Responses to Referee #1:

This paper analyzes the prediction performance of a lumped hydrological model using different time and spatial dependent parametrizations of one of its parameters. There are several errors in the paper and points that should be explained better and I have a major concern regarding the results.

Comment on the results:

A1: The value of omega looks strange to me. Assuming that the equation 1 you wrote is correct (and therefore it is a frequency and not a phase) and that the order of magnitude of omega is of hundreds (like shown in figures 8 and 9), this mean that your parameter theta1 oscillates hundreds of times per time step. This looks unreal to me since the goal of having time-variant parameters is to represent long term (seasonal) oscillations. Therefore, either there is a problem with the unit of omega or your model is not doing what it was meant for. If omega is a phase (meaning theta1 = alpha + beta*sin(t + omega)) the value of omega makes more sense but theta1 would still complete an oscillations every 6.28 time steps (the time step is days, right?). Don't you also have a frequency that multiplies "t" and have a small value?

Reply:

(1) We apologize for our mistakes. ω represents frequency rather than phase. It will be revised accordingly in the revised manuscript.

(2) We have carefully checked the results of regression parameter ω and found that the Figures 8 and 9 in the manuscript of ω should be modified as the attachmments:

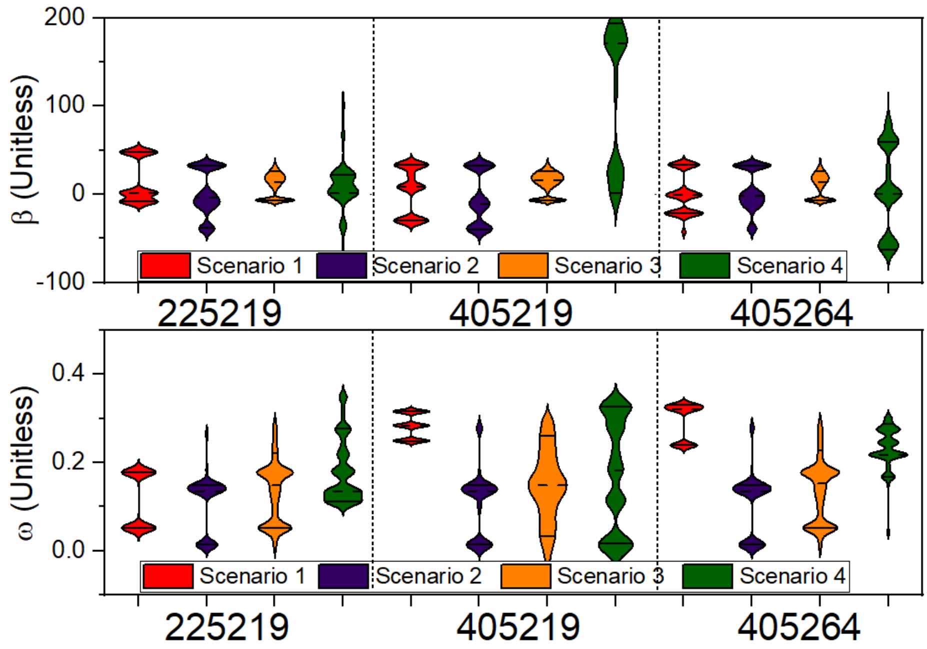

Figure 8. Posterior distributions of the regression parameters (β and ω) for the production storage capacity (θ_1) for the four modeling scenarios in all the 3 studied catchments. In this figure, parameters were calibrated in the non-dry period while verified in the dry period. The solid horizontal lines within the violin plots denote the 25th and 75th percentiles of the posterior distribution, while the dash line denotes median estimates.

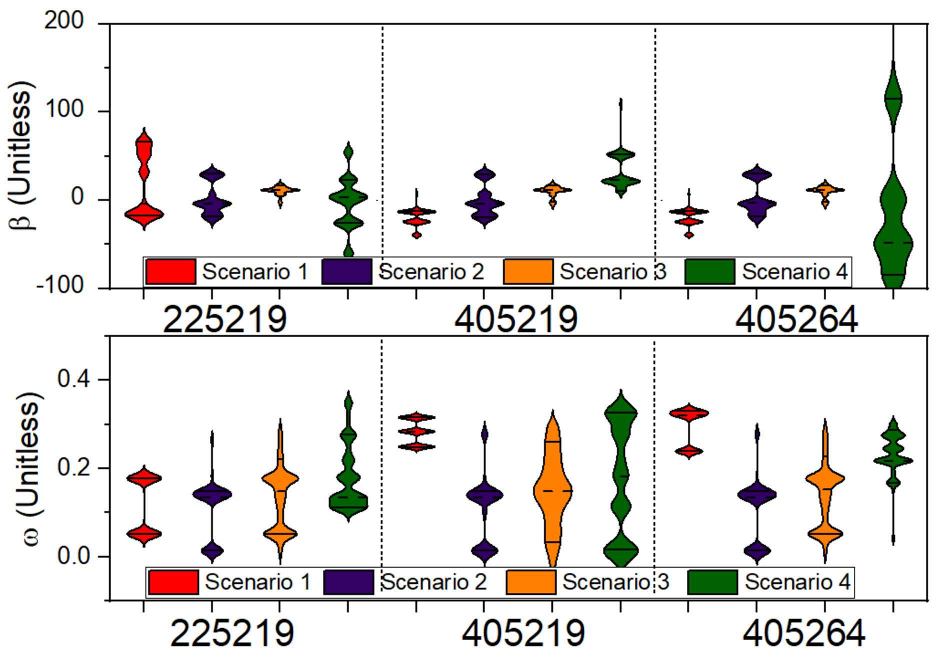

Figure 9. Posterior distributions of the regression parameters (β and ω) for the production storage capacity (θ_1) for the four model scenarios in all 3 studied catchments. In this figure, parameters were calibrated in the dry period while verified in the non-dry period. The solid horizontal lines within the violin plots denote the 25th and 75th

percentiles of the posterior distribution, while the dash line denotes median estimates.

For the first four scenarios as shown in Figure 8, the average median estimates of regression parameter ω of the 3 catchments are 0.24, 0.14, 0.15, and 0.18, respectively., and that in Figure 9 are 0.15, 0.26, 0.23, and 0.17 respectively in Figure 9. Thus, the phase of the sine term could be derived based on the regression parameter ω. The mean phase of model parameter θ_1 for each scenario is 26.2, 46.3, 41.9 and 35.2 in Figure 8, respectively. It is 42.9, 24.1, 27.4 and 38.0 in Figure 9, respectively.

Detailed comments:

A2: line 102-103: There is not a clear definition of pooling, complete pooling and hierarchical Bayesian. I would explain shortly what do they mean and which are the differences since then the paper only writes about hierarchical Bayesian.

Reply: Thank you for your comments. The following explanations (in blue) about the pooling, complete pooling and hierarchical Bayesian will be added in the revised manuscript.

In general, there are three methods to consider the spatial coherence between different catchments in parameter estimation. The first one is no pooling, which means every catchment is modeled independently, and all parameters are catchment-specific. The second one is complete pooling, which means parameters are considered to be common across all catchments. The third/last one is hierarchical Bayesian (HB) framework, also known as partial pooling, which means some parameters are allowed to vary by catchments and some parameters are assumed to be drown from a common hyper-distribution across the region that consists of different catchments.

A3: line 152-153: It would be beneficial to explain shortly how the method works even if it was already used in other studies.

Reply: Thank you for your comment. Definition of dry period is explained in the following paragraph and will be added in the revised manuscript:

Saft et al. (2015) tested several algorithms for dry period delineation, which considered different combinations of dry run length, dry run anomaly and various boundary criteria, and found that the identification results of dry period by one of the algorithms showed marginal dependence on the algorithm and the main results were robust to different algorithms. The detailed processes could be found on Saft et al. (2015) and also are as follows.

Firstly, the annual rainfall data were calculated relative to the annual mean, and the anomaly series was divided by the mean annual rainfall and smoothed with a 3 year moving window. Secondly, the first year of the drought remained the start of the first 3 year negative anomaly period. Thirdly, the exact end date of the dry period was determined through analysis of the unsmoothed anomaly data from the last negative 3 year anomaly. The end year was identified as the last year of this 3 year period unless: (i) there was a year with a positive anomaly >15% of the mean, in which case the end year is set to the year prior to that year; or (ii) if the last two years have slightly positive anomalies (but each <15% of the mean), in which case the end year is set to the first year of positive anomaly; (iii) To ensure that the dry periods are sufficiently long and severe, in the subsequent analysis, the author use dry periods with the following characteristics: length≥ 7 years; mean dry period anomaly<25%.

A4: line 159: Maybe it is more appropriate to use "cross validation" instead. I suggest to avoid making a paragraph with just one sentence and remove paragraphs 2.1.1 and 2.1.2 putting all together in section 2.1.

Reply: Thanks.

(1) Follow the Referee's comment, the phrase "Verification method will be modified as "Cross validation".

(2) Follow the Referee's suggestion, paragraph 2.1.1 and 2.1.2 will be put together in section 2.1, and the sub-titles of section 2.1.1 and 2.1.2 will be deleted in the revised manuscript.

A5: chapter 2.3: It is not clear to me what do you do with the other parameters of the GR4J model (theta2, theta3, theta4). Do you keep them fixed or do you sample them? What is their effect on the final result?

Reply: Thank you for your comment.

(1) All other model parameters (θ_2, θ_3, and θ_4, except θ_1) are not fixed, but sampled simultaneously with regression parameter α, β and ω (if present), and hyper-parameters μ_2, σ_2, μ_3 and σ_3 in the SCEM-UA algorithm. In actual calculation process, we would set a large variation interval for each unknown quantity first, parameters would converge to a small interval in MCMC calculation process, the final parameter samples that satisfy the requirement that a GR value must be smaller than a Gelman-Rubin convergence value of 1.2 (Gelman et al., 2013) would be selected as the posterior probability distribution of parameters. More information will be added in the revised manuscript.

(2) Previous studies on GR4J model showed that θ_2, θ_3, and θ_4 are less sensitive than θ_1 under changing climate (Perrin et al., 2003;Renard et al., 2011;Westra et al., 2014). Therefore, we think that it is reasonable to assume that θ_1 is time-varying while other model parameters are temporal invariant.

A6: line 199: The equation is different from the ones reported in Table 1.

Reply: We apologize for our mistakes. The fault equations in Table 1 have been revised as equation 1 in the revised manuscript.

A7: line 201: You write that omega is the phase while in the equation 1 it is a frequency.

Reply: Thank you for pointing out this mistake. The ω represents the frequency rather than the phase (see response to comment A1). The statement in line 201 is wrong and will be modified in the revised manuscript.

A8: line 202: The combination alpha=beta=omega=0 makes theta 1 to be equal to 0, that indeed it is a constant value but probably it is not what you want.

Reply: Thanks. According to the definition of the GR4J model (Perrin et al., 2003), θ_1 represents the primary storage of water in the catchment and must be a positive value. Thus, in the first four scenarios, in order to avoid this situation (θ_1=0), the combination of α=β=ω=0 would be excluded first, and other combinations that made θ_1 equal to zero would be excluded too.

A9: chapter 2.3.2: What happens to alpha? You don't write about it anymore in the rest of the paper. Do you keep it fixed or do you sample also it? What is its effect on the final result?

Reply: Thanks.

(1) The α represents the constant term in equation 1. Changes in α lead to consistent changes in θ_1 across the whole time series, which doesn't result in temporal variations of model parameter θ_1. In addition, one objective of this study is to explore the potential temporal variation of θ_1; thus, the regression parameter α is not our focus.

(2) Regression parameter α is not fixed in advance but is sampled as same as

other unknown quantities. The posterior distribution of α is derived out simultaneously with hyper-parameters μ_2, μ_3, σ_2 and σ_3, other regression parameters β and ω (if present), and model parameters θ_2, θ_3 and θ_4 in the SCEM-UA algorithm.

A10: chapter 2.3.2: It is not clear to me if linking the parameters between catchments means sampling them from the same Gaussian distribution or there is another form of linking.

Reply: We apologize for the misunderstanding. The link is that regression parameter β (ω) of different catchments is assumed to sample their values in the same Gaussian distribution. This kind of links have been widely used in the field of extreme event analysis, such as Sun et al (2015, 2016), Lima et al (2009) and Bracken et al (2018).

A11: chapter 2.3.2: How do you sample omega and beta when they are not linked?

Reply: Thanks. The ω is not linked in scenario 1, while β is not linked in scenario 2. In scenario 4, both ω and β are not linked. Spatially irrelevant parameters would be sampled and derived as independent variables. For example, in scenario 4, the ω and β of different catchments are not linked, thus values of ω and β of each catchment are calibrated from corresponding catchment inputs. In scenario 1, regression parameter $\beta(c) = N(\mu_3, \sigma^2)$, which means that β is shared with linked catchments, while independent regression parameters $\omega_{1\text{-}1}$, $\omega_{1\text{-}2}$, and $\omega_{1\text{-}3}$ are used to represent the frequency of model parameter θ_1 in different catchments. The name of all unknown quantities in different scenarios could be found in the supplementary material (at the end of this reply) , and these tables will be added in the revised manuscript.

The prior ranges of all unknown quantities in different scenarios have been added in the supplementary material.

A12: line 218: How do you choose the values of mu and sigma, the hyper-parameters of your model?

Reply: Thanks. The posterior distributions of all unknown quantities, including model parameters θ_2, θ_3 and θ_4, and regression parameters α, β and ω, and hyper-parameters μ_2, μ_3, σ_2 and σ_3 are derived simultaneously through the SCEM-UA algorithm. In actual calculation process, we would set a large variation interval for each unknown quantity first, parameters would converge to a small interval in MCMC calculation process, the final parameter samples that satisfy the requirement that a GR value must be smaller than a Gelman-Rubin convergence value of 1.2 (Gelman et al., 2013) would be selected as the posterior probability distribution of parameters.

A13: chapter 2.4.1: I wouldn't call "likelihood function" what actually is an objective function.

Reply: Thanks. As suggested, the "likelihood function" will be modified as "objective function" in the revised manuscript.

A14: line 250: You are mixing an objective function with a prior distribution of the parameters. How do you account for the prior distribution of the parameters when they are not linked?

Reply: Thanks.

(1) The objective function of Eq.1 will be modified as follows:

$$\varepsilon_c\left[\theta_1,\theta_2,\theta_3,\theta_4\right] = \text{-}RMSE\left[\sqrt{Q}\right]\left(1+\left|1+BIAS\right|\right)$$

where $\theta_1,\theta_2,\theta_3,\theta_4$ refer to four model parameters.

(2) The objective function of Eq.5 will be modified as follows:

$$Scenario\ 1:\ \Lambda=\prod_{c=1}^{C}\varepsilon_c\left[\theta_1(t,c),\theta_2(c),\theta_3(c),\theta_4(c)\middle|\alpha(c),\beta,\omega(c)\right]\bullet f_N\left(\beta\middle|\mu_2,\sigma_2\right)$$

$$Scenario\ 2:\ \Lambda=\prod_{c=1}^{C}\varepsilon_c\left[\theta_1(t,c),\theta_2(c),\theta_3(c),\theta_4(c)\middle|\alpha(c),\beta(c),\omega\right]\bullet f_N\left(\omega\middle|\mu_3,\sigma_3\right)$$

$$Scenario\ 3:\ \Lambda=\prod_{c=1}^{C}\varepsilon_c\left[\theta_1(t,c),\theta_2(c),\theta_3(c),\theta_4(c)\middle|\alpha(c),\beta,\omega\right]\bullet\prod_{n=1}^{2}f_N\left(\beta,\omega\middle|\mu_2,\sigma_2,\mu_3,\sigma_3\right)$$

$$Scenario\ 4:\ \Lambda=\prod_{c=1}^{C}\varepsilon_c\left[\theta_1(t,c),\theta_2(c),\theta_3(c),\theta_4(c)\right]$$

$$Scenario\ 5:\ \Lambda=\prod_{c=1}^{C}\varepsilon_c\left[\theta_1(c),\theta_2(c),\theta_3(c),\theta_4(c)\right]$$

where the number of catchments in the region is represented by C; *c* represents the specific catchment; the *t* is the time step.

A15: chapter 2.4.2: You don't say which settings of the sampling method you use (e.g. how many parameters you sample. . .)

Reply: Thanks. The sampling method used in this paper is the SCEM-UA algorithm. The detailed description of the settings of SCEM-UA algorithm will be added in the revised manuscript:

(1) Convergence is assessed by evolving three parallel chains with 30000 random samples, while verifying that the posterior distribution of parameters results in a value smaller than a Gelman-Rubin convergence value of 1.2 (Gelman et al., 2013).

(2) The number of unknown quantities in different scenarios are as follows: 15 in scenario 1 and scenario 2, 13 in scenario 3 and 18 in scenario 4.

A16: chapter 3.2.1: The dataset that you get is unbalanced, since there are more wet years. Is it taken into account? Does it have an effect on the calibration?

Reply: Thank you for pointing out this situation.

(1) Generally, calibration data should be longer than 3-6 years for daily hydrological modeling in order to get robust results (Perrin et al., 2003, Coron et al.,

2012). Thus, data from both dry period (15 years) and wet period (21 years) were used for model calibration to meet this requirement.

(2) Generally, a longer time series may improve the robustness of hydrological predictions. However, we tested the calibration performance with different lengths of records (> 10 years) in dry and non-dry periods and found that their results are almost the same. Therefore, we used both the length of 15 years of dry and 10 years non-dry periods into calibration in order to utilize all available data.

A17: chapter 3.2.3: Figures 7 and 8 are actually 8 and 9.

Reply: Thanks. Changes will be made as suggested.

A18: Figures 5, 6, 8, 9: Since you want to show a probability distribution I wouldn't use a boxplot but, instead, I suggest to use a violin plot (e.g.https://seaborn.pydata.org/examples/grouped_violinplots.html)

Reply: Thank you for your suggestions.

(1) Figures 8 and 9 will be modified as violin plot in the revised manuscript, which also could be found in response to comment A1 by Referee #1**.**

(2) Figures 5 and 6 will be revised as violin plot in the revised manuscript, which also could be found as follows:

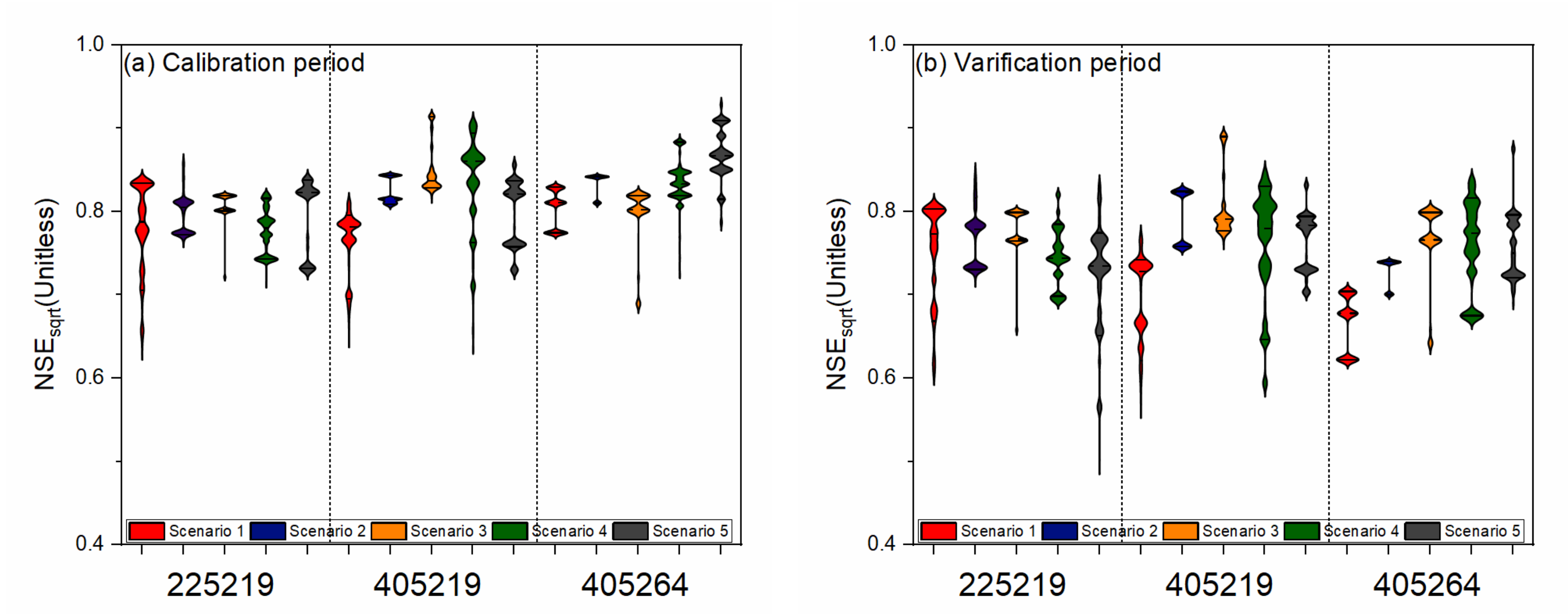

Figure 5. NSE$_{sqrt}$ for each of the five scenarios for each catchment during (a) the calibration period (non-dry period) and (b) the verification period (dry period).

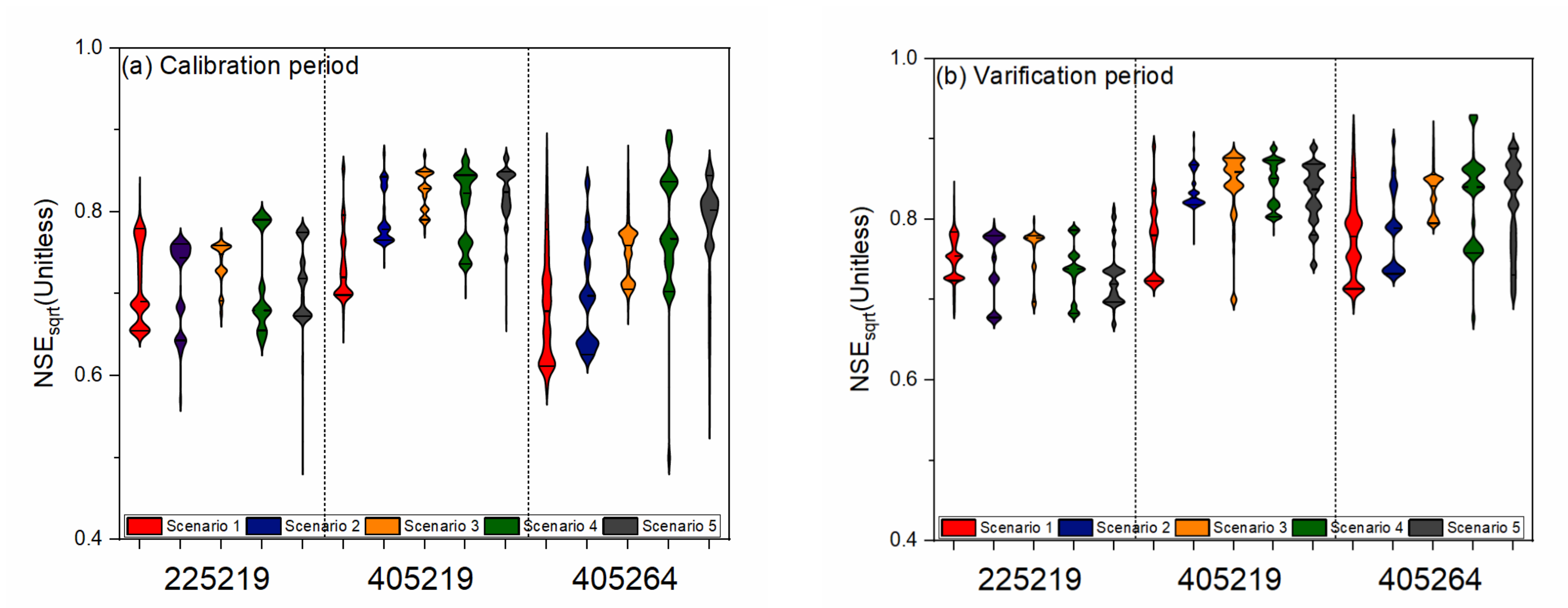

Figure 6. NSE$_{sqrt}$ for each of the five scenarios for each catchment during (a) the calibration period (dry period) and (b) the verification period (non-dry period).

A19: Figures 8, 9: Why do you change the colors between beta and omega? This makes the plot more difficult to read.

Reply: Thanks. The same color will be used to the same parameter consistently in all figures. Changes will be made as suggested in the revised figures. Please refer to response to comment A1 by Referee #1.

Supplement:

Table S1 The prior ranges of all unknown quantities in different scenarios

(1) Calibration in non-dry period and verification in dry period:

Scenario 1**:**

θ_{2-1}	θ_{2-1}	θ_{2-3}	μ_2	σ_3	θ_{3-1}	θ_{4-1}	α_{1-1}	ω_{1-1}	θ_{3-2}	θ_{4-2}	α_{1-2}	ω_{1-2}	θ_{3-3}	θ_{4-3}	α_{1-3}	ω_{1-3}
-10	-10	-10	-100	0	0.1	1	1	0.0001	0.1	0.5	100	0.0001	0.1	0.1	1	0.0001
10	10	10	100	6	200	10	600	0.4	300	20	1000	0.4	300	20	500	0.4

Scenario 2:

θ_{2-1}	θ_{2-1}	θ_{2-3}	μ_3	σ_3	θ_{3-1}	θ_{4-1}	α_{1-1}	β_{1-1}	θ_{3-2}	θ_{4-2}	α_{1-2}	β_{1-2}	θ_{3-3}	θ_{4-3}	α_{1-3}	β_{1-3}
-6	-6	-6	-0.4	0	1	0.5	1	-300	1	0.1	100	-300	0.1	2	1	-200
-6	-6	-6	0.4	0.1	500	10	600	300	300	20	600	500	400	20	800	300

Scenario 3:

θ_{2-1}	θ_{2-1}	θ_{2-3}	μ_2	σ_2	μ_3	σ_3	θ_{3-1}	θ_{4-1}	α_{1-1}	θ_{3-2}	θ_{4-2}	α_{1-2}	θ_{3-3}	θ_{4-3}	α_{1-3}
-5	-5	-5	-200	0	-0	0	1	0.5	1	1	0.1	100	1	0.5	100
5	5	5	100	8	0.4	0.1	120	10	500	300	20	500	250	20	600

Scenario 4:

θ_{2-1}	θ_{3-1}	θ_{4-1}	α_{1-1}	β_{1-1}	ω_{1-1}	θ_{2-2}	θ_{3-2}	θ_{4-2}	α_{1-2}	β_{1-2}	ω_{1-2}	θ_{2-3}	θ_{3-3}	θ_{4-3}	α_{1-3}	β_{1-3}	ω_{1-3}
-10	1	0.1	1	-300	0.0001	-10	1	0.1	0	-300	0	-10	1	0.1	0	-300	0.0001
10	500	10	800	300	0.4	10	500	10	800	300	0.4	10	500	10	800	300	0.4

(2) Calibration in dry period and verification in dry period:

Scenario 1**:**

θ_{2-1}	θ_{2-2}	θ_{2-3}	μ_2	σ_2	θ_{3-1}	θ_{4-1}	α_{1-1}	ω_{1-1}	θ_{3-2}	θ_{4-2}	α_{1-2}	ω_{1-2}	θ_{3-3}	θ_{4-3}	α_{1-3}	ω_{1-3}
-10	-10	-10	-60	0	1	0.5	1	0	1	0.5	1	0	1	0.1	1	0
10	10	10	60	6	300	10	600	0.4	300	20	600	0.4	300	15	600	0.4

Scenario 2:

θ_{2-1}	θ_{2-2}	θ_{2-3}	μ_3	σ_3	θ_{3-1}	θ_{4-1}	α_{1-1}	β_{1-1}	θ_{3-2}	θ_{4-2}	α_{1-2}	β_{1-2}	θ_{3-3}	θ_{4-3}	α_{1-3}	β_{1-3}
-10	-10	-10	0.0001	0	1	0.5	1	-300	1	0.1	1	-400	0.1	0.5	1	-400
10	10	10	0.4	0.1	200	15	500	400	300	20	600	500	140	20	600	400

Scenario 3:

θ_{2-1}	θ_{2-2}	θ_{2-3}	μ_2	σ_2	μ_3	σ_3	θ_{3-1}	θ_{4-1}	α_{1-1}	θ_{3-2}	θ_{4-2}	α_{1-2}	θ_{3-3}	θ_{4-3}	α_{1-3}
-10	-10	-10	-80	0	0	0	1	0.5	1	1	0.1	1	1	0.1	1
10	10	10	80	6	0	0.1	200	10	500	400	20	600	400	20	600

Scenario 4:

θ_{2-1}	θ_{3-1}	θ_{4-1}	α_{1-1}	β_{1-1}	ω_{1-1}	θ_{2-2}	θ_{3-2}	θ_{4-2}	α_{1-2}	β_{1-2}	ω_{1-2}	θ_{2-3}	θ_{3-3}	θ_{4-3}	α_{1-3}	β_{1-3}	ω_{1-3}
-10	1	0.1	1	-300	0.0001	-10	1	0.1	1	-300	0	-10	1	0.1	1	-300	0
10	500	10	800	300	0.4	10	500	10	800	300	0.4	10	500	10	800	300	0.4

Notes:

θ_{2-1}, θ_{2-2} and θ_{2-3} refers to model parameter θ_2 in catchment 225219, 405219 and 405264, respectively; θ_{3-1}, θ_{3-2} and θ_{3-3} refer to model parameter θ_3 in catchment 225219, 405219 and 405264, respectively; θ_{4-1}, θ_{4-2} and θ_{4-3} refers to model parameter θ_4 in catchment 225219, 405219 and 405264, respectively; μ_2, σ_2, μ_3 and σ_3 represent four hyper-parameters; α_{1-1}, α_{1-2} and α_{1-3} refer to regression parameter α in catchment 225219, 405219 and

405264, respectively; $\beta_{1\text{-}1}$, $\beta_{1\text{-}2}$ and $\beta_{1\text{-}3}$ refer to regression parameter β in catchment 225219, 405219 and 405264, respectively; $\omega_{1\text{-}1}$, $\omega_{1\text{-}2}$ and $\omega_{1\text{-}3}$ refer to regression parameter ω in catchment 225219, 405219 and 405264, respectively.

$$\varepsilon_c\left[\theta_1,\theta_2,\theta_3,\theta_4\right] = \text{-}RMSE\left[\sqrt{Q}\right]\left(1+\left|1+BIAS\right|\right)$$

$$\textit{Scenario 1}: \Lambda=\prod_{c=1}^{C}\varepsilon_c\left[\theta_1(t,c),\theta_2(c),\theta_3(c),\theta_4(c)\middle|\alpha(c),\beta,\omega(c)\right]\bullet f_N\left(\beta\middle|\mu_2,\sigma_2\right)$$

$$\textit{Scenario 2}: \Lambda=\prod_{c=1}^{C}\varepsilon_c\left[\theta_1(t,c),\theta_2(c),\theta_3(c),\theta_4(c)\middle|\alpha(c),\beta(c),\omega\right]\bullet f_N\left(\omega\middle|\mu_3,\sigma_3\right)$$

$$\textit{Scenario 3}: \Lambda=\prod_{c=1}^{C}\varepsilon_c\left[\theta_1(t,c),\theta_2(c),\theta_3(c),\theta_4(c)\middle|\alpha(c),\beta,\omega\right]\bullet\prod_{n=1}^{2} f_N\left(\beta,\omega\middle|\mu_2,\sigma_2,\mu_3,\sigma_3\right)$$

$$\textit{Scenario 4}: \Lambda=\prod_{c=1}^{C}\varepsilon_c\left[\theta_1(t,c),\theta_2(c),\theta_3(c),\theta_4(c)\right]$$

$$\textit{Scenario 5}: \Lambda=\prod_{c=1}^{C}\varepsilon_c\left[\theta_1(c),\theta_2(c),\theta_3(c),\theta_4(c)\right]$$