Review of manuscript HESS-2022-67 entitled "Subseasonal precipitation forecasts using preceding atmospheric intraseasonal oscillation signals in a Bayesian perspective" by Yuan Li, Zhiyong Yu, Hai He and Hao Yin

OVERALL RECOMMENDATION

Major revision

SUMMARY

This manuscript proposes the construction of subseasonal forecasts of pentad-mean precipitation in 17 hydroclimatic regions in China. These subseasonal forecasts are generated using a purely statistical method called Bayesian Hierarchical Modeling (BHM). This statistical model uses previously observed climate variables at the global scale as input. The performance of the statistical subseasonal forecasts is evaluated with a deterministic metric (Kling-Gupta Efficiency) and probabilistic diagnostics (CRPSS, reliability diagram). The authors show that, in terms of probabilistic verification, these forecasts are still skillful in all 17 regions up to 25-day lead time. They also claim that the most interesting results are to be found in southern China.

MAJOR COMMENTS

This is a very relevant topic to propose statistical models for subseasonal forecasting based on lagged relationships. Not only can they be used as a benchmark to assess dynamical subseasonal forecasts (e.g S2S, SubX), but they might also prove more skillful than them. This seems to be the underlying claim of the authors for the statistical forecasts in the manuscript.
Then, I consider this study might be worthy of publication. However, it suffers from a lack of details that cast a doubt on the real added value of the method. I therefore ask the authors to convince me of its benefits through major revisions, as I feel the main claims are insufficiently supported in the current version.

Here are my major concerns, by order of importance:

1) The scores that are used to claim the benefits of the method **should be compared** to the scores obtained with raw dynamical subseasonal forecasts (e.g ECMWF), and possibly with your own BJP-processed from Li et al (2020). All those scores should appear simultaneously in Figures 3 and 4.

2) The methodology should be illustrated with more figures besides Figure 2. For instance, you could show the results of the LASSO predictor selection for the Figure 2 example (Region 1) at a specific lead time. Then, you could also select a specific target week (e.g your May 1-May, 5, 1979 period) and simultaneously visualize the values of the different predictors and the predicted precipitation.

More generally speaking, my recommendation is to **open the "black box"** and give more visual information showing what the statistical model is doing and why it works.

3) A figure **summarizing the different steps of the statistical prediction** is necessary for the reader to have a complete vision of the workflow.

4) Some **spatial visualization of the scores** is missing, e.g a map where the 17 regions are colored according to their score. This is important to support the claim that the method performs best in southern China. You could also give names to the regions and indicate them on Figures 3 and 5, this would help a lot.

5) In order to compensate for the necessary additional details required in my comments 1 to 4, some parts of the manuscript **could be shortened** (e.g Introduction, Sections 4 and 5).

6) Section 2.2.4, l.316: "The reference forecasts are generated using the Bayesian hierarchical model with no predictors used for prediction."
l.318: "show no improvement over the cross-validated climatology"
→ **It is unclear to me what the reference in CPRSS is**. Is it the cross-validated climatology or the forecasts generated with no predictor? Are they the same? If so, you should state it explicitly.

MINOR COMMENTS

Figures 3 and 4: I think the graphical aspect of these figures could be improved (e.g vertical scale, colored bars, etc.).

Figure 4: The curves on Figure 4 are illegible as there are too many time steps. Personally, I can't see the red curve (model) and how it compares to the observations in blue. Actually, I'm not sure this figure is really necessary beyond the indications in the top left-hand corner (KGE, r, etc.), I suggest replacing by a table.

Figure 5: I am surprised that CRPSS does not decrease monotonically with lead time. Admittedly there can be some noisy variations at longer lead times, but I still find that some results are quite weird (e.g in Region 2, CRPSS at 20 days is better than at 0 day). Isn't there an effect of the reference that is used in the CRPSS? Some explanations should be provided.

l.414-416: Please specify what are "the BJP calibrated sub-seasonal precipitation forecasts" from Li et al. (2020). I guess it corresponds to post-processed outputs of dynamical subseasonal forecasts with a GCM, but you should remind it and give the name of the model. More generally, your assertions concerning the comparison between BHM and your previous method from Li et al (2020) should be illustrated more extensively (see Major Comment #1).

l.436-438: "*Here*, we analyzed the spatial patterns of correlations between lagged signals and filtered precipitation over Region 1 at the lead time of 0-day for each step of the leave-one-year out cross-validation".
I can't see where the results you are referring to are, e.g I don't know what "Here" stands for in this sentence.

l. 19: "owing to the *underestimation* of intraseasonal variability in this region"
l.396: "The decomposition of KGE values suggest that the intraseasonal variability is *underestimated* in these regions"
I am not sure "underestimation" is the correct word in this context. From what I understand, the important fact is that intraseasonal variability is of limited importance in those regions because it does not account for a large fraction of total variability, so the model cannot perform well in those regions. I suggest rephrasing.

l.381-382: "The results also suggest that the probabilistic forecasts are sharp at all lead times, especially for below-normal and above normal categories".
Judging by the reliability diagrams, I am not convinced by the sharpness of the forecasts. On the contrary, I think the authors should mention very limited sharpness. I guess this is intrinsic to a Bayesian approach relying on a non-informative prior.

LANGUAGE AND TYPOS

l.9: "as predictor**s**" → "as predictor"
l. 19: "owing to the underestimation of intraseasonal variability in this region". Why underestimation?
l.22: "Other sources (…) ~~would~~ **will** be included"
l. 22: "forecast skill**s**" → "forecast skill".
I think that the word "skill" is never expected to be plural in this context. Same remark at l.34, l.74, l.116 (x2), l.395, l.425
l.25: "mitigation**s**" → "mitigation"
l.28: ~~"lunched"~~ → "launched"
l.30: ~~"could not"~~ → "cannot"
l. 32: "before it ~~could~~ **can** be used"
l.41: "atmospheric-oceanic indices" → Do you mean "atmospheric or oceanic indices"?
l.43: "dominant" → I suggest using another word, what about "more performant"?
l.45: "plenty **of**"
l.48-51: "a new cluster-based empirical method (…), which the sea surface temperature (…) were included as predictors.".
The sentence is unclear, I suggest rephrasing, e.g splitting the sentence in two: "a new cluster-based method (…) European and Mediterranean regions**. This method uses** sea surface temperature (…) as predictors".
l.56: "at such a time scale" Unnecessary, please remove.
l.69: "but **in** extra-tropical regions as well"

l.77-78: “the relationships between ISO signals and precipitation are of high uncertainty for different regions at different lead times”. I suggest rephrasing, e.g “the relationships between ISO signals and precipitation **are highly uncertain and depend on the region and lead time.**”
l.79-81: “To our best knowledge, the uncertainties of relationships between preceding ISO signals and sub-seasonal precipitation have not been fully considered in sub-seasonal precipitation forecasts in previous studies.” I suggest another formulation.
l.84: Remove the CSC acronym. You never use it in the rest of the article.
l.87: “Bayes-theorem based statistical models” → “**Bayesian** statistical models”
l. 91: Idem
l.104: “is frequently influenced by” → “is frequently **subject to**”
l.111: “the model performance (…) ~~are~~ **is** evaluated”
l.115-116: “the deterministic and probabilistic forecast **skill is** presented”
l.127: “is area-weighted averaged over 17 hydroclimatic regions”
l.134: “to monitor~~ing~~” → “to monitor”
l.139: “proved to be capable of reflecting the MJO structure as the zonal wind” Unclear → “proved to be **as** capable of reflecting the MJO structure as the zonal wind”?
l.148: “calculating efficiency” → “computational efficiency”?
l. 194, l.196: “~~the~~ Africa” → “Africa”
l. 234: “in (Nardi and Rinaldo, 2011; Mcneish, 2015)”. Typo, remove parentheses.
l.301: ~~“A full discussion of the KGE-statistics sees Gupta et al (2009)…”~~ → “For a full description of KGE-statistics, see Gupta et al (2009)...”