Review: "Future changes in the mean and variability of extreme rainfall indices over the Guinea Coast and role of the Atlantic equatorial mode"

Authors: Koffi Worou, Thierry Fichefet, and Hugues Goosse

This study utilizes CMIP6 future climate projections to understand how Guinean Coast extreme rainfall events are likely to change in the future under the SSP5-8.5 scenario. More specifically, this study focuses on the relationship between the Atlantic Equatorial Mode (AEM) and extreme Guinean Coast rainfall. Results indicate that the extreme rainfall responses to the AEM appears to be captured by the ensemble mean of 24 different CMIP6 models for the present day, though there are biases especially in terms of magnitudes of the response. In the future, they find that the decreased variability of the AEM leads to a reduced magnitude of the rainfall extreme responses over the Guinean Coast. Thus, while there is an overall projected increase in the variability of most extreme rainfall indices, the contribution via the AEM to this variability weakens in the future warmer climate.

Overall, the paper is well written and will be a welcome addition to the scientific community. Before acceptance, I have a few issues detailed below that the authors need to think about and address.

Major Comments:

- Missing from this manuscript is analysis and/or explicit reference to past studies that have evaluated the ability of the CMIP6 models to realistically represent the AEM in terms of the SSTAs. I think the authors may have done this analysis in their prior work (Worou et al. 2022), but the outcome is not explicitly mentioned here in this manuscript. This dawned on me by the time I got to lines 359 – 360 as I started to ponder whether the poor skill scores were just an artifact that the coupled GCMs cannot replicate this mode of variability very well, so of course the relationship between the AEM and the extreme rainfall would not be well captured.
- The authors select CHIRPS daily rainfall data to evaluate and validate the present day CMIP6 results against. That being said there is no evaluation regarding the uncertainty in the observations that you are trying to match up the CMIP6 models against, so it is unclear what biases CHIRPS has in respects to extreme rainfall. I would suggest drawing in additional daily rainfall datasets for this comparison, maybe ARC2 and/or TAMSAT. Both of these datasets extend back to 1983 which would give the authors close to similar temporal coverage to CHIRPS. Doing so would allow the authors to comment on the extent to which the differences between the observations and CMIP6 models is due to the model compared to the uncertainty in the observations. Similarly, you could evaluate multiple SST datasets, though this is less of an issue compared to the rainfall field in my opinion.
- Around line 159: I didn't pick up on this subtlety at first, but realized later on in the manuscript that the authors are evaluating the SSTs for the AEM at monthly timescales, yet they are evaluating rainfall presumably at daily timescales for extreme rainfall events. I guess this may be appropriate, though I wonder if it would be more appropriate to evaluate SSTs at sub-monthly timescales instead (weekly?), as presumably there are sub-monthly variations in equatorial Atlantic SSTs that could be important. Anyhow, to avoid this confusion I suggest the authors explicitly motivate in the text why they choose to evaluate SSTs at the monthly scale to help clarify this decision for the reader.
- I'm unclear why the authors chose to detrend the SST data. I would think they would want to keep it in raw form because they calculate their extreme indices for each year and then average the yearly totals up over the 20 years evaluated over for each time slice if I understand correctly. This decision needs to be explained better in the manuscript. Also – with all of this data manipulation (detrending, normalizing, standardizing), it would be beneficial if the authors would compare these

among the models and observations, so the reader can understand what is being removed/changed, and how alike/dislike are the various models exactly are.

- I know there is a lot of information to convey in a confined article, but I really think it would be useful here in this study to not only analyze results from the ensemble means, but also evaluate individual models to identify which produce more realistic distributions/frequency/intensities of extreme rainfall events. This information could be used to eliminate inclusion of specific models that are judged to perform "poorly" for a given index (or overall), and thus could potentially increase the accuracy/realism of your ensemble mean by eliminating them from consideration before formulating your ensemble mean.

  An example of what I mean involves section 3.2 (Fig. 1 & Table 2). You could evaluate the individual models here and report on whether most models are outliers, or whether most models are close to the ensemble mean with a few outliers on each side. Knowing this could really strengthen the results and the reader's confidence in the ensemble mean. While I understand it is unfeasible to show Fig. 1 for all 24 models, but maybe you could consider expanding/reimagining Table 3 to include info for all of the models as well as the ensemble mean and organize the individual models from those that perform the best to the worst for your selected indices.

  Likewise, you could do the same for Tables 4 and 5, expanding them to include individual models ranked from best to worst to comment on their relationship to the ensemble mean results already shown.

Minor Comments:

- Line 145: I presume you are using daily SST and rainfall data, correct? Or is it 3-hourly? Suggest you explicitly mention this in the text here for better clarity.
- Lines 165 – 168: Would be helpful to include a figure to orient the reader here, showing the countries and a box of the analysis region you defined in the text.
- Table 1: Include information on the spatial resolution for each model evaluated so that information is conveyed to the reader.
- Line 168: I am confused why you are selecting 2.8° resolution for use here. Maybe it is due to the coarsest resolution of the GCMs? Suggest motivating this choice better here at first mention (adding the spatial resolutions for each model to Table 1 would help here too). Actually – I see you have done this later on line 175. I'd suggest moving it up to this line to avoid confusion.
- Section 2.3.1: Motivate why you chose 1° resolution here. I suspect the CMIP6 models are coarser, so how exactly did you interpolate to a higher resolution?
- Line 175-176: Do you mean "….averaged over the JAS season **for each year**"? It would also be informative if the authors could calculate and report on the standard deviation over each time slice interval you evaluate over to see how its variability is changing.
- Line 188 – 190: "performs better" in terms of what exactly? Can you clarify what you mean here better? Also – do you still intend to evaluate individual models to identify which produce more realistic distributions/frequency/intensities of extreme rainfall events? This would be important to include I would think. Furthermore, this information could be used to eliminate inclusion of specific models that perform "poorly" for a given index, and thus could potentially increase the accuracy/realism of your ensemble mean by eliminating them from consideration.
- Figure 2 – again how does this change when a different target other than CHIRPS is used? If it is the same, you could just comment on it in the text without adding additional figures. If it is

different, it may be useful to include/expand a figure showing the changes if a different target is used.

- Line 460: Suggest expanding this to include discussion in terms of the uncertainty that exists in the observations by using more than 1 rainfall dataset to evaluate the CMIP6 models against (see prior comment earlier).
- Lines 464 – 465: Comment on how the ensemble mean relates to the spread of the individual members that are used to determine the ensemble mean. Are most models close to the ensemble mean, or are most outliers and they average out to the mean?
- Lines 483 – 485: So how well can we expect the CMIP6 coupled models to replicate the AEM, and what are the implications for this on your findings here?
- On all spatial map figures in the manuscript it would be helpful if the country outlines were included in each panel.