**Review of "*Improving hydrological projection performance under contrasting climatic conditions using spatial coherence through a hierarchical Bayesian regression framework*" by Pan et al.**

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 transfered 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 myprevious comments and made improvements based on that.

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?

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?

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.

**Minor comments**

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".

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

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.
L274. Please add what T and t represent for completeness.
L328. Please describe all your variables for completeness.
L333. Please describe all your variables for completeness.
L335. Please describe all your variables for completeness.
L347. Do you mean potential evaporation? Or do you use an estimate of actual evaporation?
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?
L396. What do you mean with the variation?
L411. This is scenario 4 still, correct?
L417. Aren't scenarios 4 and 5 both higher than scenario 3 for 405264? Hard to see in the plot.
L427. Do you mean Figure 6?
L427-431. I agree, but it's quit normal that the period used for calibration outperforms the verification.
L439. This is hard to see in violin plots
L440. Isn't the range for 405219 larger?
L441. Do you mean the second smallest in variation?
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.
L460. When compared → when comparing
L468. How can I see this? It is just made bold, please add the observed values or present an error measure in the table.
L471. Idem as above.
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 note 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 for example the medians.
Fig. 5 Varification –> Verification
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.