Reply to referee #1

"This paper evaluates the ECMWF seasonal forecast system (S4) on the global meteorological drought forecasts. The dynamical forecast shows advantage against climatological forecast. I am impressed by the detailed supplementary materials. The paper can be published in HESS after addressing a few comments below"

The authors wish to thank the reviewer for the positive comments. We have addressed in the following the reviewer comments. A point by point response follows.

" 1) My big concern is the use of two datasets (GPCC and ERAI) to *quantify* the influence of initial conditions on the drought forecast. I agree such comparison can reflect the uncertainty to some extent, but it can not support the conclusions such as "The memory effect of initial conditions was found to be 1 month lead time for the SPI-3, 3 to 4 months for the SPI-6 and 5 months for the SPI-12". To rigorously investigate the effects of initial conditions, I suggest using climatolgical mean (GPCC) during the antecedent period (plus following S4 forecast) when calculating SPI, and compare it with the result from GPCC S4. Then we may have a sense how an initial anomaly affect the SPI forecast. Otherwise, I suggest the authors change their statement about the effects of initial condition throughout of the paper (e.g., P931 L1, P935 etc). The evidence shown in this paper only demonstrates the impacts of differences in the two datasets on the forecast. "

We thank the reviewer for calling our attention to this point. We agree that comparison that we made in the paper only allows to measure the impact of the two datasets (GPCC and ERAI) in the forecasts, and the generalization to the "memory effect" should not be done.

Following the reviewer suggestion, we generated a new configuration of seasonal-reforecasts using GPCC climatology for the monitoring and S4 for the forecasts (GPCC_CLM S4). We produced a similar figure 2 (in the manuscript), by computing the ROC skill score of GPCC S4 using GPCC_CLM S4 as reference, and is presented in the end of this reply (Figure R1). For the SPI-3 and SPI-12 both figures show similar values for the effect of initial conditions, but for the SPI-6, ranking GPCC S4 with GPCC_CLM S4 gives a consistent value of 4 months globally while when ranking with ERAI S4 we had values between 3 and 4 months. These results show that in regions where ERAI has a similar evolution to GPCC we have higher forecast skills than using climatological initial conditions. This new result will be included in the manuscript, and discussed in the text to clarify the "memory effect" of the initial conditions in the forecast skill.

" 2) Following the above comments, I think the statement like "ERAI has higher RMS errors" (P929, L11), which using GPCC as the truth, is unfair for the ERAI. How do we know there is no uncertainty in GPCC, especially for sparsely gauged area like Africa? It seems that the authors do not compare apples with apples. I mean, you may expect less RMS for ERAI if using ERAI as reference. Actually, global meteorological drought forecast has many issues, and the uncertainty of observation is one of them. I think the authors can make their points along this line."

We agree with the reviewer comment. The companion part 1 paper discusses in more detail the impact of different datasets in drought monitoring along with the uncertainties of GPCC, in particular the reduction of rain gauges used in GPCC during the last decade. In this part 2 we decided to assume GPCC as the reference dataset for drought forecasts verification, therefore the statement "ERAI has higher RMS errors". If we would verify the forecasts against the SPI calculated with ERAI the results would be different, but this would make the manuscript longer and difficult to follow. In that paragraph, a new sentence was added"

"The forecasts using ERAI for monitoring are penalized when compared with the forecasts using GPCC for monitoring, since GPCC is used as reference dataset (for the forecasts verification). These results do not consider the uncertainties in GPCC that are discussed in more detail in the companion part 1 paper."

" 3) P926, equations (3) and (4). Any reason using such ACC instead of the common one: correlation both in space and time? Any reference?

It is a common practice to calculate the temporal mean of the spatial anomaly correlation, as the reviewer suggests, for upper air fields such as the geopotential height, and when evaluating medium-range forecasts. In this study, we follow the methodology used to verify S4 forecasts (see table 4.2.1 of Molteni et al. (2011)) by computing the area-mean of the grid-point anomaly correlations.

4) About the comments on whether global drought onset is a stochastic forecasting problem (last sentence in the paper), I think what Yuan and Wood (2013) want to illustrate is the low skill averaged globally. For example, if the GPCC S4 has hit rate as 0.3, but false alarm ratio as 0.5 (Table 2), that means if 10 droughts in the historical, the model give you 6, but 3 of them are false alarms. I do not think Yuan and Wood (2013) neglect the fact that over some region, the models are quite skillful."

We thank the reviewer for raising this point. Our results are coherent with the findings of of Yuan and Wood (2013), but their conclusions might sound contradictory.

The sentence: "Yuan and Wood (2013) begged the question as to whether seasonal forecasting of global drought onset was largely or solely a stochastic forecasting problem only. However, our results show that this is not a global result: within several regions in the world drought onset forecasting is feasible and skilful."

Was changed to:

Yuan and Wood (2013) raised the question as to whether seasonal forecasting of global drought onset at local scale (for 1 degree grid cells) was largely or solely a stochastic forecasting problem. Our results are coherent with the findings of Yuan and Wood (2013), but when the grid cells are pooled into a region, our regional analysis highlight that, within several regions in the world, drought onset forecasting is feasible and skilful

References:

Molteni, F., Stockdale, T., Balmaseda, M., BALSAMO, G., Buizza, R., Ferranti, L., Magnunson, L., Mogensen, K., Palmer, T., and Vitart, F.: The new ECMWF seasonal forecast system (System 4), ECMWF Tech. Memo., 656, 49 pp, 2011

Yuan, X., and Wood, E. F.: Multimodel seasonal forecasting of global drought onset, Geophys. Res. Lett., 40, 4900-4905, doi: 10.1002/grl.50949, 2013

Figures

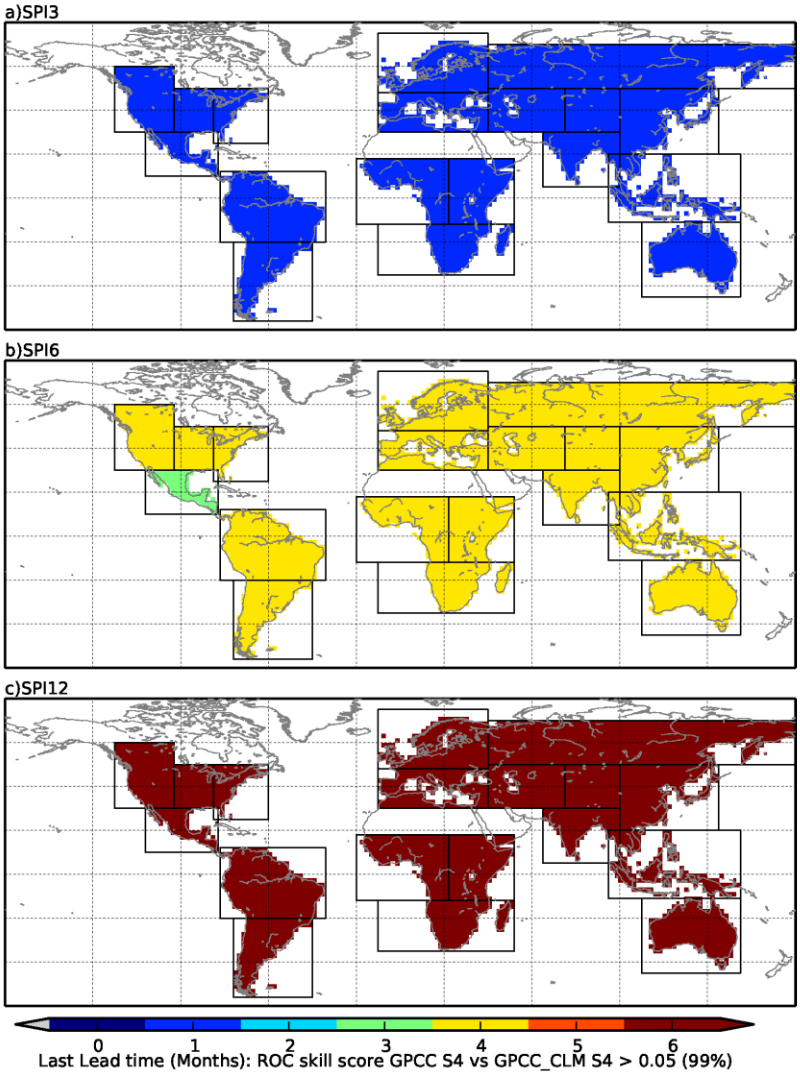

Figure R1. Last forecast lead time (months) where the ROC skill score of GPCC S4 (using GPCC_CLM S4 as reference forecasts) is higher than 0.05 with 95% confidence and the ROC of GPCC S4 is higher than 0 with 95% confidence. Seasonal forecasts of the (a) SPI-3, (b) SPI-6 and (c) SPI-12. The forecasts are verified in each region for the calendar month presented in Table 1 (in the manuscript).