S. M. Vicente Serrano (Referee) #2

Review of: Seasonal forecasts of drought indices in African basins. This is an interesting study showing the results of a drought forecasting system applied in Africa using precipitation outputs from the ECMWF. I find the results very useful since they stress the difficulty of obtaining reliable datasets for the climate monitoring in Africa and how the climate forecasting outputs are within the range of validity of the currently used precipitation datasets. The main drawback of the manuscript is a very poor conclusions section. I think that the authors should stress and discuss in more depth the limitations existing for drought monitoring and forecasting in Africa, and how the available information is seriously affecting the accurate assessment of the current drought forecasting products in the region.

I recommend the publication of the article in HESS, but I would suggest some issues that authors could include in the revised manuscript to improve its quality.
We thank Sergio Vicente-Serrano for his positive and encouraging review. A paragraph was added to the conclusions (and also section 3.2 - following reviewer #1 comments) highlighting the limitations in monitoring and forecasting droughts due to the available precipitation data.

11095. Line 15. Together WMO 2009, the authors should cite Hayes et al. (2011) Bull. Amer. Meteor. Soc., 92, 485, in which it is recorded the selection of the SPI as a reference global drought indicator.
Added

11096. Line 3. The recent study by Vicente-Serrano et al. (2012). Earth Interactions 16 (10) shows a global empirical example of the comparison of different drought indices.
Added

11096, Line 24. ECMWF must be defined the first time it appears in the text.
Changed

11097. Line 20. Please, cite Guttman (1999) J. Amer. Water Resour. Assoc., 35, 311–322., which directly focused on this issue.
Added

11099: An assessment of the influence of the three different precipitation datasets on the SPI would be really useful. Although authors include correlations and maps in Figs 4 and 5, some SPI regional time series for the entire Africa and the four analysed regions would be welcome.
The time series of the SPI-12 for each different basin from the different datasets are represented Figure 4, and the time series for the SPI-3 and SPI-6 were added to the supplementary material in figures S1 and S2, respectively. In addition to this, the new information in table 2 that includes the 95% confidence interval of the correlation between the GPCP SPI against the remaining datasets (for each basin and SPI time-scale), shows that in most of the cases, ERAI and CAMS-OPI are very similar, and S4L0 has significantly higher correlations than ERAI/CAMS-OPI in the NG/BN for the SPI-12, lower correlations in all SPI time-scales in the LP, and all three datasets compare similarly to GPCP in ZB.

11102. It is not clear how the verification metrics are obtained. What the authors are comparing? Predicted vs. observed SPI? Over which period?. This may be clearly stated in the methods section. Although in the results section the readers may see clearly what is the approach followed, the procedure must be clear from the beginning. A reference where to find these indicators or the equations would be welcome since they are not commonly used.
The details of the use of each metric follow the results presentation, since they depend if we are evaluating precipitation or SPI and also the initial forecast initial date and lead time. Section 2.3 was changed to include more details on the scores and pointing to references to their calculations. The ACC is widely used, the ROC and REL are recommended by WMO for the verification of probabilistic long-range forecasts, and the CRPS is an extension of the brier score falling to a mean absolute error in case of a deterministic forecast.

11105. Reference on the GRDC and more details on the quality of this dataset is also required. The streamflow data, as the climatic information, is not free of errors. Authors should be aware that several

perturbations (damming, water management, land cover change, etc.) may affect streamflow. Therefore, the streamflow data cannot be considered as the true reference for the validation of the climatic datasets. A discussion of this issue is needed.
This was removed from the manuscript, see general reply.

Table 3. Independently of the dataset used, correlations are low between streamflow and precipitation. It stresses the poor quality of the climate datasets currently available in Africa and the reliability of using this data for drought forecasting. Authors should stress this issue in the discussion and to state the need of a strong investment on climate monitoring. On the contrary, forecasting may be largely constrained. Figure 5 clearly shows how correlations tend to be higher in areas with better climate networks (Maghreb and South Africa).
Table 3 was removed from the manuscript, see general reply. The quality of the observations is now further included in the conclusion, also following the comments of reviewer #1.

Fig 4. How the discharge series were standardized? This issue is a bit more complex than for the precipitation series. See Vicente-Serrano et al. (2012). Journal of Hydrologic Engineering.
See *general reply*.

Figure 6. The white cells must be detailed in the caption. Do they correspond to nonsignificant coefficients?
Clarified in the caption.

Figure 7. Given the low agreement between the precipitation datasets and the uncertainty in observational data and predictions, I have doubts on how the forecasting showed at 12-month SPI is reliable or not, given the strong memory of the index. For the 3-month SPI the forecasting model shows a lower reliability, but probably as a consequence of the lower memory of the 3-month time scale. It would be interesting to compare and/or discuss these results with a simple statistical model based on the SPI transitional probabilities (e.g., Moreira et al., 2008, J. Hydrol 354: 116) to check whether the dynamic model shows a relevant improved skill in comparison to autoregressive models and/or average precipitation values.
In figure 7 (and Figs. 8, 9 and 10), the forecasts using the precipitation fields from the dynamical model (S4), are compared with climatological forecasts created by randomly sampling different 15 years (same ensemble size as S4) of GPCP (CLM). The S4 based forecasts are always benchmarked against CLM forecasts, showing the basins/lead times where the use of the dynamical model improves the skill in comparison with climatological based forecast. This is detailed in the beginning of section 3.3.2.