Responses to all the Referees:

Dear Editor and Reviewers:

We sincerely appreciate the comments from the Editor and Referees. Detailed responses to the comments raised by the Editor and Referees have been presented below. In the following Responses, there are several simplified labels. For example, A1 represents the comment 1 made by Reviewer #1, B1 represents the comment 1 made by Reviewer #2. Please also note that the page and line numbers mentioned in reviewers comments refer to the original version, while in the authors' response they refer to the revised version.

Sincerely,
Pan Liu, PhD, Professor
School of Water Resources and Hydropower
State Key Laboratory of Water Resources and Hydropower Engineering Science
Wuhan University, Wuhan, Hubei Province, 430072, P. R. China
E-mail: liupan@whu.edu.cn

Responses to Editor:

Although the reviewers are generally positive with the paper, there are still some points of concern. Please address the remaining comments thoroughly before resubmission.

Reply: We sincerely appreciate the comments from the Editor. Great efforts have been made to address both reviewers' comments.

Responses to Referee #1:

The authors generally addressed the concerns of the reviewers (I am referring here mostly to the concerns expressed by myself, Anonymous Reviewer 1) and the resulting paper represent an improvement of the previous version. Nevertheless, there are still some points to discuss:

A1: There are still some imprecisions in the naming of the characteristics of the sinusoidal functions. In the equation $y = a \sin(bx + c)$, a represent the amplitude, b the frequency, and c the phase. The period T can be calculated using the equation $\mathrm{T} = 2\pi/\mathrm{b}$. Please keep it in mind and check carefully the paper before submission since there are still some imprecision (e.g. line 523)

Reply: Thank you for your comments.

(1) Follow the reviewer's comment, the "regression parameter ω" and "regression parameter β" have been modified as "frequency ω" and "amplitude β" in the revised manuscript. Please refer to line 245, 249, 266 on page 12, line 529 on page 25, *etc*. in the revised manuscript.

(2) The "phase T" has been modified as "the period T". Please refer to line 548, 550, 551 on page 26.

A2. Since $\theta_1 = \alpha + \beta sin(\omega t)$ the combination expressed in line 224 α =β=ω=0 produces $\theta_1 = 0$ for any value of t. I am still convinced that, since θ_1 represent a storage in the model, setting it equal to 0 it is not a good choice. I think that the right values to express what you mean are β = 0, α = const and the value of ω becomes irrelevant.

Reply: We are really sorry for this error. According to the definition of the GR4J model (Perrin et al., 2003), the value of model parameter θ_1 should be larger than zero for any value of *t*. Therefore, this sentence has been modified as "According to the definition of the GR4J model (Perrin et al., 2003), the value of θ_1 must be a positive value. If model parameter θ_1 is constant then $\beta = 0$ and $\alpha > 0$ suffice in Eq.1 Meanwhile, the value of ω becomes irrelevant, thus the resulting model simplifies to a stationary hydrological model". Please refer to lines 224-228 on page 11 in the revised manuscript.

A3. I still have some doubts on the value of the parameter ω: as you write in the paper, your sinusoidal function oscillates with a period around 40 days: I think that the objective of using time varying parameters was to capture seasonality effects (e.g. high storage in winter and low in summer); is it what you want to represent with your model? If yes, I would expect the period to be around 365 days and not 10 times less.

Reply: Thank you.

(1) The purpose of introducing parameter ω in Eq.1 was to represent the periodical variation of model parameter θ_1, which might be monthly, seasonal, every half a year, or annual, etc. It should be noted that seasonality is only one of the potential time-varying schemes.

(2) Based on the results from the studied catchments, the mean periods of different scenarios are within 24~47 days, nearly every 0.8~1.5 month. In addition, we used the Hilbert-Huang Transform method (Huang et al., 1998) to identify the potential periods of the series of several climate variables (including the daily rainfall, daily potential evapotranspiration, daily maximum temperature and daily minimum temperature in the studied catchments). It was found that these series have periods of 22.2~49.1 days. Thus, we guess that the potential periods of these climate variables may be the possible reasons for the periods of time-varying parameters. It also should be mentioned that the adopted Hilbert spectrum method is one of the most popular methods for analyzing nonlinear and non-stationary data. Huang et al. (1999) indicated that this method is better than the Fourier transform method and Wavelet Transform method in processing nonlinear and non-stationary data. These analyses also have been added in the revised manuscript. Please refer to lines 552-562 on page 26.

Reference:

Huang, N. E., Shen, Z., Long, S. R., Wu, M. L. C., Shih, H. H., Zheng, Q. N., Yen, N. C., Tung, C. C., and Liu, H. H.: The empirical mode decomposition and the Hilbert

spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. A-Math. Phys. Eng. Sci., 454, 903-995, 10.1098/rspa.1998.0193, 1998.
Huang, N. E., Shen, Z., and Long, S. R.: A new view of nonlinear water waves: The Hilbert spectrum, Annu. Rev. Fluid Mech., 31, 417-457, 10.1146/annurev.fluid.31.1.417, 1999.

A4. Line 261: it would be beneficial to list what are the "unknown quantities"

Reply: Thank you. Unknown quantities consist of model parameters (θ_2, θ_3, and θ_4), regression parameters α, β and ω, and hyper-parameters μ_2, σ_2, μ_3 and σ_3. Details about the "unknown quantities" also have been added in the revised manuscript. Please refer to lines 271-277 on page 13.

A5. I think that overall the paper has made one step further towards publication. The only major concern that I have against it is verifying if the model is doing what it was meant to (see point 3).

Reply: Thank you for your comments.
(1) As illustrated on lines 211-218 in pages 10, regression parameter ω is used to account for the periodical variation (rather than seasonality) of model parameter θ_1. Based on the results from the studied catchments, the mean periods of are within 24~47 days, nearly every 0.8~1.5 month. Furthermore, the seasonality effect is not significant in these catchments.
(2) We agree with reviewer #1 that the seasonality effect (e.g. high storage in winter and low in summer) of model parameter θ_1 is a reasonable guess before calculation. However, our results did not prove this point. The reasons behind might be that since a daily hydrological model is adopted, the remarkable day-to-day variation in climate variables (e.g., daily rainfall) masked the seasonality of time-varying parameter.
(3) Explanations also have been made in response to comment A3, please refer to comment A3.

Responses to Referee #2:

De study of Pan et al. applies a hierarchical Bayesian framework in three Australian catchments. The HB-framework involves estimating the spatial and temporal coherence of model parameters by a regression equation. Five scenarios are tested in the study, with different degrees of spatial and temporal coherence. The authors conclude that the time-varying setting improved performance but increased uncertainty, spatial coherence reduced uncertainty and that performance decreased when parameters were transferred from dry periods to wet periods.

The article shows quite some improvements compared to the previous version of the manuscript. I am also happy that the authors addressed my previous comments and made improvements based on that.

B1: Therefore, I appreciate the effort of the authors to clarify their method, but, to be honest, I'm still a bit confused. It may be just me, and my lack of knowledge here, but I still wonder where the Gaussian distributions come in. The authors state that, in paragraph 2.4.2 and their response, that all parameters, including the regression parameters (L.296-298), are sampled simultaneously and come from a uniform distribution (L.310). So where are the Gaussian distributions coming in? Are the regression parameters not samples from these Gaussian distributions, which are defined by the hyper-parameters? So, shouldn't it be 1) sample hyper-parameters and spatially irrelevant parameters from a uniform distribution, and 2) sample the spatially relevant parameters from the Gaussian distributions? I believe this is mainly a textual issue which the authors can easily clarify, because when I look at S1 in the supplement, as an example, beta is not mentioned for scenario 1, which makes me think it is sampled based on the Gaussian defined by the hyper-parameters. So can you clarify this a bit more?

Reply: We apologize for unclear descriptions and thank you for your suggestions. Now clarifications have been made in the revised manuscript, please refer to lines 304-308 on pages 14-15. Hyper-parameters and spatially irrelevant parameters are sampled from the uniform distributions, while spatially relevant parameters are sampled from the Gaussian distributions. It should be noted that the prior ranges for unknown quantities are different.

(1) For instance, in scenario 1, regression parameter β is not sampled from the uniform distribution but it is the output from the Gaussian distribution $\beta = N(\mu_2, {\sigma_2}^2)$, in which hyper-parameters μ_2 and σ_2 are sampled from the uniform distributions with different ranges. While, other unknown quantities, i.e., model parameters (θ_2, θ_3, and θ_4), regression parameters α and ω (no β), are sampled from the uniform distributions. In conclusion, in scenario 1 (S1), β is estimated from the Gaussian distribution and ω is sampled from the uniform distribution, while hyper-parameter μ_3 and σ_3 are not considered.

(2) In scenario 2, ω is not sampled from the uniform distribution but it is the output from the Gaussian distribution $\omega = N(\mu_3, {\sigma_3}^2)$, in which hyper-parameter μ_3 and σ_3 are sampled from the uniform distributions. In contrast to scenario 1, β is sampled from the uniform distribution, while μ_2 and σ_2 do not exist in scenario 2. More details about the unknown quantities of different scenarios are presented in Supplement.

B2: I am happy with the additional criteria of mean annual maximum flow and mean annual minimum flow. However, as described and presented in the tables now, these are just the numbers obtained by the model. How are these values compared with the observations? Or are these numbers the error between modelled and observed annual maximum and minimum flow?

Reply: Thank you for your comments.

(1) These numbers represent the observed values and modeled values. The scenario with the minimum absolute difference between its modeled and observed

values is the best scenario.

(2) The percentage errors between modeled and observed annual maximum and minimum flow have been added to Tables 6 and 7. The scenarios with the minimum absolute errors are recognized as the best scenarios. Please refer to Tables 6 and 7 in the revised manuscript.

B3: With regard to my previous remark on the choice of the Gaussian distribution, and the authors response on that, I fully understand the reason why the authors used a Gaussian distribution. However, in my point of view, it is just really interesting to look a bit further, as there should be a physical reason why storage capacities (and/or their trend) are spatially related by a Gaussian distribution. Maybe the authors can just add some thoughts on the physical reasons for their findings in the discussion, as this is a bit missing in general.

Concluding, I am happy the authors found most of my comments useful and addressed all of them. When the authors also address the minor issues raised above, I would recommend publication of the manuscript.

Reply: Thank you.

(1) As a response to reviewer #2 in the previous round, the Gaussian distribution is one of the widely used distributions for describing the prior layer within the HB framework and has been applied in many previous studies, such as Sun et al (2015, 2016) and Chen et al (2014).

(2) The studied catchments are adjacent and have similar climate characteristics, e.g., the similar precipitation pattern and drought period anomaly (see Table 3). It is expected to have similar variation pattern of the catchment storage capacity for these catchments. However, there are still uncountable factors that may have impacts on the spatial coherence between adjacent catchments, which makes the coherence between β and ω tend to converge a central value but with finite variance (central limit theorem). The Gaussian distribution is the most likely distribution to describe the variables that obey the central limit theorem. These discussions also have been added on lines 254-258 on page 12.

Minor comments

B4: Generally, the terms dry and wet period are a bit confusing, as it makes me think of a wet season and dry season. The authors mean a longer period of dry years and wet years though. Maybe it is better to replace "dry period" and "wet period" throughout the manuscript with "dry years" and "wet years".

Reply: Thank you. Changes have been made as suggested. "dry period" and "non-dry period" have been modified as "dry years" and "wet years" in the revised manuscript.

B5: L176. Why should the anomaly be less than 25%?

Reply: Thank you. Sorry for this typo.

(1) According to Saft et al. (2015), the number here should be "-5%" rather than "25%". The sentence has been modified as "mean dry years anomaly<-5%.". The mean dry years' anomaly should be smaller than -5%, which is to identify dry years with more

than 5% less rainfall than wet years.

(2) It should be noted that "-5%" is an experimental parameter in Saft et al. (2015). Saft et al. (2015) tested several algorithms for dry years delineation, which considered different combinations of dry run length, dry run anomaly and various boundary criteria, and found that one of the algorithms, i.e., the method adopted in our study, showed marginal dependence on the algorithm and the main results were robust to different algorithms. Please refer to lines 166-177 on page 8.

B6: L224. When theta is constant, I think alpha needs to have a value, as written alpha is also zero, and then theta becomes zero too.

Reply: Thank you. As a response to comment A2 by reviewer #1, this sentence has been modified as follows.

According to the definition of the GR4J model (Perrin et al., 2003), the value of θ_1 must be a positive value. If model parameter θ_1 is constant, $\beta = 0$ and $\alpha > 0$ suffice in Eq.1. Meanwhile, the value of ω becomes irrelevant. Thus, the resulting model simplifies to a stationary hydrological model.

Please also refer to lines 224-228 on page 11.

B7: L274. Please add what T and t represent for completeness.

Reply: We apologize for this negligence. The meanings of T and *t* have been added in the revised manuscript. T represents the number of the time series while *t* represents the time step. Please refer to lines 287-289 on pages 13-14.

B8: L328. Please describe all your variables for completeness.

Reply: Thank you. The meanings of all variables in Eq.7 have been added in the revised manuscript. $Q_{sim}(t)$ and $Q_{obs}(t)$ represent the simulated and observed daily streamflow values of the t^{th} day, respectively. *T* refers to the length of the study period. Please refer to lines 334-336 on page 16 in the revised manuscript.

B9: L333. Please describe all your variables for completeness.

Reply: Thank you. The meanings of all variables in Eq.8 have been added in the revised manuscript. *p* refers to probability, q represents the observations of streamflow and ξ denotes the time series of model input, e.g., rainfall and potential evapotranspiration. Please refer to lines 348-349 on page 17.

B10: L335. Please describe all your variables for completeness.

Reply: Thank you. The meanings of all variables in Eq.9 have been added in the revised manuscript. Please refer to lines 348-349 on page 17.

B11: L347. Do you mean potential evaporation? Or do you use an estimate of actual evaporation?

Reply: We are sorry for this mistake. It should be "potential evapotranspiration" rather than "evapotranspiration" in this sentence. Change has been made in the revised manuscript. Please refer to line 363 on page 17.

B12: L381-387. This paragraph seems a bit odd to me. Why divide your timeseries into a dry and wet period if the change cannot be larger then 11%? In my view, if you want to test the robustness of the model you should actually even have higher differences then 11%. The discussed results of Vaze et al. (2010) only proof that those models were not robust and can not model extreme cases of droughts. Or arguing from the other end, if the change between rainfall in the dry and wet period is just hypothetically 0.0001%, what is the whole point of splitting into dry and wet periods?

Reply: Thank you. The "11%" is the catchment average reduction from wet years to dry years of the three study catchments. It should be noted that it is statistical results rather than evaluation criteria to divide the time series into dry years and wet years. The criteria for segmenting the time series have been presented in section 2.1. Please refer to lines 166-177 on page 8 in the revised manuscript.

Furthermore, this paragraph has been modified as follows:

In terms of changes in rainfall, on average catchments had an 11% reduction from the wet years to the dry years (Table 3). Meanwhile, these catchments experienced a 17.6% decrease in runoff during the dry periods, which is more severe than the reduction in rainfall. The similar findings can be derived out from the comparison of runoff coefficients of different periods, that is, all catchments experienced a decrease in its runoff coefficients during the dry years.

Please refer to lines 397-402 on page 19.

B13: L396. What do you mean with the variation?

Reply: Thank you. This sentence has been modified as "Furthermore, the magnitude of performance loss increases along with the variation in rainfall between the calibration and verification periods." Please refer to lines 410-412 on pages 19-20.

B14: L411. This is scenario 4 still, correct?

Reply: Yes. To improve the clarity of the manuscript, this sentence has been modified as "However, the introduction of additional regression parameters (α, β and ω) at the same time amplified the model projection uncertainty in two of three catchments (225219 and 405264) when comparing results from scenarios 4 and 5." Please refer to lines 428-431 on page 20.

B15: L417. Aren't scenarios 4 and 5 both higher than scenario 3 for 405264? Hard to see in the plot.

Reply: Thank you. During the verification period, the median estimate of scenario 4 is a little higher than that of scenario 3. Conversely, the median estimate of scenario 5 is inferior to that in scenario 3. In addition, the Figures have been improved, white dots have been added to represent the median estimates of the results in the violin plots.

B16: L427. Do you mean Figure 6?

Reply: We are sorry for this typo. It should be Figure 6 rather than Figure 5. Changes have been made in the revised manuscript accordingly.

B17: L427-431. I agree, but it's quite normal that the period used for calibration outperforms the verification.
Reply: Thank you. We agree with reviewer #2 that it's quite normal that the period used for calibration outperforms the verification. The purpose of this sentence is to verify the rationality of the results, which is the basis for further analysis.

B18: L439. This is hard to see in violin plots
Reply: Thank you. The violin plots have been modified to improve clarity. The white dots have been added to represent the median estimates of the results in the violin plots.

B19: L440. Isn't the range for 405219 larger?
Reply: We are sorry for this typo. It should be catchment 225219 rather than 405219 here. The sentence has been modified in the revised manuscript. Please refer to lines 463-474 on page 22.

B20: L441. Do you mean the second smallest in variation?
Reply: Yes. The phrase "in variation" now has been added in this sentence. Please refer to the response to comment B19.

B21: L442. Are you comparing here just scenario 4 with 5, or 1-4 with 5? In the last case, this statement is not always true, as far as I can see, and following the discussion above.
Reply: Thank you. In this sentence, we compared the results from scenarios 4 and 5 rather than compared scenarios 1-4 with 5, because the only difference (Principle of a single variable) between these two scenarios is that scenario 4 adopted the time-varying parameters while scenario 5 adopted the temporal invariant parameters. We did not compare scenarios 1-3 with 5 here either, because the former not only adopted the time-varying parameters but also used the spatial coherence, and it would be not easy to distinguish the impacts by each individual.

This sentence has been modified as follows:

These results demonstrate that the time-varying scheme (scenario 4) for model parameters improved the median NSE_{sqrt} performance but also amplified the projection uncertainty compared with the results from the stationary scheme (scenario 5) for model parameters.

Please refer to lines 460-463 on page 22.

B22: L460. When compared→ when comparing

Reply: Thank you. Change has been made as suggested. Please refer to line 486 on page 23.

B23: L468. How can I see this? It is just made bold, please add the observed values or present an error measure in the table.

Reply: Thank you. The observed MinF and MaxF were presented at the first row of Tables 6 and 7. As a response to comment B2 by reviewer #2, Tables 6 and 7 have been modified, the % variation between the modeled value and the observed value have been presented in the revised manuscript. The scenarios with the least absolute variation between the modeled values and the observed values are recognized as the best scenarios.

Further explanations have been made in the revised manuscript, please refer to lines 358-360 on page 17.

B24: L471. Idem as above.

Reply: Thank you. Changes have been made to address the reviewer's comment. Please refer to the response to comment B23 by Reviewer #2.

B25: Figs. 5,6,8,9. I am not sure if the violin plots are much more helpful compared to the boxplots in the previous version of the manuscript. It becomes more complicated to find the median values, which the authors often refer to, especially as some posterior ranges are noted nicely equally distributed. I believe the median values are a dashed black line, but this is hard to see. I would suggest different colors or line types for example the medians.

Reply: Thank you. Follow the reviewer's comment. whites dots that represent the median estimates of the results have been added in the revised figures. We hope that it can help to relieve the reviewer's doubt. Please refer to Figure 5,6,8 and 9 in the revised manuscript.

B26: Fig. 5 Varification –> Verification.

Reply: Thank you. Changes have been made as suggested.

B27: Fig.7. I guess the bars are the reference 10-year flow, but please make sure it is clear which x-axis (left-right) belongs to which graph. Add the bars also to the legend.

Reply: Thank you. The main scales of the figures have been added, which are helpful to find out which x-axis belongs to which graph easily. In addition, the bars have been added in the revised Figure 7. Please refer to Page 44.