EGUsphere, referee comment RC2
https://doi.org/10.5194/egusphere-2022-449-RC2, 2022

Comment on egusphere-2022-449

Anonymous Referee #2

Referee comment on "Seasonal forecasting skill for the High Mountain Asia region in the Goddard Earth Observing System" by Elias C. Massoud et al., EGUsphere, https://doi.org/10.5194/egusphere-2022-449-RC2, 2022

Summary

This study evaluates the subseasonal predictions from the NASA GEOS5-S2S hindcasts for 1981-2016 over the High Mountain Asia (HMA) domain with a focus on a set of hydrometeorological variables including 2-m air temperature, precipitation, snow cover fraction, snow water equivalent, soil moisture, and total water storage. The evaluation was done against two reanalyses and other independent datasets, and the evaluation focuses on monthly time scale with lead times up to 3 months. Unbiased root mean square error and anomaly correlation are the major metrics used in this evaluation.

Overall, the study provides useful information about the predictive skill of the NASA GEOS5-S2S hindcast over the HMA region. The manuscript is well written and easy to understand, and the quality of the visualizations is generally good. However, the study falls short in several important aspect regarding forecast verification at subseasonal to seasonal time scales. The value and the contribution of this study to our understanding about predictability of the climate system at S2S time scale is very limited. I believe at the minimum a major revision is needed. I list my major concerns and some specific comments below.

Major issues

- For prediction beyond the typical weather scale (i.e, 1-2 weeks), probabilistic forecast is more appropriate and useful than deterministic forecast given the chaotic nature of the climate system, which is why S2S forecast with numerical models needs to produce ensemble predictions. In this study, only unbiased root mean square error (ubRMSE) and the anomaly correlation (ACC) of the ensemble mean were used, which is useful but only shows very limited aspects of the forecast quality. There are many verification metrics that can be used for ensemble predictions such as those listed at

https://www.cawcr.gov.au/projects/verification/#Methods_for_probabilistic_forecasts. I'd highly recommend that a few more meaningful metrics are included in this study.

- My biggest concern regarding the analysis is its over-simplified approach to deal with the spatial heterogeneity within the study domain. The study domain is quite large; more importantly, it is very heterogeneous with distinct climates and land surface characteristics including elevation, land cover type, etc. As shown in Figures 7-12, temperature, precipitation and other hydrometeorological variables and model's skill in predicting these quantities can vary drastically across the domain. Spatially averaging them across high mountain ranges, the Tibet Plateau, Taklamakan desert, and the Indian subcontinent does not make much sense, and evaluating the spatially averaged quantities is not very meaningful and insightful. It is not clear what these spatial averages physically mean and how verification at such a level can help us to understand the model deficiency in a meaningful way. Although Figure 7-12 highlight the spatial heteorogeneity, the evaluation is only limited to the ensemble mean, spread, and ubRMSE. I'd suggest that the authors divide the domain into multiple smaller regions that are more homogeneous or multiple watersheds where the spatial averages are more meaningful, and conduct the forecast verification of these regional quantities using multiple metrics (probabilistic and deterministic).

Minor issues

- line 13: "where water resource needs change depending on ..." although this sentence is correct, it could be a little confusing as either "needs" or "change" can be interpreted as the verb, thus resulting in different meanings.

- line 13: how is intensity of the hydrological cycle defined? It was not mentioned in the study.

- line 30: "a range of factors", the predictability itself is also an important factor.

- line 34: remove the comma before "where"

- line 35-36: This sentence reads a little awkward, please consider rephrase.

- line 40: Part of the study domain is heavily populated, but the majority of HMA do not have much population, such as Tibet Plateau and dessert.

- line 43: The term "water tower" of the Earth have been used for many years among researchers in Asia, so some earlier literature needs to be cited here to be more appropriate.

- line 144: This is only over the real-time forecast period, isn't it? Please clarify that these 6 additional members are not available in the hindcast period and thus not used in the evaluation.

- line 147: "a long period for forecast validation" "validation" and "verification" are different terms although they are related. One can verify if a forecast is correct or wrong, but you cannot validate a forecast when the forecast is wrong. So it would be more appropriate to say "forecast verification" or "forecast evaluation" here.

- line 233: remove "in our evaluation" as it is redundant with "in this study" at the beginning of the sentence.

- line 234-235: Does this mean the dataset is heterogeneous in space and time? If that is the case, how does this affect the evaluation? Please explain.

- line 282: It would be useful to give the equation for R_anom too. Does this includes both space and time dimensions?

- line 291-292: Is the ensemble spread also lead-time dependent?

- Section 3.2: Since the evaluation metrics are based on anomalies, what purpose does this section serve in the paper?

- Figure 4 and others: Since the gridded model forecast is spatially averaged over the large domain with different masks for different variables, it would be useful to show the masks for these variables in Figure 1 so that readers know how the spatial average is calculated.

- Section 3.3: This section is about the absolute error. Because of the seasonality discussed in the previous section, it is not surprising that errors are generally larger during the season when the absolute value of variable is also large. So it will be necessary and more informative to discuss the relative errors beyond the absolute error.

- Figure 6: For each panel,the y-axis should be set to the same range as that in the corresponding panel in Figure 5.

- Line 489-490: The results in this study do not seem to back up this statement.

- Line 518-520: This statement is speculative. It would be more appropriate to provide justifications.

- Line 529: How is 4% cold bias calculated? Using different units such as Kelvin, Celsius will certainly result in different percentage change? So a statement like this does not make much sense.

- Line 603-604: This statement assumes that the ensemble spread of the forecast is informative. The assumption may or may not be true. Linking a smaller forecast spread with higher skill is unjustified and questionable.

- Line 633-634: It is not clear how this study achieve this as it does not provide much insights that can guide model improvements.