Responses to reviewer 2

We would like to thank reviewer 2 for his constructive comments and remarks that help to improve our current study. The questions of the reviewer are recalled in black, and our responses are written in purple.

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.
  Thank you for this excellent comment. Current GCMs can replicate the SST variability associated with the AEM over the tropical Atlantic Ocean, but they struggle to reproduce the spatial distribution of the observed rainfall response over the tropical Atlantic and the Guinea Coast. This has been highlighted in our prior work (Worou et al. 2022), and we will recall this point in the revised manuscript.
- 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.
  Thank you for the suggested rainfall data. We will add a total of five additional observed rainfall data in the revision of this work. These data are: ARCv2, PERSIANN, REGEN, TAMSAT and GPCC_FDD_v2022. As an example, Fig. R2.1 shows that the model multi-median performs well in the simulation of some characteristics of the rainfall over the Guinea Coast. However, it overestimates the frequency of wet days (FRQW) and the maximum consecutive wet days (CWD). In our revised manuscript, we will assess the performance of the models against the six different observed rainfall datasets. When needed, the multimodel median values across the different observations and the different models will be indicated (for the mean biases map for instance). On the other hand, we won't include additional observed

SST data as we consider that this is not a critical point for out analyses.

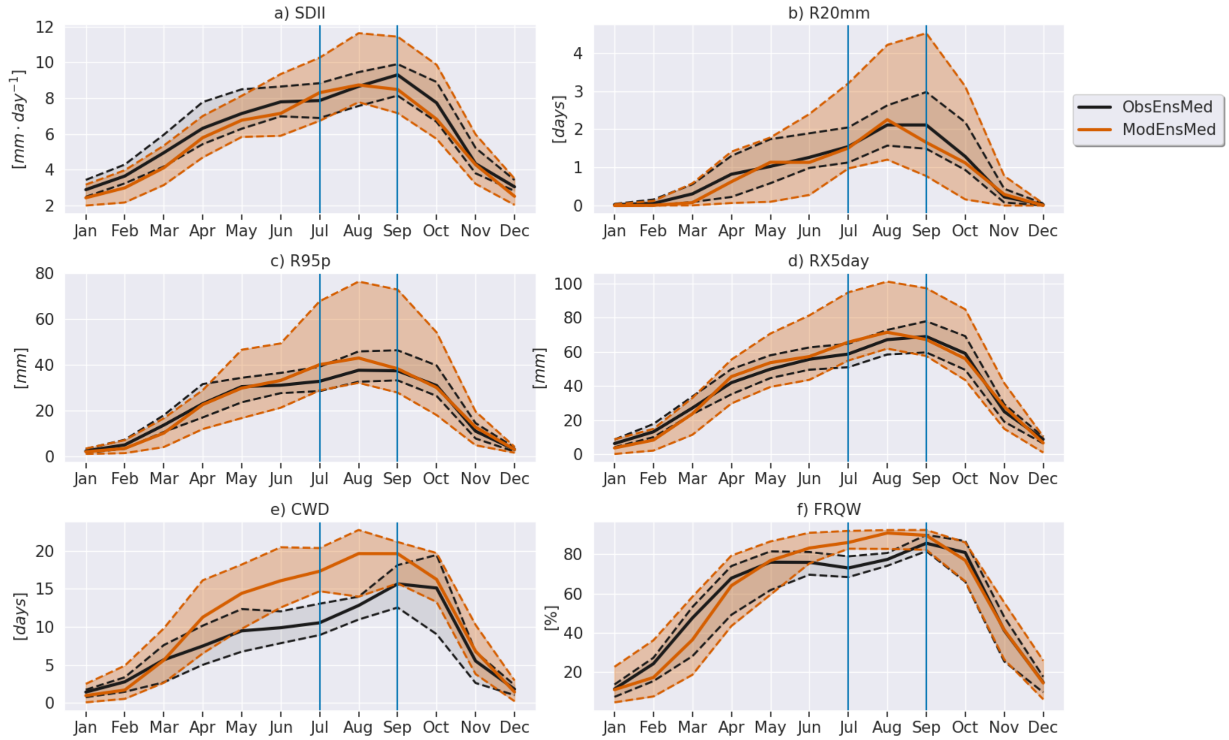

Figure R2. 1 Annual cycle of the extreme rainfall indices over the Guinea Coast (1995 – 2014): median (black and brown curves), 10^{th} - 90^{th} percentile range (gray and brown shades) are indicated in black for 6 observations and in brown for 24 CMIP6 GCMs. Vertical lines indicate the season considered in our study.

- 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.
  To clarify the point, on the one hand, we are computing monthly extreme rainfall indices from daily rainfall values. Then, we consider an average over the boreal summer season, from July to September. This means that, at the end, we have seasonal values for the rainfall indices for each year of the considered period. On the other hand, we are interested in the variability of the Atlantic equatorial mode (also called Atlantic Niño), which peaks from the end of spring to the boreal summer and occurs mainly on the interannual timescale. Our study does not address the variability on weekly timescales, though they are also important for sub-seasonal to seasonal coupled atmosphere-ocean processes such as the Madden-Julian Oscillation, which have also a large impact on the West African hydroclimate. We will clarify our choice in the revised manuscript accordingly.
- 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.

There are two main points in our study. First, we look at the mean changes in the extreme variables as well as the changes in their variability. The extreme rainfall indices are kept in their raw seasonal values for each year. Second, we address changes in the variability of the extreme indices around their mean values. We remove any linear drift in the 20-year period timeseries, from each monthly series before averaging over the July-September season. Our aim is not to study the trend in the data, but rather the interannual variability in the extreme indices that is partly explained by the Atlantic equatorial mode. Similarly, the Atlantic equatorial mode positive or negative phases are defined by SST anomalies exceeding at least one standard deviation around the SST mean values in the eastern equatorial Atlantic. Any linear trend (due to global warming for example) needs to be removed to focus on the AEM phenomenon during the 20-year period, since climate change could have an impact on the AEM variability. We propose to include in the online supplementary material, for the present-day period, for each extreme rainfall index, for the AEM SST index, for each model, and for each observation, a figure illustrating the JAS average of the raw indices, as well as the trend of these indices. This will help the reader to have an idea of the behavior of the trend in each index. We will add this clarification in the revised manuscript.

- 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.

Thank you for this interesting suggestion. First, we will replace Tables 3, 4 and 5 by a box-whisker plot to highlight the variability among the models in their representation of the different extreme rainfall characteristics. When needed, we will use the multi-model ensemble median instead of the multi-model ensemble mean, which is less impacted by outliers.

We will also show some diagrams of the models' performance relative to the 6 observations: for a given metric, the different models will be on the horizontal axis and the different extreme indices will be on the vertical axis. This was termed the "portrait diagram" in Faye et al. (2021). This will help the reader to see, for each model, the skill in representing the different variables.

Furthermore, a model which poorly represents a particular index does not necessarily struggles in representing another index. This choice of models' stratification in different categories could be easily done for one variable but it will make difficult the presentation of the models' performances for all the variables. We will therefore do this stratification only when studying the teleconnection patterns to the Guinea Coast. In that case, we will simply split the models into two groups: the first group will have a good sign of the extreme rainfall response to the Atlantic equatorial mode phases, and the second group will have a sign opposite to the observations.

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.
  We are using monthly SST data and daily rainfall data. We will mention explicitly the temporal resolution of the datasets as suggested. Thank you.
- 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.
  We will add countries on a map and a box showing the different regions mentioned.
- Table 1: Include information on the spatial resolution for each model evaluated so that information is conveyed to the reader.
  The spatial resolution of each model will be added in Table 1 of the revised manuscript.
- 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.
  The 2.8° resolution corresponds to the coarser resolution of the 24 models. We will move the information to the right place, as suggested. Thank you.
- 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?
  There are models with a low resolution that only have a few grid points within the ATL3 region (20°W-10°E,3°S-3°N). If the choice was made on each model grid, this would imply to use different regions for different models. We therefore use a bilinear interpolation method with the climate data operator routine (CDO) to interpolate all the sea surface temperature datasets to a common grid of 1° of resolution. We will add this information to the revised manuscript. A similar procedure has been applied in Worou et al., (2022) and Kucharski et al., (2017) for instance.
- 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.
  Thank you for this remark. In the revised manuscript, we will improve the description of the method. In the submitted version, we computed the rainfall indices over each season, for each year. In the revised manuscript, we will compute the rainfall indices for each month (to show the annual cycle). For uniformity, we will keep the monthly values of the extreme indices, and when we need a seasonal index, we will average over the months in the season (e.g. July-August-September) **for each year**. This will also be the case for seasonal SST and rainfall anomalies. Table 2 will also be modified accordingly by reducing the number of variables and by replacing the JAS season by monthly values.
  Moreover, as suggested, we will add a box-whisker plot of the standard deviation of the different seasonal extreme rainfall indices, probably in the supplementary material. We show below the distribution of the changes in the standard deviation (Fig. R2.2):

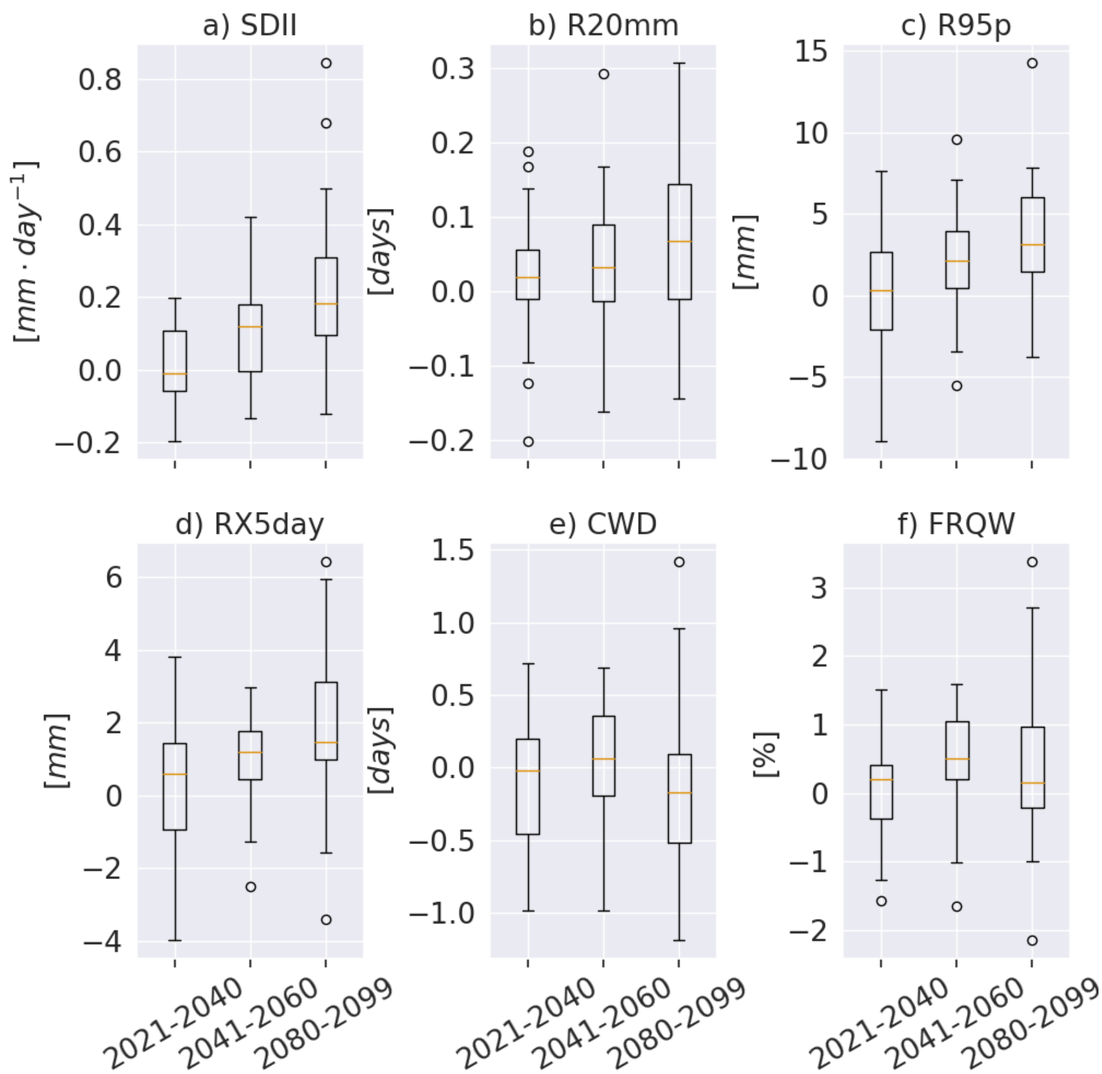

Figure R2. 2 Near-term (2021-2040), mid-term (2041-2060) and long-term (2080-2099) mean changes (relative to 1995-2014) in the standard deviation of the timeseries of JAS extreme rainfall indices averaged over the Guinea Coast. Each box-whisker plot represents the changes from 24 different GCMs, in the SSP5-8.5 scenario relative to the historical simulations.

- 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.

  By "performs better", we mean that the multimodel ensemble mean values are closer the observations than each individual model. This can be seen in Fig. R2.1, where we show the multimodel ensemble median instead of the multimodel mean. We will modify the text accordingly. We will also add portrait diagrams showing the skills of each model, compared to the different observations.

  Furthermore, we have decided to include a stratification of the models into "poor" and "good" categories in the analysis of the AEM impact of the extreme indices over the Guinea Coast, to have a more specific analysis.

  Finally, we will consider the multimodel ensemble median when needed, instead of the multimodel ensemble mean, as it is more resistant to outliers.

- 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.
  Figure 2 of the submitted manuscript will be modified. For each model, the biases will be computed relative to each of the six observed datasets. Then, the new figure that will resume the models biases relative to the different observations will be the multi-model ensemble median along all the different models' biases computed. The portrait diagram that we will provide will show the biases for each model, averaged over the Guinea Coast, which will give another complement information at an individual model level.
- 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).
  We will update in the revised manuscript the discussion of uncertainties related to the observations.
- 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?
  We will provide in the supplement material six different Taylor diagrams relative to the six different indices, to illustrate the spread of the models' spatial distributions compared to the multimodel ensemble median. This information will be taken into account in the revised manuscript.

- 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?
  The CMIP6 models can simulate reasonably well the SST pattern associated with the AEM over the tropical Atlantic. However, they show some difficulties in simulating the rainfall responses over the tropical Atlantic and the Guinea Coast. To get more insight into the models' performance, we will focus on two groups of models as discussed above, i.e. models with a correct response in the rainfall and models with a wrong response. We think that this is the most adequate way to discuss the models' ability to simulate the AEM-related patterns in the present framework. This will be considered in the revised article.
- On all spatial map figures in the manuscript it would be helpful if the country outlines were included in each panel.
  Thank you for this comment. We will add country contours in the different maps.

References

Worou, K., Goosse, H., Fichefet, T., and Kucharski, F.: Weakened impact of the Atlantic Niño on the future equatorial Atlantic and Guinea Coast rainfall, Earth Syst. Dynam., 13, 231–249, https://doi.org/10.5194/esd-13-231-2022, 2022.

Kucharski, F. and Joshi, M.K. (2017), Influence of tropical South Atlantic sea-surface temperatures on the Indian summer monsoon in CMIP5 models. Q.J.R. Meteorol. Soc, 143: 1351-1363. https://doi.org/10.1002/qj.3009

Faye, A., Akinsanola, A.A. Evaluation of extreme precipitation indices over West Africa in CMIP6 models. Clim Dyn 58, 925–939 (2022). https://doi.org/10.1007/s00382-021-05942-2