Final Authors Comments to Reviewer 1

Comment on egusphere-2022-449
Anonymous Referee #1

Referee comment on "Seasonal forecasting skill for the High Mountain Asia region in the Goddard Earth Observing System" by Elias Charbel Massoud et al., EGUsphere, https://doi.org/10.5194/egusphere-2022-449-RC1, 2022

Review of Massoud et al. entitled “Seasonal forecasting skill for the High Mountain Asia region in the Goddard Earth Observing System”

General comments:

This manuscript evaluates the seasonal forecast skill for hydrometeorology over the High Mountain Asia (HMA) region from the Goddard Earth Observing System (GEOS). As the author suggests, “S2S forecasting for HMA is in its infancy”. Their results show that the GEOS-S2S system’s ability to forecast HMA hydrometeorology on the seasonal timescale is limited. The authors raise some issues of the GEOS for seasonal hydrometeorological forecasts. These results help to improve the ability of the seasonal forecast model in the future. Therefore, the scientific questions are of interest. Their introduction provides context and objectives for their work, which catches the reader's interest. The data and methods are described in detail and are reasonable. The results of the evaluation and discussion are well written. In general, this paper is well prepared and fits the scope of ESD.

Author Comments: We very much thank the reviewer for the time spent on our manuscript and for the generous and kind description of our paper. Our comments below indicate where we plan to make changes in the manuscript to satisfy some of these concerns.

Specific comments:

1. The credibility of verification data is a great challenge. Even though the authors use multiple data, I still think the current results are quite uncertain due to credibility of verification data.

Author Comments: We thank the reviewer for this insight. This is the reason for using multiple sources of observed and verification data in our study. Relying on solely one source of information may be misleading, hence we used 2 different products for each climate variable to get a sense of the uncertainty in the forecast skill for each variable. We will include a statement in Section 2.3 to state this clearly, “The datasets used for evaluation in our study have their own biases and issues, particularly over the mountainous regions of our study.”

2. The S2S (subseasonal to seasonal) prediction project database

(http://www.s2sprediction.net/) provides reforecasts by many operational forecast systems. Can the forecasting skills of the GEOS be compared with models that participate in the S2S prediction project?

Author Comments: We agree with the reviewer that this would be an interesting and useful comparison. GEOS-S2S does participate in the North American Multi-Model Ensemble (NMME, https://www.cpc.ncep.noaa.gov/products/NMME/). We will provide some literature review in our paper describing the reported skill of the GEOS-S2S system in previous studies, with the goal of showing GEOS-S2S as a state-of-the-art system. We will include comments in the discussion that refers to this.

3. It appears that the area of focus includes some low-elevation areas within the range of Figure 1 (e.g., parts of India). Are the authors calculating some statistics (e.g., Figure 2, Table 2) for the entire area of Figure 1? Should low-altitude areas be masked out?

Author Comments: The reviewer makes a good point; different regions may have different skill metrics. Our results in Figures 5-12 show maps of the skill and the ensemble spreads for the whole domain, which can be used to analyze spatial differences in the results and to pinpoint specific regions with higher or lower skills in general. Figures such as these can be used to interpret the skill level for different regions within the domain, such as low- or high- elevation regions. For example, Figure 7 shows that there is a higher ensemble spread as well as a higher error for temperature in the India subregion compared to other regions in the domain. To help further satisfy this concern, we will provide new plots showing the evaluation metrics for smaller sub-regions within the domain.

4. Section 3 describes the results in detail. However, there seems to be a lack of an indepth scientific explanation. For example, what are sources and effects of the forecast errors.

Author Comments: Section 3 is intended to provide a thorough quantitative analysis of the results. The more in-depth scientific discussion is in Section 4, which digs deeper into the reasoning and qualitative analysis behind the results of Section 3. For example, Section 4.1 discusses the role of model initialization as well as the persistent memory of the physical system and the impact these characteristics have on the skill, and Section 4.2 discusses the role of model resolution or the representation of different processes within the model and how these characteristics impact the skill.

5. I noticed that the skills of GEOS vs. MERRA-2 and observation are quite different (Figure 2a vs. 2b). The ubRMSEs in Figures 5d and 6d show the issue. Which result should I believe? Why are there such obvious differences in skill when using different verification data (especially SM, TWS)? How do the authors interpret such differences of results?

Author Comments: We agree. There are distinct spatial and temporal differences in GEOS-S2S skill when compared to MERRA-2 or other observations. There are several reasons that these

differences arise for different variables. For example, the fact that GEOS-S2S and MERRA-2 are close in their architecture makes it necessary to have other data for verifying our results, and this is explained in more detail in Sections 4.1 (Lines 475-477, Lines 485-487) and Section 4.2 (Lines 538-555). We will make changes to the text that highlight the potential causes of these differences. In addition, we will add a line to the conclusions explaining our relative confidence in the forecast skill.

6. High anomaly correlation or low ubRMSE indicates better forecasting skills. Both the anomaly correlation and ubRMSE represent the correspondence between forecasts and observations. It looks like it is acceptable to use just one metrics. Why use both anomaly correlation and low ubRMSE?

Author Comments: Although we agree with the reviewer that the ubRMSE and the anomaly correlation may be showing redundant information in some cases, there are cases in which having both evaluation metrics can be useful. While correlation is a non-dimensional metric that is invariant to changes in the mean and variance, the ubRMSE has units and is sensitive to the variance in the data. In general, most studies that report on a model evaluation show various evaluation metrics. Here, we simply report two metrics that cover complementary skill aspects to make understanding the results of the evaluation more accessible.

7. Section 3.2 and Figure 4: It appears that the annual cycles have large uncertainties, mainly hydrological variables. The anomalies are derived by removing the annual cycle. This might greatly affect the credibility of the results. How does the author address this issue? There should be an explanation.

Author Comments: We thank the reviewer for this insightful comment. As a reminder, the anomalies are created by subtracting the mean value for each variable in each respective month (e.g., mean January T2M subtracted from all the January T2M's, mean February T2M subtracted from all the February T2M's, etc.). Section 3.2 shows the annual cycle, which portrays how each variable changes through the course of the year. We see that for some variables, such as T2M and PRECTOT, there is more agreement in the annual cycle between the various products. However there tends to be a higher spread in the annual cycle of the other variables. Some of the reasoning behind this spread are explained in Section 4.2.2. For example, on Line 518-520, we state:
"*For SM and TWS, error patterns in Figure 5E and 5F and Figure 6E and 6F can primarily be related to monsoon representation in the S2S system, but the errors can also be associated with the observational difference in the seasonal cycles shown in Figures 4E and 4F*".

8. The reviewer did not get the point of Figure 3. This figure depicts the difference in skill between variables and between forecast lead times. Different variables have different predictability. Forecast skill decreases with forecast lead time as a matter of course. What is the purpose of comparing their relative skills?

Author Comments: We thank the reviewer for this comment. This figure was included to visually depict the difference in skill between variables and between forecast lead time. This kind of figures makes it easier to directly compare results between the relative performance skill of the variables/lead times. The usefulness of Figure 3 in this paper is that in Figure 2 we show anomaly correlation as the main metric to visualize the differences, whereas Figure 3 uses ubRMSE as the main metric. When a specific box is blue, it means that for that variable and at that lead time the skill is higher than that of the skill for that variable at the other lead times, and when a box is red that means the skill is lower for that variable at that lead time compared to the skill for that variable at different lead times. We will revisit and clarify the relevant text in response to this comment.

Minor comments:

Line 23 and 25: "ranges" shoud be "range".

Author Comments: We thank the reviewer for this comment. We will fix this in the paper.

Line 118: "…five mountain ranges, including the Himalayas, Inner Tibetan Plateau, Karakoram, and Hindu Kush." Shoud be "four"?

Author Comments: We thank the reviewer for this comment. We will fix this in the paper.