Responses to Comments on "Sub-seasonal precipitation forecasts using preceding atmospheric intraseasonal oscillation signals in a Bayesian perspective" (Referee #2)

Anonymous referee #2 reported on 11 May 2022.

Our responses are in blue and proposed manuscript revisions underlined.

General 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.

Thanks for your comprehensive review and recognition of the study contribution. The constructive comments will help us improve our manuscript after revision. We provide detailed responses to your comments and our proposed manuscript revisions in the subsequent sections.

Major comments:

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.

Thanks for this comment. We agree that it is important to compare the skill scores of the statistical model we built in this study and the raw dynamical models. However, we also note that the configurations of the statistical model are not the same as the raw dynamical models. Consider, for example, predicting pentad mean precipitation during the period of 1th May and 5th May 1979. In this case, the pentad mean ISO signals during the period of 26th April and 30th April 1979 are used to predict the pentad mean precipitation at a lead time of 0-day (could also be referred as a lag time of 0-day). The pentad mean precipitation and corresponding ISO signals during the period of 1980-2016 from May to October are pooled together to make parameter reference. On the other hand, the raw dynamical models are not always able to provide pentad mean precipitation forecasts for the same period of 1th-5th May 1979 as the hindcast initial time, hindcast period, and hindcast frequency are different (Table 1). The comparison may be unfair if the predictand of the statistical model and raw dynamical models are not the same.

To overcome this problem, we would like to compare our results with the NCEP model. Although the NCEP model is not the top scoring model for sub-seasonal precipitation forecasts (De Andrade et al., 2019), the hindcast frequency of the NCEP model makes it able to generate pentad mean precipitation forecasts for the same period as the statistical model from 1999 to 2010 (Table 1).

Table 1. Configuration of S2S model hindcasts

S2S model	Time range (days)	Spatial resolution	Hindcast frequency	Hindcast period	Ensemble size	Ocean coupling
ECMWF*	46	Tco639/Tco319, L91	2/week	Past 20 years	11	Yes
NCEP	**44**	**T126, L64**	**Daily**	**1999-2010**	**4**	**Yes**
JMA	33	TL479/TL319, L100	3/month	1981-2010	5	No
KMA*	60	N216, L85	4/month	1991-2010	3	Yes
UKMO*	60	N216, L85	4/month	1993-2016	7	Yes
CNRM	61	T255, L91	2/month	1993-2014	15	Yes
ECCC*	32	0.45°X0.45°, L40	Weekly	1998-2017	4	No
ISAC	31	0.75°X0.56°, L54	Every 5 days	1981-2010	5	No
BOM	62	T47, L17	6/month	1981-2013	33	Yes
CMA	60	T106, L40	Daily	1994-2014	4	Yes
HMCR*	61	1.1°X1.4°, L28	Weekly	1985-2010	10	No

*Hindcasts are produced on the fly (model version is not fixed)

Moreover, we will develop a STPM2-BHM statistical model as suggested by the Anonymous Referee #1. We will no longer define potential predictors by averaging ISO signals in the areas of significant correlations. Instead, the predictors will be defined by extracting the coupled patterns between pentad precipitation anomalies and atmospheric intraseasonal oscillation signals, which is also known as STPM2 in Fig. 1 (Hsu et al., 2015). The BHM model is then built to address the parameter uncertainty in the transfer function (Fig. 1). A more detailed description of the newly built STPM2-BHM model can be found at https://hess.copernicus.org/preprints/hess-2022-67/hess-2022-67-AC1-supplement.pdf.

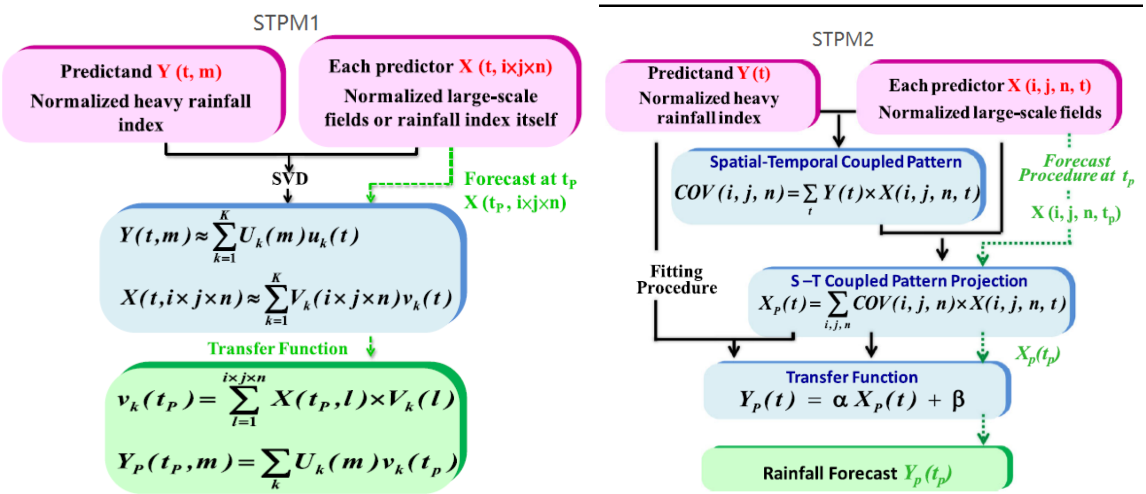

Fig. 1. Major steps of STPM1 and STPM2 prediction model (Hsu et al., 2015).

The predictors we defined above could be used to predict pentad mean anomalies and pentad mean precipitation amount as well. Our previous responses to comments of the Anonymous Referee #1 had shown that the STPM2-BHM model was capable of predicting pentad mean anomalies. Here, we use the same predictors, including the U850, U200, OLRA, H850, H500, and H200, to predict pentad mean precipitation over different hydroclimatic regions. The results will then be compared with the NCEP model.

Figure 2 compares the pentad mean precipitation forecasts of the NCEP model and the STPM2-BHM model over Region 1 (Inland rivers in Xinjiang). The predictors are firstly defined by the spatial-temporal coupled covariance patterns of U850, U200, OLRA, H850, H500, and H200, seperately. All covariance patterns are then pooled together to predict pentad mean precipitation. It is not surprise that the NCEP model outperforms the STPM2-BHM model when the lead time is within 5-10 days. However, the STPM2-BHM model shows much higher forecast skill compared to the NCEP model at longer lead times. In addition, the ensemble spread of the NCEP model is too narrow to provide reliable sub-seasonal forecasts as the ensemble size of dynamical models is always limited. In comparison, the STPM2-BHM model is of much higher reliability compared to raw dynamical models.

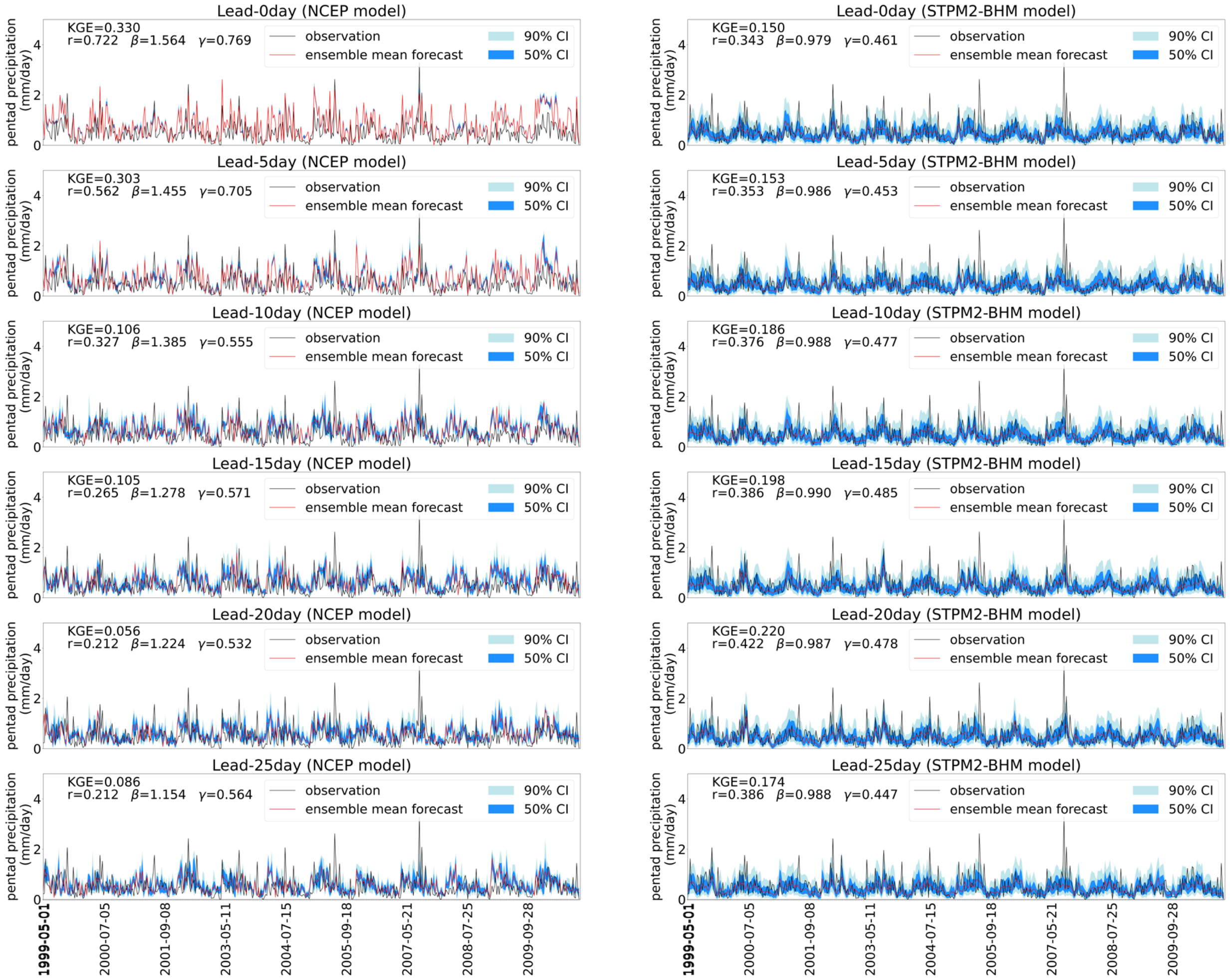

Fig. 2. Comparison of pentad mean precipitation forecasts of the NCEP model and the STPM2-BHM model over Region 1 (Inland rivers in Xinjiang).

We will incorporate the above proposals in the revised manuscript. A more detailed comparison of the NCEP model and the STPM2-BHM model will be given to provide evidence of the added value of the statistical method we proposed in this study.

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.

Thanks for this comment. As suggested by the Anonymous Referee #1, we will no longer define potential predictors by averaging ISO signals in the areas of significant correlations. The LASSO and stepwise regression approaches will not be used to select potential predictors. Instead, we will follow the steps of STPM2 (Fig. 1) to define coupled pattern projection coefficients as predictors. We first construct spatial-temporal coupled covariance patterns (COV) where the predictand Y (pentad mean anomalies or pentad mean precipitation amount) and predictors X (10-60-day component of atmospheric ISO signals) are significantly correlated,

$$cov(X_{i,p},Y)=\frac{1}{T}\sum_{t=1}^{T}(y_t-E(y))(x_{i,p,t}-E(x_{i,p})). \tag{1}$$

where t is the number of pentads during the training period, y_t is the pentad mean anomalies or pentad mean precipitation amount, and $x_{i,p}$ is p^{th} ISO atmospheric field of U850, U200, OLRA, H850, H500, and H200 for significantly correlated grid *i.*

The coupled pattern projection coefficient X_p , which can also be regarded as the predictor, is then obtained by <u>multiplying the covariance and predictors X, and summing the product for each grid point where a 95 % significant level is reached</u>,

$$X_p=\sum_{i=1}^{N}cov(X_{i,p},Y)*X_{i,p} \tag{2}$$

where N is the total number of grid points where a 95% significant level of correlation coefficient is reached. Thus, there will be only one predictor for each ISO atmospheric field of U850, U200, OLRA, H850, H500, and H200.

To open the "black box" of the STPM2-BHM model, we will compare the forecast skill of the STPM2-BHM model with different predictors. Figure 3 shows the KGE values of the NCEP model and the STPM2-BHM model with different predictors over Region 1 (Inland rivers in Xinjiang). It is clear that the ISO signals of U850, U200, and H850 fileds contributed most to the sub-seasonal precipitation forecasts over Region 1 (Inland rivers in Xinjiang). In contrast, the OLRA and H200 may have limited effects on predicting sub-seasonal precipitation forecasts over this region.

<u>We will have a more comprehensive analysis of the results by comparing the forecast skill of the STPM2-BHM model with different predictors as mentioned above.</u>

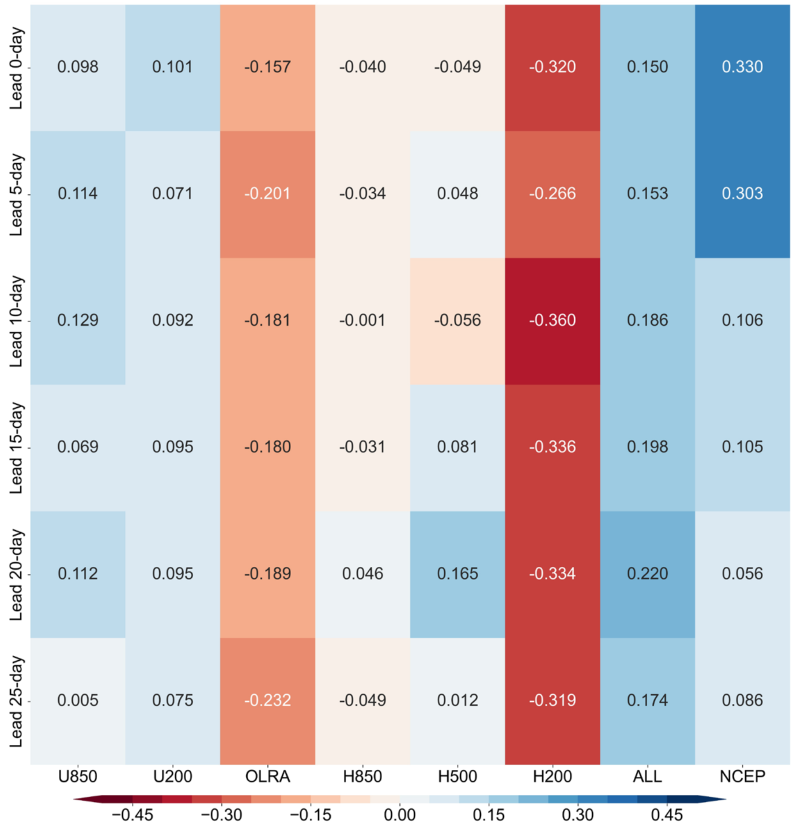

Fig. 3. Comparison of KGE values of the NCEP model and the STPM2-BHM model over Region 1 with different predictors (Inland rivers in Xinjiang).

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

Thanks for this comment. We have summarized the major steps of the STPM2-BHM model in Figure 4. <u>We will add this flow chart in the Data and methodology section.</u>

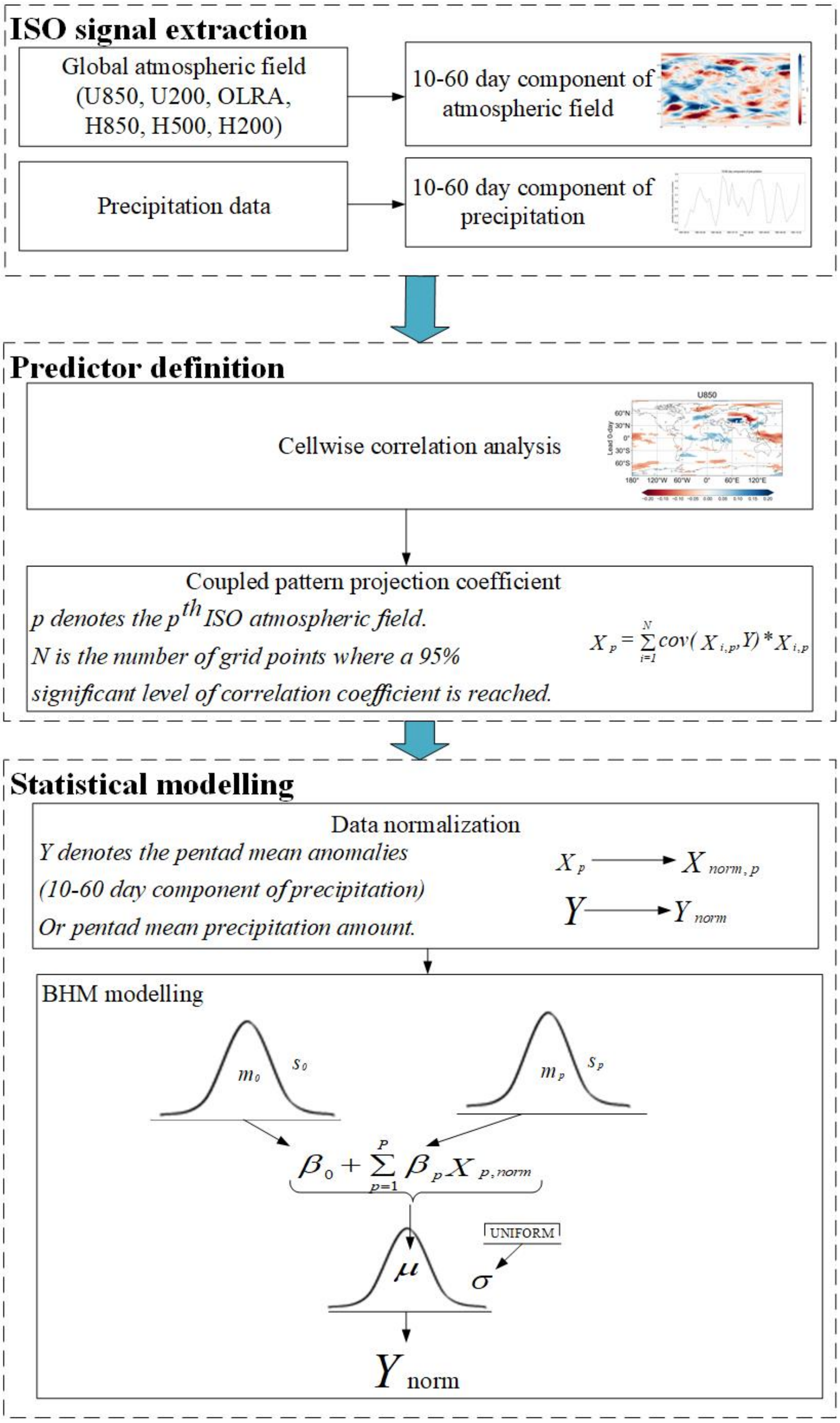

Fig. 4. Flow chart representing the major components of the STPM2-BHM model.

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.

We agree with the referee that the spatial map of forecast skill over 17 regions could help to realize the spatial pattern of sub-seasonal predictability. However, this would lead to large number of figures as we also compare our forecast skill using different predictors. <u>To address this comment, we will plot the heatmap to have a better illustration of forecast skill as shown in Figure 5</u>. The region names are also presented. The NCEP model shows higher forecast skill in most regions when the lead time is within 10 days. However, the

STPM2-BHM model performs better compared to the NCEP model at longer lead times in most regions, except Region 15 (Lower Yangtze River), Region 16 (Pearl River), and Region 17 (Southeast rivers).

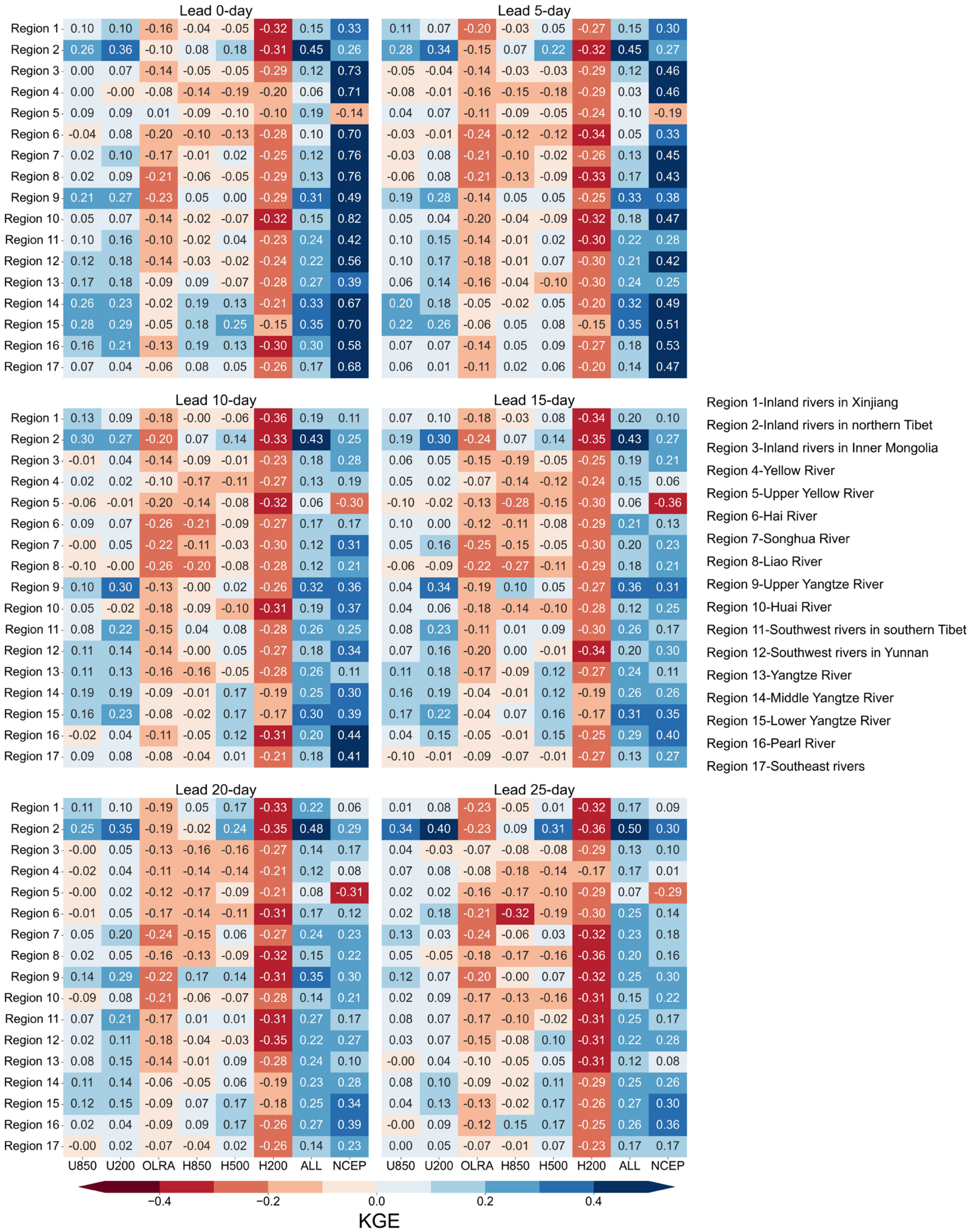

Fig. 5. Comparison of KGE values of the NCEP model and the STPM2-BHM model over 17 hydrological regions at different lead times.

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).

Thanks for this comment. We will incorporate this suggestion in the revised manuscript.

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.

Thanks for this comment. In this study, the reference forecasts are generated with no predictors. This is the same meaning as the cross-validated climatology, which the mean and standard deviation of the predictand is only determined by the cross-validated precipitation data. We will have a more detailed description of the reference forecasts in the Data and methodology section.

MINOR COMMENTS

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

Thanks for this comment. We will plot the heatmap to have a better illustration of forecast skill as shown in Figure 5.

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.

Thanks for this comment. We will remove Figure 4 in the revised manuscript, and the curves will be provided as supplementary files. The results of KGE values will be presented as shown above in Figure 5.

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.

Thanks for this comment. It is true that the forecast skill decreases as lead time increases for dynamical models. This is also can be found in Figure 5 and our previous study (Li et al, 2021).

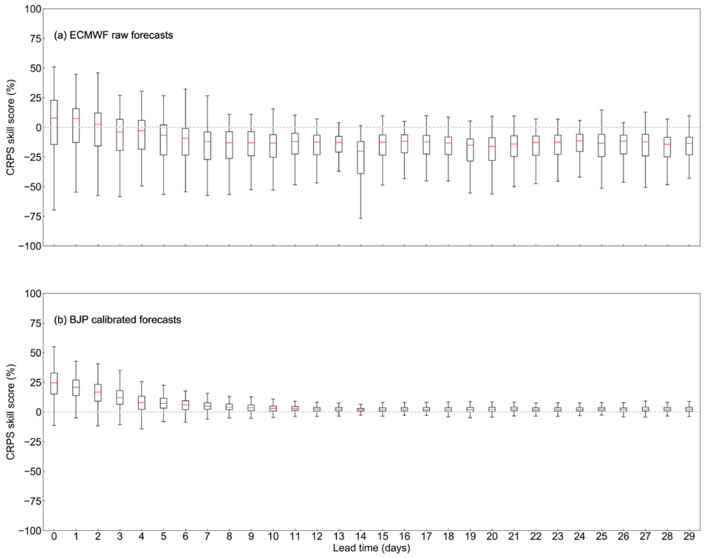

Fig. 6. Boxplot diagrams of CPRS skill scores of ECMWF raw ensemble forecasts (top) and the BJP calibrated forecasts (bottom) at different lead times during the boreal summer monsoon. (Li et al., 2021)

However, we should note that the STPM2-BHM model is a purely statistical model. The forecast skill of the STPM2-BHM model is mostly determined by the relationship between precipitation and atmospheric ISO signals. The concurrent relationship between precipitation and atmospheric/oceanic signals may not be as strong as lagged signals. For example, Shukla et al. (2011) found that the Nino-3 index had strongest relationship with Indian Summer Monsoon Rainfall Index (ISMRI) with a lag of 5 sesasons (MAM). Thus, the forecast skill of ISMRI were found to be higher at a lag of 5 seasons compared to a lag of 4 seasons when using the Nino-3 index. This is also found by many other studies, which the relationship between precipitation and large scale circulation signals may be stronger at longer lags (Kirono et al., 2010; Piechota et al., 1998). Thus, it is not surpise that the skill scores of the STPM2-BHM model are higher at longer lead times (could also be referred as longer time lags). To address this comment, we will provide comprehensive explanations of the results and compare with previous statistical modelling studies in the discussion section.

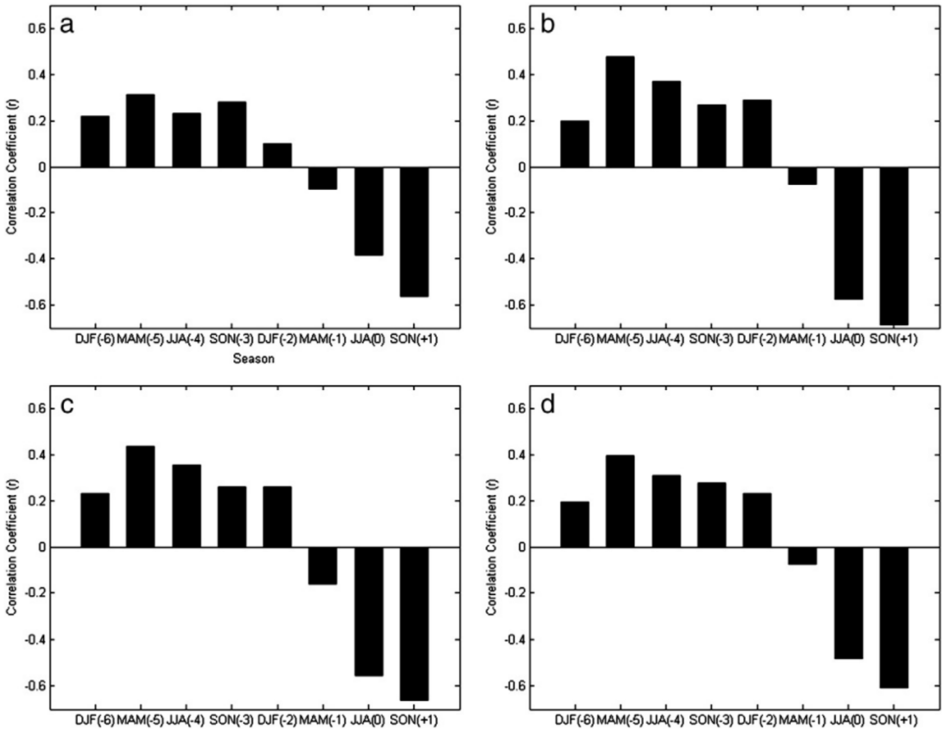

Fig. 7. Correlation coefficient (r) between India Summer Monsoon Rainfall Index (ISMRI) and (a) Niño 1+ 2 index, (b) Niño 3 index, (c) Niño 3.4 index and (d) Niño 4 index with Niño indices lagging by 1–8 season(s) (Shukla et al., 2011).

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).

Thanks for this comment. Our previous study post-processed ECMWF model outputs using the Bayesian Joint Probability (BJP) approach. In this study, we will compare the STPM2-BHM model with the NCEP model as we introduced previously.

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.

Thanks for this comment. We have analyzed the non cross validated correlation and cross validated correlation between ISO of U850 and precipitation over Region 1 in Figure 8. However, the results are not presented in the manuscript as the cross validated correlation could be generated for each year for the leave-one-year-out cross validation strategy. <u>To address this comment, we will provide the cross validated correlation maps as supplementary files.</u>

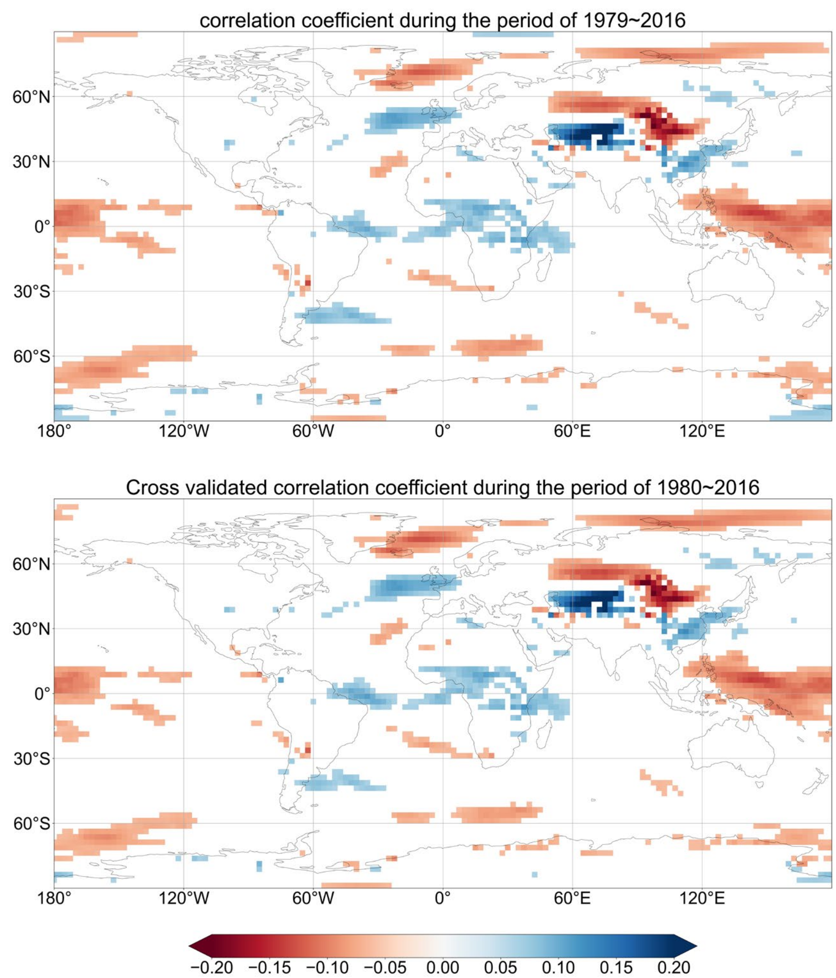

Fig. 8. Comparison of non cross validated correlation and cross validated correlation between preceding ISO signals of U850 and 10-60-day component of precipitation over Region 1 (Inland Rivers in Xinjiang) at a lag of 0-day.

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.

We agree with the referee that the intraseasonal variability is of limited importance in these regions. We will rephase these sentences in the revised manuscript.

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.

The authors agree with the reviewer that the Bayesian approaches may have difficulty in predicting extreme events when non-informative prior is used. The copula-based statistical approches will be used in the future to see whether the shapness of forecasts could be improved. To address this comment, we will provide more discussion on the sharpness of sub-seasonal forecasts in the Discussion section.

LANGUAGE AND TYPOS

l.9: “as predictors” → “as predictor”

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

l. 19: “owing to the underestimation of intraseasonal variability in this region”. Why underestimation?

Thank you for this comment. We agree with the referee that the intraseasonal variability is of limited importance in these regions. We will rephase these sentences in the revised manuscript.

l.22: “Other sources (…) would will be included”

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

l. 22: “forecast skill” → “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

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

l.25: “mitigations” → “mitigation”

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

l.28: “lunched” → “launched”

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

l.30: “could not” → “cannot”

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

l. 32: “before it could can be used”

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

l.41: “atmospheric-oceanic indices” → Do you mean “atmospheric or oceanic indices”?

Thank you for this comment. The atmospheric-oceanic indices mean atmospheric or oceanic indices here. We will incorporate this suggestion in the revised manuscript.

l.43: “dominant” → I suggest using another word, what about “more performant”? l.45: “plenty of”

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

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”.

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

l.56: “at such a time scale” Unnecessary, please remove. l.69: “but in extra-tropical regions as well”

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

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."

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

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.

Thank you for this comment. We will rephrase this sentence in the revised manuscript as follows: To our best knowledge, the uncertainty in relationship between preceding ISO signals and sub-seasonal precipitation has not been fully considered yet.

l.84: Remove the CSC acronym. You never use it in the rest of the article.

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

l.87: "Bayes-theorem based statistical models" → "Bayesian statistical models"

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

l. 91: Idem

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

l.104: "is frequently influenced by" → "is frequently subject to" l.111: "the model performance (…) are is evaluated"

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

l.115-116: "the deterministic and probabilistic forecast skill is presented" l.127: "is area-weighted averaged over 17 hydroclimatic regions"

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

l.134: "to monitoring" → "to monitor"

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

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”?

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

l.148: “calculating efficiency” → “computational efficiency”?

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

l. 194, l.196: “the Africa” → “Africa”

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

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)...”

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

References

de Andrade, F. M., Coelho, C. A. S., and Cavalcanti, I. F. A.: Global precipitation hindcast quality assessment of the Subseasonal to Seasonal (S2S) prediction project models, Climate Dynamics, 52, 5451-5475, https://doi.org/10.1007/s00382-018-4457-z, 2019.

Hsu, P.-C., Li, T., You, L., Gao, J., and Ren, H.-L.: A spatial–temporal projection model for 10–30 day rainfall forecast in South China, Climate Dynamics, 44, 1227-1244, 10.1007/s00382-014-2215-4, 2015.

Kirono, D. G. C., Chiew, F. H. S., and Kent, D. M.: Identification of best predictors for forecasting seasonal rainfall and runoff in Australia, Hydrological Processes, 24, 1237-1247, https://doi.org/10.1002/hyp.7585, 2010.

Li, Y., Wu, Z., He, H., Wang, Q. J., Xu, H., and Lu, G.: Post-processing sub-seasonal precipitation forecasts at various spatiotemporal scales across China during boreal summer monsoon, Journal of Hydrology, 125742, 2021.

Piechota, T. C., Chiew, F. H. S., Dracup, J. A., and McMahon, T. A.: Seasonal streamflow forecasting in eastern Australia and the El Niño–Southern Oscillation, Water Resources Research, 34, 3035-3044, https://doi.org/10.1029/98WR02406, 1998.

Shukla, R. P., Tripathi, K. C., Pandey, A. C., and Das, I. M. L.: Prediction of Indian summer monsoon rainfall using Niño indices: A neural network approach, Atmospheric Research, 102, 99-109, https://doi.org/10.1016/j.atmosres.2011.06.013, 2011.