baseflow parameters (Ws, Ds, and Dm) are less sensitive than other three parameters (Demaria et al., 2007)”. Demaria et al., (2007) used a slightly different implementation of the baseflow formulation introduced by Nijssen et al., 2001 (see Nijssen, B., G. M. O’Donnell, D. P. Lettenmaier, D. Lohmann, and E. F. Wood, 2001, Predicting the discharge of global rivers, J. Clim., 14, 3307–3323.). Although this implementation has in principle the same equation, the parameters are different. I think you should mention to avoid confusion to future readers.

The parameter sensitivity will be referenced to another paper.

29. Page 7026 Line 15: replace “and are different in different sub-grid.“ with “and are different in different sub-grids.”

The sentence has been revised.

30. Line 15: it is not clear what you mean by “But using calibration methodology, parameters will be set as same value in the whole catchment. Therefore, using this framework, baseflow parameters will be distributed and more relatively authentic.” Does it mean in the calibration procedure you a-priori parameter values will be the same for each grid cell? What do you mean by “Therefore,: …). Please clarify.

Under calibrations conditions, model parameters value will be same for each grid cell to avoid over parameterization. But under this framework, the three baseflow parameters could be estimated in each grid with Eq. 8-11. The parameters will be different in different grid, and be more physically distributed.

31. Line 20: why 24 catchments were used, it seems to me that selecting a large number of basins detriments the quality of the MC sampling, i.e., fewer parameter sampling. Wouldn’t it be more beneficial to use one basin from each represented climate as in (Demaria et al., 2007; Van Werkhoven, et al.,2009. Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. Advances in Water Resources, 32(8), 1154-1169.) ?

The 24 catchments under different hydro-climatic conditions in China were selected to study PUB. The number of Monte Carlo simulations in three represented catchments (Gaoqitou, Taolinkou, and Minhe catchment) was increased to 2000.

32. Page 7027 Line 6: replace “Most available streamflow data are more than 20 yr.” with “Most available streamflow data are archived for at least 20 years.”

The sentence has been revised.

33. Line 14: replace “One is estimating all six parameters through calibration, called 6 parameters methodology. Another one is estimating three baseflow parameters by physical properties of soil and topography, and the remaining three parameters are calibrated, called 3 parameters methodology.” with “The first one consists in estimating One is estimating all six parameters through calibration, called 6 parameters methodology. The second one estimates three baseflow parameters using the physical properties of soil and topography, and the remaining three parameters are calibrated, called 3 parameters methodology.”

The sentence has been revised.

34. Line 15: what are the values of the parameters (Dm, Ws and Ds) that were estimated using physical properties? You need to include them in the paper. Line 15: how the parameter values computed with equations 8, 9 and 11 compare to the ones obtained through the sensitivity analysis? I would be nice if the Authors include the values in Figure 6 for reference (with a star for example).

The values of the parameters estimated with Eq. 8-11 are shown in figure R2. (Major comments 2 and 6)

35. Line 18: I think three bar plots showing for each catchment and each objective function

the model performance will be a better way to show the results of table 3. You can do 1 figure with 3 subplots.

Table 3 will be replaced by a figure.

36. Line 20: replace “No matter for 6 parameters methodology or 3 parameters methodology,” with “Regardless of the method used: 6 parameters methodology or 3 parameters methodology,”

The sentence has been revised.

37. Line 20: By looking at Table 2 and 3 it is difficult to know what basins are humid and which ones are arid. I recommend adding an additional column in Table 2 showing the Dryness Index (Ep/P). In table 3 it would be useful to order the basins based on the Dryness Index from drier to wetter.

Dryness index (E_p/P) will be added to Table 2. The 24 catchments are ordered by China’s “Hydrological Year Book”.

38. Line 20 to 27: What can be driving the differences in Nsc and RE in the Haihe and Yellow river? Is it climate, is it the size of the basin? Please explain.

Haihe River and Yellow River are located in semi-humid and semi-arid regions. The annual runoff coefficient index in some catchments are lower than 0.2. It is difficult for hydrologic simulation in these regions. The *Nsc* will be lower and absolute *Re* will be higher than it in other humid catchments.

39. Page 7028 Line 8: replace “The model parameters sensitivity is estimated by MC simulation, and the results in three kinds of hydro-climatic catchments: Gaoqitou, Taolinkou, and Minhe catchment, are illustrated in Fig. 5. ” with “The model parameters sensitivity is estimated using a MC approach, and the results for three different of hydro-climatic environments: Gaoqitou, Taolinkou, and Minhe catchment, are illustrated in Fig. 5.”

The sentence has been revised.

40. Line 10: what do you mean by : “ i.e., the model can perform best within an extensive range of parameters space”, please explain

The model can be optimized at a wide range of parameters spaces for the three baseflow parameters and parameter d_3. For example, in Gaoqitou catchment, the highest value for *Nsc* can be obtained no matter what value of parameter W_s is set during (0,1) (Fig. R1 (a)).

41. Line 10-11: The Authors need to explain what a Sensitive parameter means? How the reader should interpret Figure 5? Am I looking to a maximum or a minimum for Sensitivity? Please explain.

Sensitive parameter means a small change of parameter value could impact the *Nsc* value significantly. Non-sensitive parameter means *Nsc* value will not nearly be influenced by the parameter value.

42. Line 12: Why parameter b is more sensitive in an arid catchment?

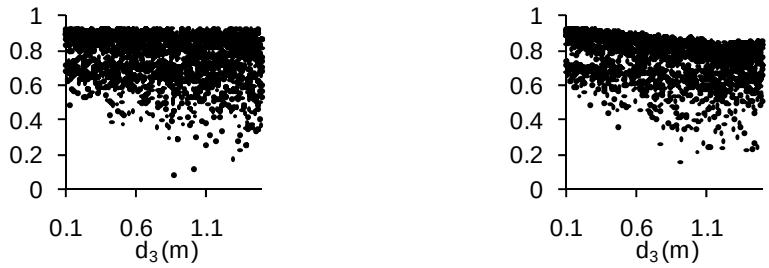
In humid catchment, Dunne runoff generation mechanism is the dominate one; otherwise, Harton runoff generation mechanism is the dominate one in arid regions. In arid regions, soil is difficult to be saturated. The shape of soil moisture capacity curvy could control how much runoff will be generated. Then parameter *b* is more sensitive in an arid catchment.

43. Line 13: Why parameter d2 is the most sensitive one? Please explain

Parameter d_2 controls how much water could be stored in the top two soil layers, and directly influences the amount of surface runoff generation. Therefore, this parameter is the most sensitive one.

44. Line 18: I cannot see the increase in sensitivity in parameter d3 for the 3-parameter method. Please explain. The authors may want to revisit (Wagener, T., D. P. Boyle, M. J. Lees, H. S. Wheater, H. V. Gupta, and S. Sorooshian (2001), A framework for development and application of hydrological models, Hydrol. Earth Syst. Sci., 5(1), 13– 26) for details.

For example, in Gqoqitou catchment, under 6-parameter methodology, highest Nsc value can be obtained, no matter what value of parameter d_3 is set (Fig. R4 (a)). But under 3-parameter methodology, highest value of *Nsc* can be obtained when parameter d3 is near 0.1 (Fig. R4 (b)).



(a) 6-parameter methodology (b) 3-parameter methodology

Fig. R4. Comparison of sensitivity of parameter d_3.

45. Line 19: what do you mean by: Meantime, some original sensitive parameters become more sensitive. What parameters? Please explain.

Here, the parameters are parameter *b* and d_2. The detail is expressed in major comments 6.

46. Line 20: What are you showing in Figure 6? Are these simulated values? Do you think the differences between Extent 1 and Extent2 are statistically significant? Please explain the figure or remove it.

As same as major comments 6.

47. Line 25: How did you choose the threshold Mnc equal to 0.6? You need to use a different color or shading for the humid and arid basins in Figure 7. It is not intuitive which is the humid basin and which is the arid one.

Generally, in Glue methodology, *Nsc*>0.6 is used as a threshold. Here, *Mnc*>0.6 is also used as a threshold.

48. Page 7029 Line 4-5: the differences in variances between the 3 and 6-paremeter method is almost negligible.

The variance will be replaced by box plot. See details in No.51.

49. Line 23: It would be nice to see the relative change between the 6 and 3-parameters in a third row. For example for Gaoqitou basin: (310.11-336.39)/336.39*100 = -7.8124% decrease with respect to the 6-parameter method.

The relative change will be added.

50. Line 24: same comment as above.

As same as No. 49.

51. Line 4-9. I am concern the differences in the variances are coming from using samples with dissimilar lengths. For example parameter d2 shows more well-behaved simulations in the Humid basin than in the Arid basin, hence the variance in the former will be larger due to insensitivity of this parameter. I think you need to weight the variance by the length of the sample or find another way to measure the variability

The variance of parameter b and d_2 is replaced by a box plot (Fig. R5). The average variability (the maximum value minus minimum value) of *b* parameter was 0.803 and 0.783 using the 6-parameter methodology and 3-parameter methodology, respectively. There were only 5 out of 24 catchments, in which the variability (the maximum value minus minimum value) of *b* parameter using the 3-parameter methodology was higher than that using the 6-parameter methodology. The average variability (the 75^{th} value minus 25^{th} value) of d_2 parameter was 0.526 and 0.512 using the 6-parameter methodology and 3-parameter methodology, respectively. There were only 8 out of 24 catchments, in which the variability (the 75^{th} value minus 25^{th} value) of d_2 parameter using the 3-parameter methodology was higher than that using the 6-parameter methodology.

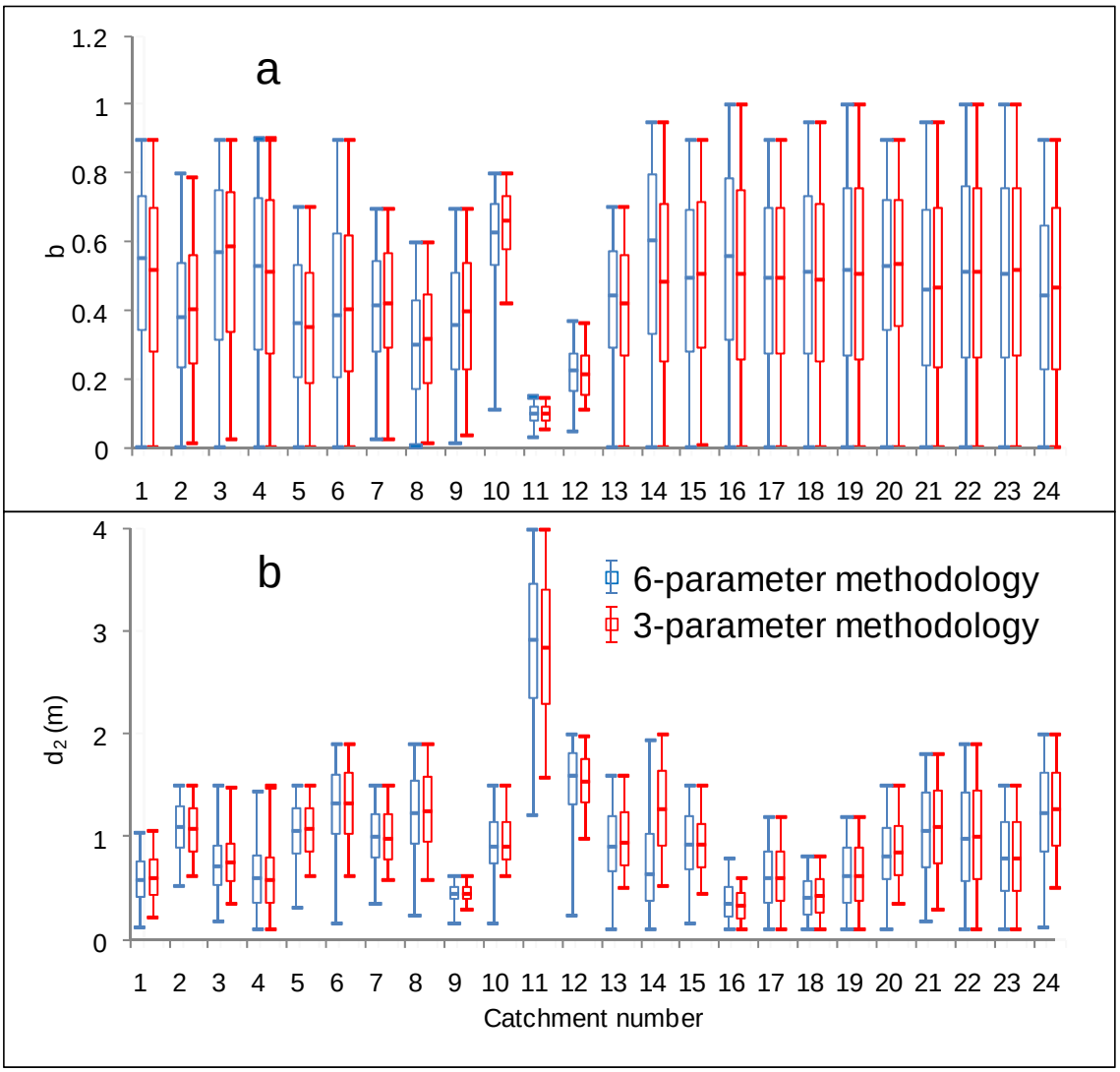


Fig. R5. Comparison of the variability of parameter *b* (a) and d_2 (b) under two kinds of parameters setting methodologies, when *Mnc* is higher than the threshold. The boxes indicated the 25th and 75th, percentiles; the whiskers indicated the lowest and highest data value; and "-" indicated the 50th percentiles value.

52. Page 7030 Line 1: Shoudn't it read Summary and Conclusions?

The sentence has been revised.

53. Line 4-5: It is not clear what the Authors mean by: Therefore, the equifinality of the three baseflow parameters is higher than other three parameters.

That means the interdependence of the three baseflow parameters is higher than other parameters.

54. Page 7031 Line 3-11: I wonder what is more computationally and time efficient, the 3-parameter or the 6-parameter methodology? It seems to me that obtaining the parameter values with equations 8, 9 and 11 can be troublesome and time consuming especially in poorly-instrumented basins. I am not convinced that the 3-parameter methodology offers any

advantage over the 6-parameter (conventional) procedure because: 1- Improvements in parameter sensitivity is minimum as seen in Figure 6 and 7. 2- Table 4 shows that the accuracy gain is quite small (-7.8124 % for the Gaoqitou basin with respect to the 6-parameter methodology for example) which is smaller than the error in a rating curve (~25%). 3- The number of MC simulations seem to be to small to accurately capture the sensitivity of the parameters.

Time cost is expressed in major comments 3. 2. Figure 6 and 7 has been revised (major comments 6 and minor comments 51). 3. Overall, the uncertainty of streamflow simulation in the 3-parameter methodology is lower than that in the 6-parameter methodology. 4. The number of Monte Carlo simulations has been increased (major comments 1).