Anonymous Referee #3

This paper analyzes the monitoring and short-term forecast of standardized precipitation index spatially-averaged over 3 medium-sized river basins (Niger, Blue Nile, Zambezi and Limpopo) in Africa. The SPI is computed on gridded datasets and also reanalyses from ERA interim and coupled seasonal predictions from Hadley Center S4 system. A main conclusion is that monitoring and short-term forecasts are difficult and spatially variable across Africa, partly because of the difficulty to measure the true rainfall and to get it in realtime. In that context, using the short-term forecasts from S4 could be a valuable alternative. It is an interesting study but I found that some paragraphs are unclear and also that authors do not pay enough attention to significance of some statistical scores and their confidence intervals.

We thank the reviewer for the important suggestions, in particular on the importance of a robust statistical evaluation. We believe that this helped improving the results interpretation. Below follows the detailed reply for the comments. Just a remark on the reviewer summary: we are using seasonal forecasts from the ECMWF S4 system and not from Hadley Center.

Two major concerns

1. The authors used the GPCP as a benchmark for the truth. We know that monitoring rainfall variability across the tropics is a difficult task and such gridded dataset merging rain-gauges and satellite estimates is probably a valuable choice. But the fact is that we have no estimate of the error between GPCP and the true rainfall. Is there a simple mean, using for example few stations from GHCN dataset to get an empirical estimates of the SPI computed on stations and thus validate someway GPCP for the 4 basinsÂ˜a? In the same context, I am intrigued by the correlations between CAMS-OPI and GPCP (Figure 5b-e) since it is rather noisy. Are the correlations related to the number of rain gauges included in CAMPS-OPI (highest values when a large number of rain-gauges are included both in CAMS-OPI and GPCP ?): if so, can we be really confident that GPCP is a better product (closer to ground truth) than CAMS-OPI ? In summary, I think that the authors should provide more information to illustrate the quality of GPCP as a benchmark for the true variations of rainfall and SPI. If it is not possible to do that at large scale, it is perhaps possible for the 4 river basins emphasized in this study.

The reviewer suggestion of estimating the GPCP error using some stations from the GHCN dataset is very interesting. However, such estimation would require a significant amount of work, and GPCP already provides an error estimate of precipitation (displayed in Figure 2 along with the number of stations used).

The noisy correlations between CAMS-OPI and GPCP in figure 5 are mainly due to the different number of stations used in each dataset and different processing algorithms. For example, in the 4 analysed basins, CAMS-OPI was generated with a lower number of stations, when compared with GPCP (see Figure 2). This is also evident in the temporal correlations in Table 2, and in the new results of Table 3, in particular the significant reduction of the temporal correlations between GPCP and CAMS-OPI in the ZB basin from the first to the second half of the record. In fact, Figure 2 shows that from 1999 onwards there were no stations used by CAMS-OPI in the ZB basin.

It is beyond the scope of the manuscript to address the reviewer concern in this publication in detail, but we added new results in table 3, discussion and added text in the conclusions in order to highlight the uncertainty in the precipitation datasets that constrains drought monitoring and forecasting in Africa.

2. A second major concern is more statistical: you discuss almost all results about differences (for example in correlation with GPCP), without really considering the confidence interval which should be rather large for your rather short records. I think that you need to be perhaps more balanced about your conclusions and the fact that one product is better than another when the difference is in the confidence intervals of both correlations. So you need to pay attention to this source of uncertainty as well as differences between the products.

We thank the reviewer for stressing the importance of a proper statistical evaluation of the different analysis. We have removed some of the results that were uncertain (see general reply to comments), and addressed the

reviewers concerns by clarifying several sentences according to the statistical significance of the results (please see below).

More minor concerns

- title is not fully adapted to the study. I think you should clearly state that you use SPI to monitor and forecast drought since it is a well-known index and also because there are other alternatives.
The title was changed from "Seasonal forecasts of drought indices in Africa basins" to "Seasonal forecasts of droughts in African basins using the Standardized Precipitation Index"

It makes also sense to state that you consider wery short lead times (mostly 0-month and 1-month delay) and it is not a study of classical seasonal prediction.
We do not understand this comment. Throughout the manuscript we always considered the full range of the seasonal forecasts length: 0 to 5 months lead time (e.g. Figs: 7, 8, 9, 10 and 11)

- There are instances where a significance level is included as in Figure 6. You need to assess systematically the significance of the correlations all along the text (Table 2, Table 3, Figure 5) and restrict your comments to significant values.
In Figure 5 the white colours denoted statistically insignificant ACC. Clarified in the caption. Table 2 includes the 95% confidence intervals, and table 3 was removed (see general reply).

- In the abstract, state clearly that the "integrated drought" index is indeed SPI. I found that several parts of sentences could be more precise; for example, what is exactly the sense of "temporally extending", "all the datasets show similar patterns" (patterns of what ?), "in the tropical region" (do you mean equatorial area ? It is perhaps better to state clearly a band of latitudes or a specific area by its name). "larger time-scales" is perhaps less appropriate than "longer time-scales".
The different points were corrected/clarified in the abstract.

- line 18 page 11098: "including precipitation provided every 3 hours"?
Corrected

- line 23 page 11103: GPCP version 2.2.
Corrected

- line 26 page 11104: Zambezi and Limpopo are more in southern Africa than in Eastern Africa that usually refer to the horn of Africa till Kenya/Tanzania or so. In the same sentence it is better to refer to "Austral summer" rather than to "Boreal summer" for the wet season in these basin rivers.
Corrected

- Line 1 page 11105: do you have a precise reference to corroborate this statement (dipole between western and eastern (or southern) Africa)?
We were referring to the rainy season and not interannual variability. This was clarified.

- Line 19 page 11105: I don't what you mean by "a better intra-seasonal ...". please add a specific reference for that topic since there are no real illustration in this paper.
"intra-seasonal" was removed following the confidence intervals added in Table 2, and the sentence was rephrased.

- Table 2: add a significance level. Can you speculate on the decrease of the correlations from 3-month to 12-month integration for ERAI and CAMS-OPI for Niger and Blue Nile basins?
All correlations in table 2 are significant at 99% ($p<0.01$), and 95% confidence intervals were added. The decrease of the correlations with increasing SPI time-scale are associated with problems in the inter-annual variability of ERAI and CAMS-OPI that are mainly visible when accumulating precipitation on long time-scales.

- Page 11106: you use discharge for Niger and Zambezi as an independent measure of SPI-12. You state that the agreement is reasonable for Niger and lower for Zambezi. I disagree with this specific statement since (1) you have a shorter record for Niger (seen from figure 4a and c) and (2) it is well-known that Niger has more decadal and longer variability, thus inflating the correlation. I am not sure

at all that a correlation of 0.65 (Niger) could be considered as significantly larger than a 0.54 (Zambezi) one considering these two facts. In that context, I think that stating "a much lower correlation" for Zambezi is exagerated.
See general reply.

- Page 11106: "CAMS-OPI tends to have a better performance than ERAI". This statement seems weird when we look at the values of Table 2 and if we consider the confidence interval for such length of record.
Sentence was removed.

- Figure 5: add a significance level. Can you speculate about the fact that SPI-12 does not give higher correlations (in mean) than SPI-3. We expect that a longer time integration would remove some of the noise/uncertainty but it seems not to be the case.
The longer accumulation periods will be more sensitive to inter-annual to decadal changes in precipitation, and both ERAI and CAMS-OPI have limitations (when compared with GPCP) in most of the African continent.

- Page 11109 line 15:I don't really understand the part "by the accumulated skill of the S4 precipitation forecasts".
Sentence changed

- Page 11110 alinea 1: It is better to clearly state a band of latitudes or a named region rather than "outside the tropical region" :
Changed to: "Drought monitoring in Africa with ERA-Interim has limitations in the central equatorial region".

- Page 11111 line 5: "climatology" instead of "climate"
Changed

Typos

- In introduction, "Mckee" should be "McKee"
corrected

- Line 12 page 11104: "except"
corrected