#Reviewer 2- RC2

We want to thank referee 2 for the critical review of our manuscript and for the positive words about our paper and its contribution. We have reflected on the comments and below are our point-to-point response to the questions raised.

Major comments

- First of all, the question of the uncertainty relating to the data used should be addressed. Since the research work deals with modeled, re-analyzed and gridded data, coming from different temporal and spatial aggregation scales and from different models, each data has its own uncertainty and in their combined use it is not possible to verify this question, not even the different uncertainties can be compared each other. GLEAM for example uses MSWEP as input, but what can we say about GloFAS? What rainfall input does it use? If different from MSWEP, how can we compare the modeled flow data with the rainfall data of MSWEP?

The source and quality of the data we use has indeed been a topic we discussed a lot and deserves more explicit inclusion in the manuscript. Our main objective was to work with observational data, or data that is as close to observations as possible for the entire region. For precipitation there were various good products available. We chose for MSWEP because that is consistent with GLEAM (avoiding accumulation of uncertainty resulting from different sources), which in turn estimates soil moisture using satellite imagery and a re-analysis approach. The MSWEP precipitation data was also selected due to its demonstrated good performance across Western, Eastern, and Southern Africa. It has shown a strong correlation with in-situ observations and substantial agreement with CHIRPS precipitation data (has been popularly applied in the region because it has been found to show a good depiction of rainfall seasonality, and in a study by Musie et al., (2019), it was also found to capture daily and monthly streamflow simulation). MWEP has better results when compared to ERA-Interim precipitation data (which was originally applied in the generation of GloFAS river discharge data). These findings are reported in studies by Cattani et al., (2021) and Beck et al., (2017). We chose to not use ERA5 precipitation because the quality is not good and there are no rain-gauge data assimilated into the product.

For discharge, we initially also wanted to use observation data or a data product that is as close to observations as possible (like GLEAM for soil moisture). However, the spatial and temporal coverage of observed discharge data in the region is too low for our analysis. Therefore we decided to use modelled data for discharge, but there is no dataset available that uses MSWEP for precipitation input. GloFAS uses ERA5 Land total precipitation data from EMCWF, distributed by the hydrological LISFLOOD model. We did not want to use ERA5 Land precipitation because it has been found to highly underestimate/overestimate the precipitation values in the region. Fessehaye et al., (2022) tested the product for Eritrea region and found it highly underestimated precipitation values. Gleixner et al., (2020), on the other hand, tested the product against CHIRPS dataset and found it overestimated precipitation in East Africa (see a copy of figure 3). GloFAS is calibrated and evaluated against in-situ river discharge, but mainly for perennial rivers at mid-latitudes (Harrigan et al., 2020; Hirpa et al., 2018). When we compared GloFAS dataset with GRDC in-situ observations in the study region, we found that there often was a strong bias in absolute values (see figure 1), but that the anomalies (value divided by annual mean discharge) are much better captured (see below figure 2 for two different stations in Ethiopia). As we mainly work with relative indices for our drought study (either standardized, or with a relative threshold), the absolute bias is less of an issue in our application so we decided to use GloFAS. We will suggest

that we include these figures in supplementary material and make our reasons for the use of GloFAS and uncertainty generated more explicit in manuscript.

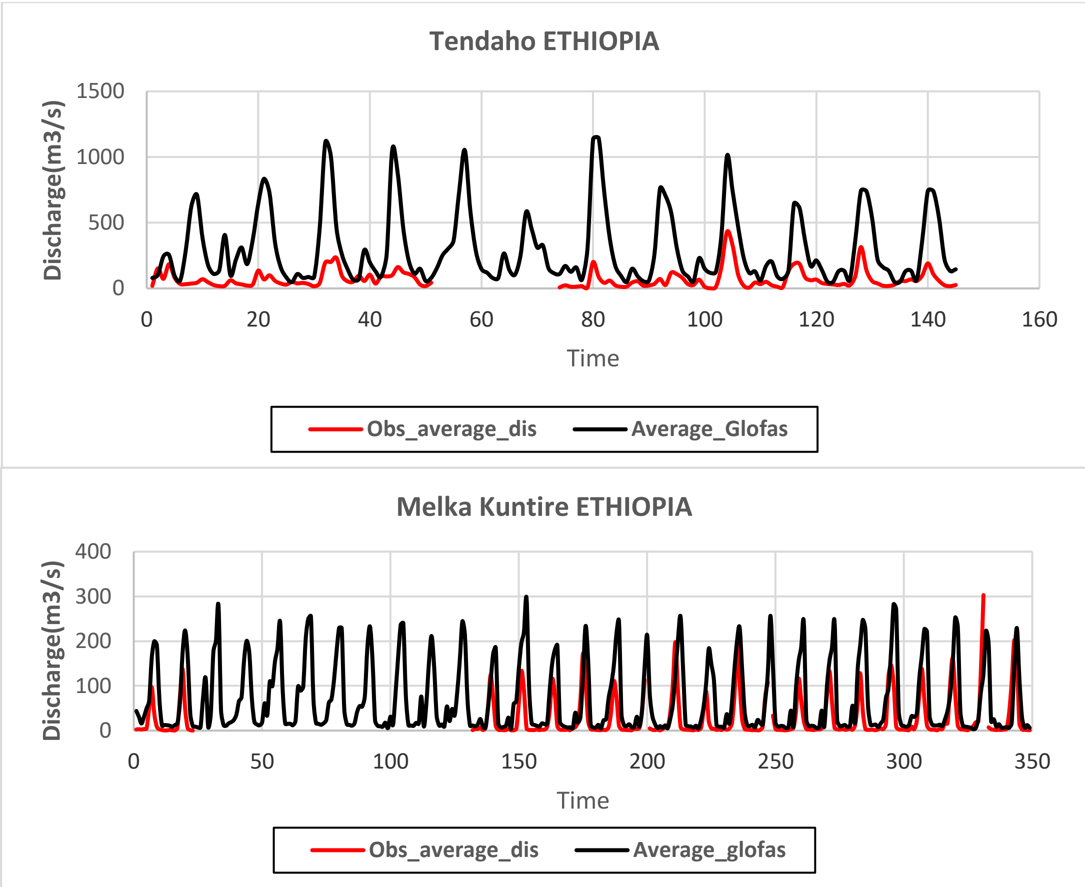
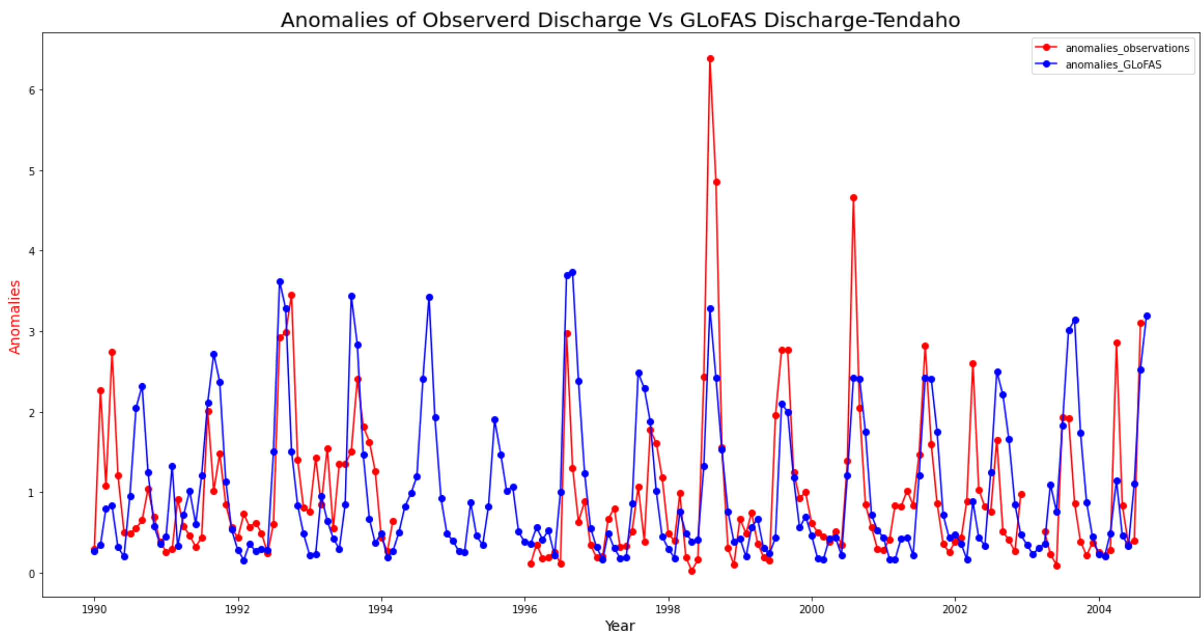


Fig. 1: Plots of GloFAS river discharge against observations



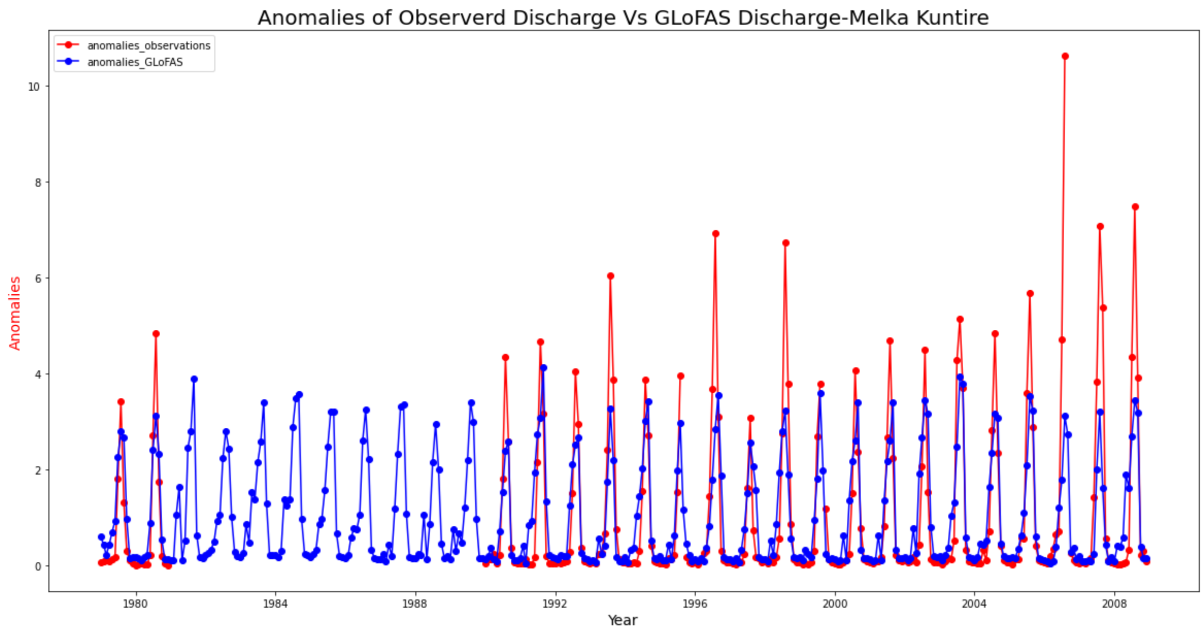


Fig. 2: Plots showing the discharge anomalies between the observed data and GLoFAS discharge. The deviations are similar as stated above

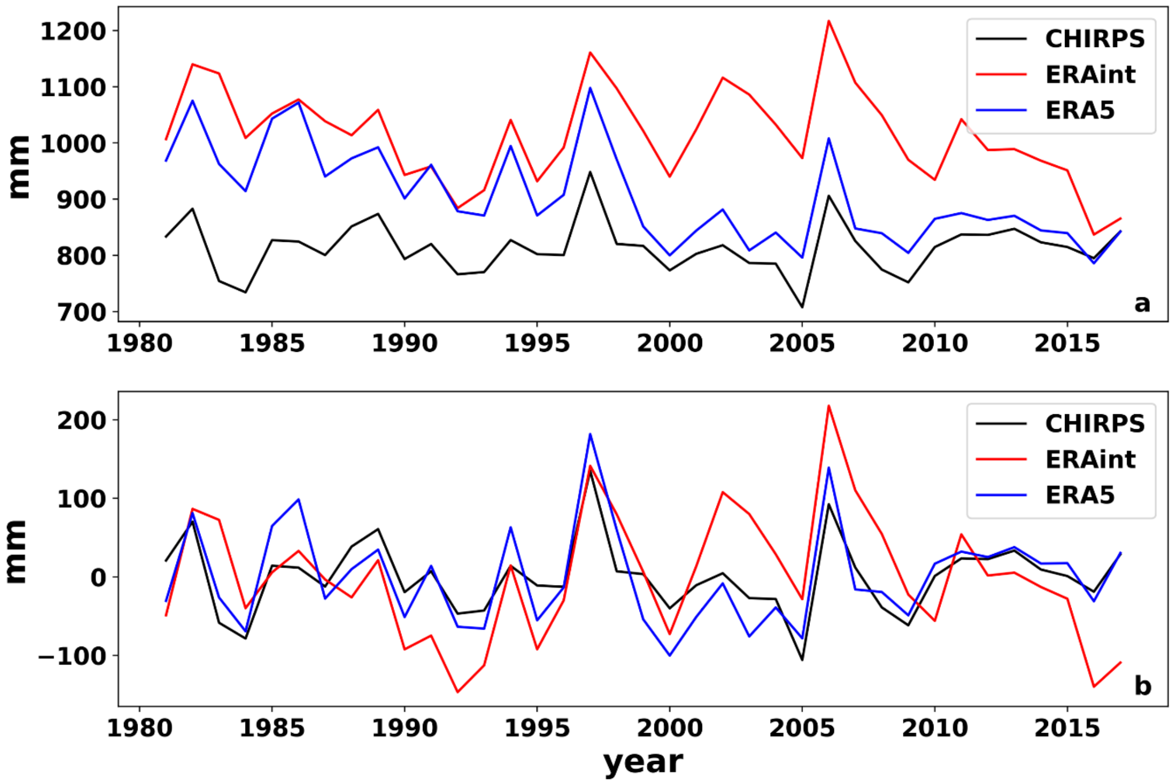


Fig. 3: Plots of ERA5 Land precipitation against CHIRPS dataset (Glexner et. al., 2020)

- Another source of uncertainty is also probably due to the fact that the different standardized indices (SPI, SSMI, SSI) are calculated for each basin with reference to different probability laws. How this would have impacted the analysis?

We indeed applied different distributions and picked the one with the best fit to calculate each of the indices. By doing this, we are not consistent between catchments, which cause uncertainties when comparing values between different catchments. As we are mainly looking to compare within catchments, we prefer the better fits to the spatial consistency. However, we only compare standardized values within catchments.

By the nature of the different indices, different distributions are best suited to fit the different data types. This is recognized in literature and as such we used the distributions suggested by Stagge et al., (2015) for calculation of SPI, distributions suggested by Ryu and Famiglietti, (2005) for calculation of SSMI and distributions suggested by Vicente-Serrano et al., (2012) for calculation of SSI. [we will explain in manuscript]

We agree that fitting a different distribution for each catchment causes some uncertainty, especially in the tails (i.e. drought extremes). However, this is only relevant if one is directly comparing drought onset or severity of the same event between catchments. In our case we do not analyze the droughts themselves, but instead we look at the propagation in the hydrological system. Therefore, we correlated the time series of the meteorological indices with several accumulation periods to the hydrological indices to fit the best correspondence. In our opinion, in this case the benefits of having the best fit to the data for each catchment outweigh the differences in extreme droughts between catchments.

Another way to verify our results is the comparison with threshold-based indices. Our propagation analysis done with the threshold-based duration ratios is very similar to that based on accumulation periods of standardized indices. This indicates that the distribution fitting is expected to have a minor role.

To give the reviewer more details about the distribution fitting, we here include some extra analysis. For example looking at the fitting of precipitation data for calculation of SPI-1 for the different catchments, we can see in the figures below that the depiction of the drought months is not different. This was similar for the other indices.

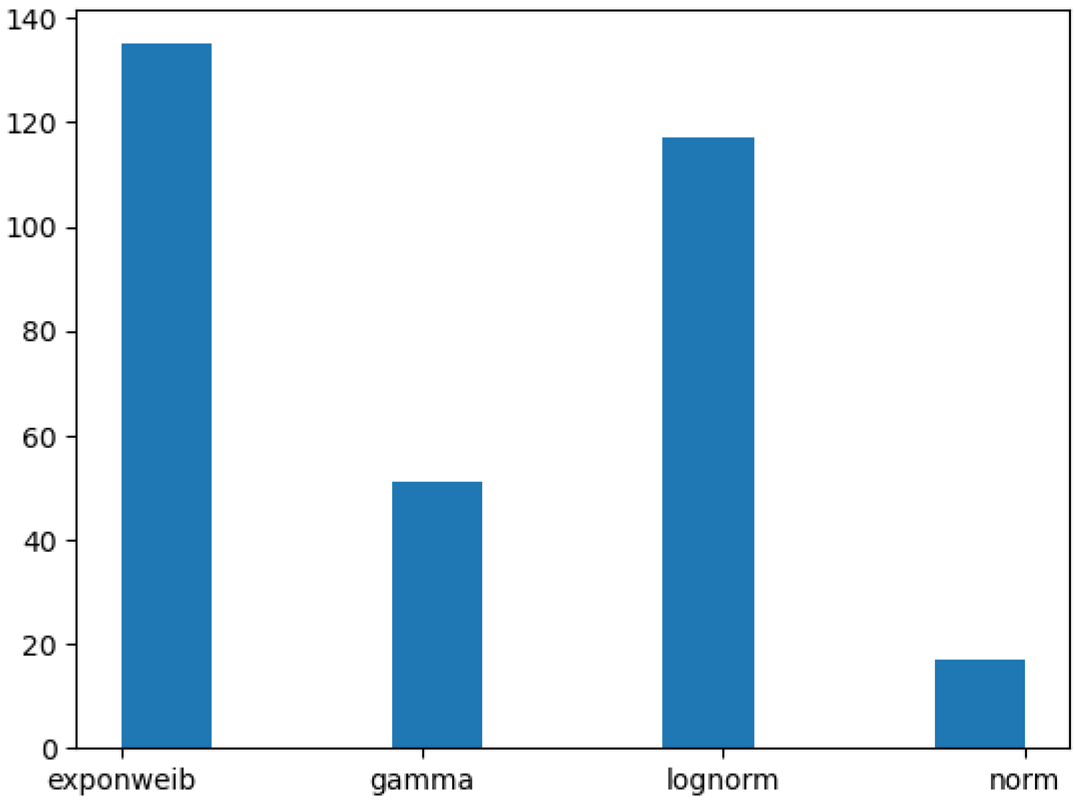


Fig. 4: Histogram showing the number of times each distribution was used in the calculation of SPI-1

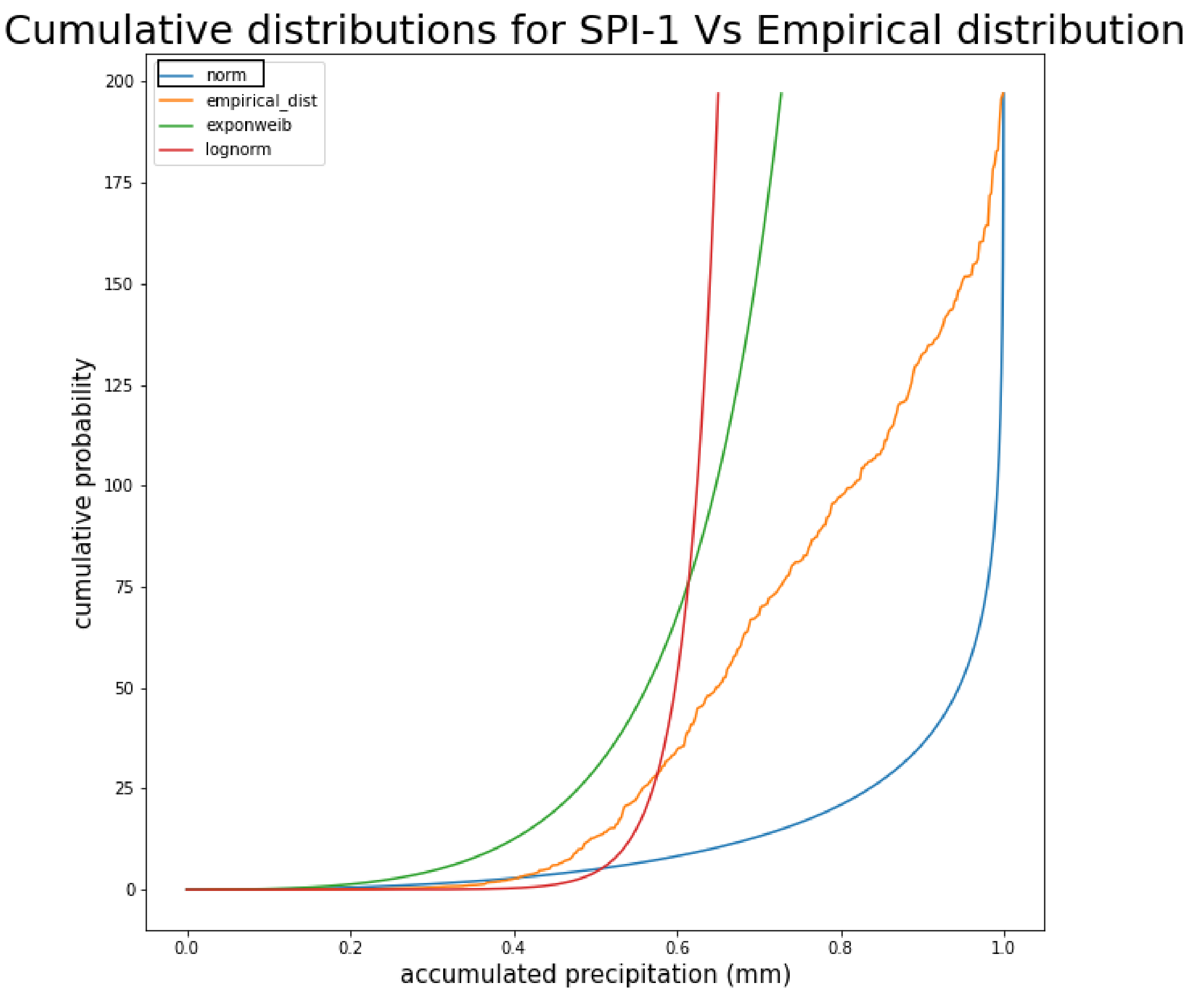
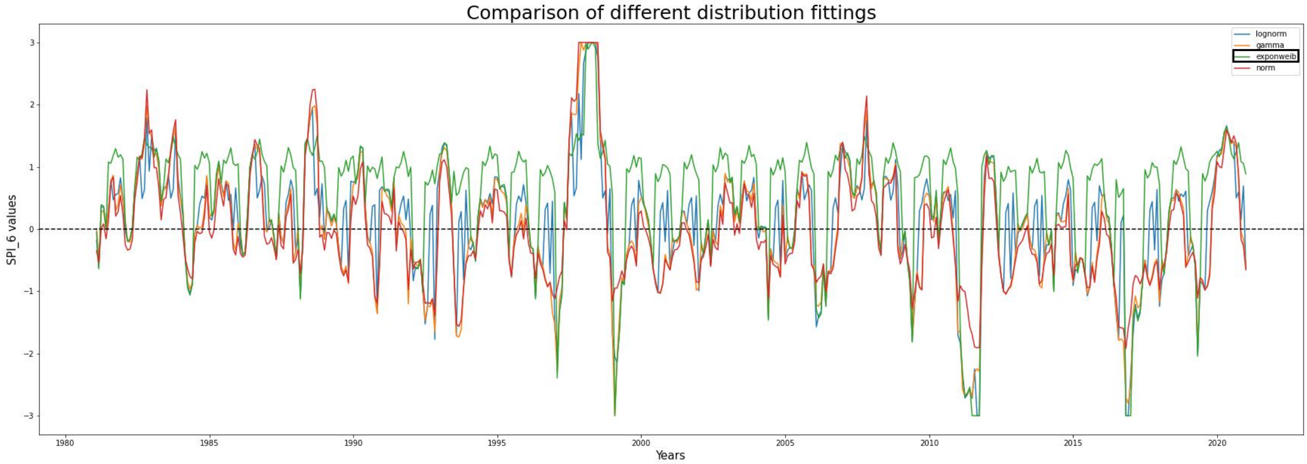


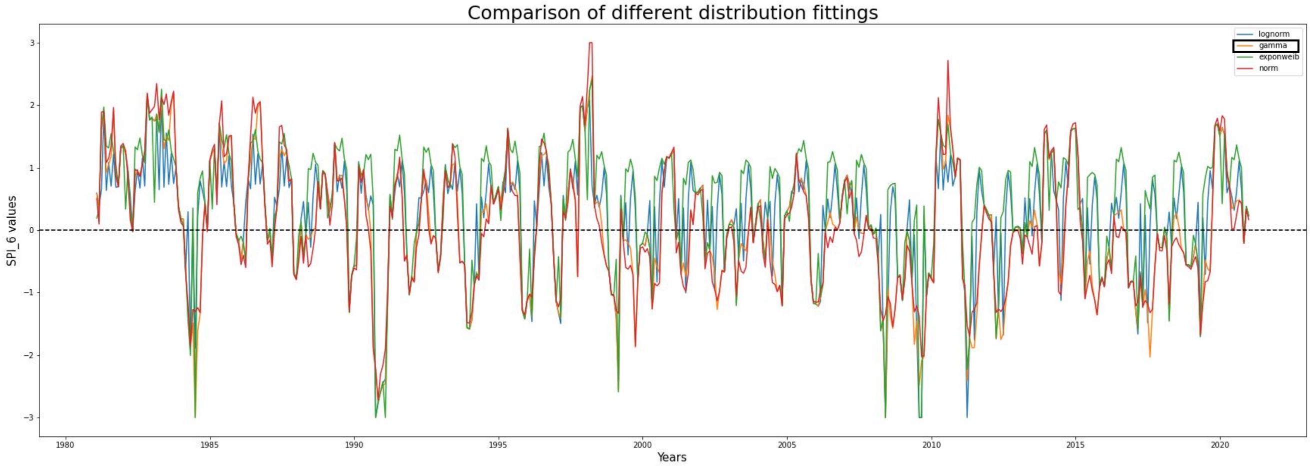
Fig. 5: shows the fitting of the distribution for the calculation of SPI-1-> selected best distribution is framed in black

The fittings for the distributions in the catchments in general were not very good as in figure 5, hence we decided to not force fit data for each catchment into the same distribution to avoid introducing artifacts. This is because if we do force fit a consistent distribution throughout the catchments, especially in dry catchments with lots of zero values, we anticipate worser fitting distribution and even worser results. Additionally, we are rather more interested in the catchment propagation than the spatial pattern existing among the catchments, because we would like to better identify the droughts accurately within each catchment.

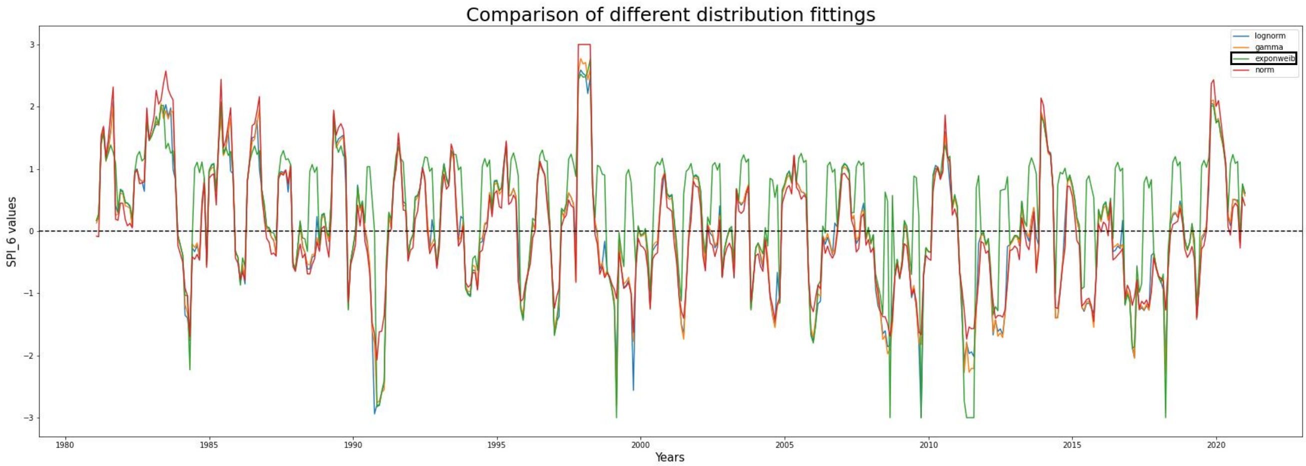
See below plots for four different catchments.



Comparison of different distribution fittings
lognorm
gamma
exponweib
norm
SPI_6 values
Years



Comparison of different distribution fittings
lognorm
gamma
exponweib
norm
SPI_6 values
Years



Comparison of different distribution fittings
lognorm
gamma
exponweib
norm
SPI_6 values
Years

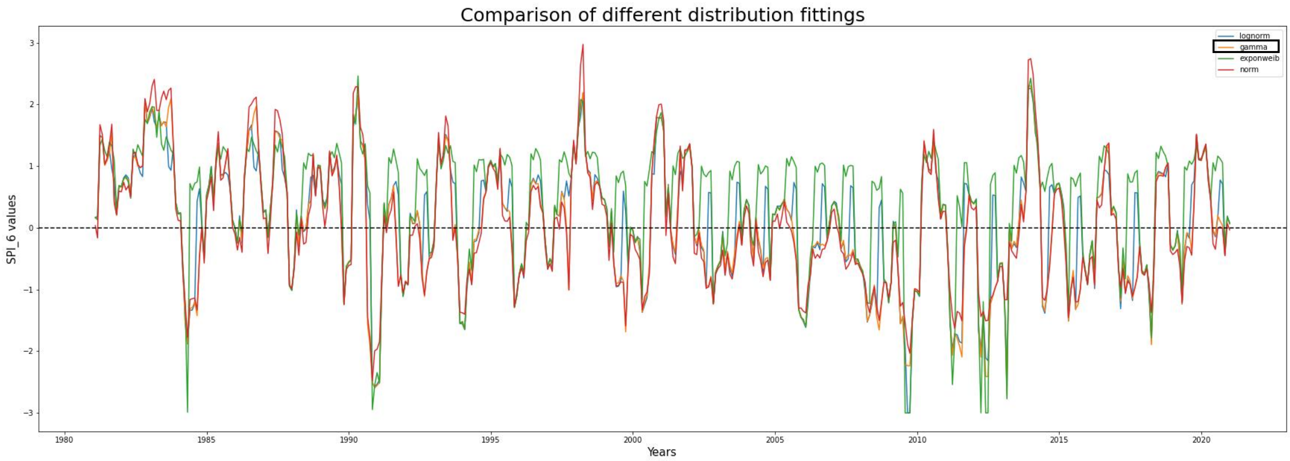


Fig. 6: shows the SPI-6 after fitting different distribution through the precipitation data for different catchments->selected best distributions are framed in black

There was a similar representation when looking at the SSMI-3 distribution fitting. There is very minor difference in the distributions (see below plot). Therefore, we concluded that the use of the different distributions will not affect the results in the long run. We suggest that we add a few lines about this in the discussion of the manuscript.

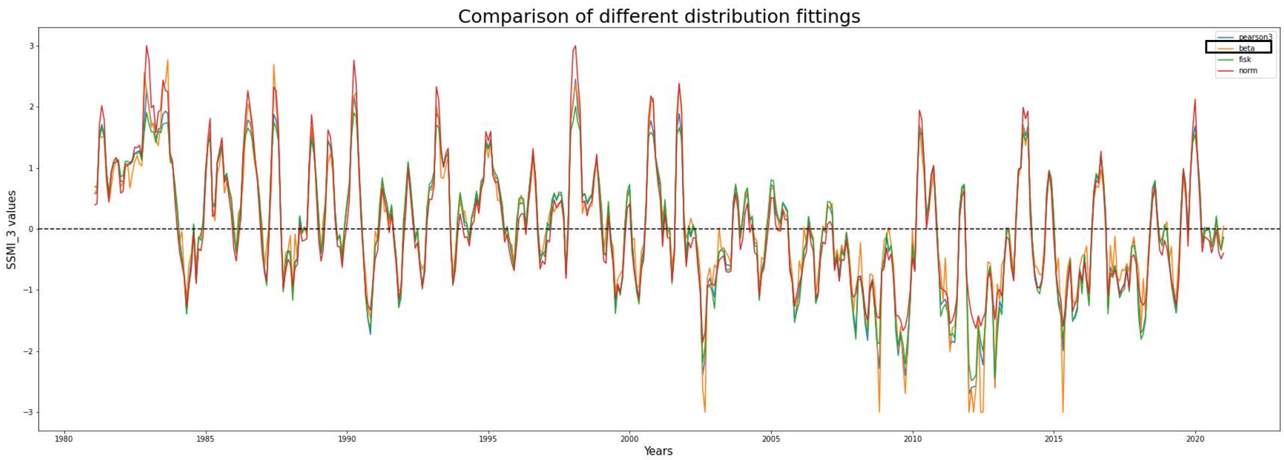


Fig. 7: shows the SSMI-6 after fitting different distribution through the soil moisture data ->selected best distribution is framed in black

- I understand that the research idea is the propagation of drought, but perhaps a risk analysis (see Figure 2), which also showed the quantitative evaluation and the relative characteristics of the indices, also from a spatial point of view, would have helped to sort out the problem. For example, on line 443 it is said, if I understand correctly, that the drought indices used "fail to capture the water deficit amount…" how can we know this if we do not have an illustration of the quantitative assessment of the indices?

Line 443 was based on a literature review on the advantages and disadvantages of using standardized indices against the threshold-based indices. Which index would be better able to capture absolute water deficits was not tested in this manuscript. Looking into drought characteristics themselves and their spatial representation was beyond the scope of our analysis. It would have made the manuscript substantially longer and would have taken away the focus from the propagation results, which in our view are more interesting to study. Additionally, as we discussed in our response to the previous comment, the spatial comparison of the drought

characteristics themselves has large uncertainties due to the fitting of different distributions in each catchment. We do agree to include the plots in the supplementary material for different catchments located in different places within the region of study for the readers, but would be hesitant to draw any conclusions from these.

- In general, the quality of preparation should be improved. Some sections of the manuscript were particularly difficult to read and sometimes repetitive. For example, section 3.3 says little more than 3.3 so maybe they could be merged. In general, the paragraph on methodology should be improved. In the same way, the presentation of figures (figure 5) and tables (table 1) in points of the body of the text where the analyses have not yet been inserted are misleading.

We will carefully read through the manuscript and remove repetitions while also merging the suggested sections.

- In general, even the figures should be improved in their quality and also in the formatting of the characters, which even when printed are too small for easy reading. I wonder why in figure 3 the same colours are always used except for panel (a) and the same for figure 6. Why don't change them or just have always the same colour? In figure 3 and 6 there are no units of measurement. Furthermore, in table 1 it is not clear to me (not even from the caption) what is the difference between the two SPI-to-SSMI (months) columns. The two columns have the same header but different values.

In Figure 3a represents the correlation values obtained during the propagation analysis while Figures 3b-e represent propagation indices maps, hence, the different colours. Same explanation applies to Figure 6. For this reason, we suggest to keep the same colour scheme in both figures for differentiation purposes.

We agree to include the units of measurements in both figures. We did try the suggested format for table 1 but realized this made it harder to read.

In Table 1, the different values are for each of the climate and catchment characteristics (the table has been split into 4 columns to reduce the length) . The values represent the mean aggregation of the selected indices of each catchment per each class of the climate and catchment characteristics. This was done to show the average selection in months of each of the indices per class. We agree to include a similar explanation in the manuscript for better understanding.

Minor comments

- Line 94: what is "level 6 boundaries of HydroBASINS?

HydroBASINS is a system that divides a large water basin into smaller sub-basins at points where two river branches meet and each has a minimum upstream area of 100 km². The sub-basins are also grouped and coded to allow for the creation of smaller sub-basins at different scales, or to move from upstream to downstream within the sub-basin network. To make this possible, HydroBASINS uses the "Pfafstetter" coding system, offering 12 nested sub-basin levels globally. We will expand the explanation of the selected level in the manuscript.

- Line 110: You probably selected the study region not because it is an "interesting" region but because, given the large number of catchments and relevant catchments attributes, it would have been suitable for the purpose of the study.

We agree and we will give a better reason behind the selection of the study region.

- Line 116: “followed by the calculation of the indices”. Which indices?

SPI, SSMI,SSI and drought duration indices. We will add.

- Line 146: “three hourly temporal”, please clarify

The MSWEP precipitation data has a resolution of 0.1 degrees grid with 3 hour intervals. Though for the analysis we used the daily precipitation data.

- Line 169: what “HAD” is?

This should have been Horn of Africa (HOA). We will change.

- Line 209: it is questionable the fact you selected 70th percentile because otherwise you have a too small number of events. The selection of the percentile deals with the severity of the drought events. Probably another message would have come from the analysis

We agree to rephrase.

- Line 215-218: it is not clear to me the explanation provided when P/Q is close to zero. This should happen when the duration of the meteorological drought is significantly smaller than the one for the hydrological drought, and this would happen when you have a small number of meteorological drought events or short meteorological drought events. Also I believe authors should better explain why they decided to use P/Q and P/SM ratio as indicators of drought propagation, what does this ratio means and why they did not use more convention drought propagation index (line 233 it is said that you did not consider the lag because in some previous study it is said that the largest correlation occurs at lag 0)

We agree to give a more detailed description and explanation in the manuscript detailing the reasons for the use of the duration ratios and their representation.

- Line 242-243: reformulate

  We agree to rephrase.

- In my opinion figure 3 panel (b) and do not properly give the same message, much better in the case of streamflow (figure7).

We do not understand what the reviewer meant.

- Line 319: in my opinion it is improper to discuss about “precipitation variability” as you only considered the mean annual precipitation for each catchment and this is not a variability index for precipitation (also on line 425)

We use mean annual precipitation as a descriptor of the precipitation climate as has been used in previous studies (Barker et. al., 2016) and to refer to the spatial differences in precipitation. We agree to avoid using this term.

- Paragraph 5.1 and 5.2 present different obvious sentences, not really of interest for the research (lines 451-456, …)

We agree to remove the paragraph

REFERENCES

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G., Brocca, L., Pappenberger, F., Huffman, G. J., and Wood, E. F.: Global-scale evaluation of 23 precipitation datasets using gaugeobservations and hydrological modeling, Global hydrology/Instruments and observation techniques, https://doi.org/10.5194/hess-2017-508, 2017.

Cattani, E., Ferguglia, O., Merino, A., and Levizzani, V.: Precipitation Products' Inter–Comparison over East and Southern Africa 1983–2017, Remote Sens., 13, 4419, https://doi.org/10.3390/rs13214419, 2021.

Fessehaye, M., Franke, J., and Brönnimann, S.: Evaluation of satellite-based (CHIRPS and GPM) and reanalysis (ERA5-Land) precipitation estimates over Eritrea, Meteorol. Z., 31, 401–413, https://doi.org/10.1127/metz/2022/1111, 2022.

Gleixner, S., Demissie, T., and Diro, G. T.: Did ERA5 Improve Temperature and Precipitation Reanalysis over East Africa?, Atmosphere, 11, 996, https://doi.org/10.3390/atmos11090996, 2020.

Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Barnard, C., Cloke, H., and Pappenberger, F.: GloFAS-ERA5 operational global river discharge reanalysis 1979 present, GloFAS-ERA5 Oper. Glob. River Disch. Reanalysis 1979- Present, 1–23, https://doi.org/10.5194/essd-2019-232, 2020.

Hirpa, F. A., Salamon, P., Beck, H. E., Lorini, V., Alfieri, L., Zsoter, E., and Dadson, S. J.: Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data, J. Hydrol., 566, 595–606, https://doi.org/10.1016/j.jhydrol.2018.09.052, 2018.

Musie, M., Sen, S., and Srivastava, P.: Comparison and evaluation of gridded precipitation datasets for streamflow simulation in data scarce watersheds of Ethiopia, J. Hydrol., 579, 124168, https://doi.org/10.1016/j.jhydrol.2019.124168, 2019.

Ryu, D. and Famiglietti, J. S.: Characterization of footprint-scale surface soil moisture variability using Gaussian and beta distribution functions during the Southern Great Plains 1997 (SGP97) hydrology experiment, Water Resour. Res., 41, https://doi.org/10.1029/2004WR003835, 2005.

Stagge, J. H., Kohn, I., Tallaksen, L. M., and Stahl, K.: Modeling drought impact occurrence based on meteorological drought indices in Europe, J. Hydrol., 530, 37–50, https://doi.org/10.1016/j.jhydrol.2015.09.039, 2015.

Vicente-Serrano, S. M., López-Moreno, J. I., Beguería, S., Lorenzo-Lacruz, J., Azorin-Molina, C., and Morán-Tejeda, E.: Accurate Computation of a Streamflow Drought Index, J. Hydrol. Eng., 17, 318–332, https://doi.org/10.1061/(ASCE)HE.1943-5584.0000433, 2012.