

Review of “Physical Climate Drivers of East Africa’s March-April-May (MAM) seasonal rainfall Identified through Machine Learning Analysis”

This paper explores the use of machine learning techniques to improve the forecasting of the East Africa Long Rains, an incredibly important season for the region, yet has long suffered from low forecast skill. This study presents an important advance, as it shows two methods that lead to an increase in forecast skill over the region. This would have a significant impact on local livelihoods and economies, and thus represents an important advance. However, in the current form the manuscript is confusing and doesn't present enough evidence to be scientifically robust. With a number of revisions, the manuscript would represent an important advance in the field. I thus recommend major revisions.

My overall comment is that there are far too many regions selected, with insufficient evidence to show how these regions physically drive MAM rainfall. There is no evidence of analysis of independence between the different regions. There is no clear link for each individual region and the physical drivers that link that specific region to MAM rainfall – this is OK if the paper is just showing that ML with a range of regions can improve forecasts, but the paper title talks about “physical climate drivers” and I am not sufficiently convinced that all of these regions are shown to be “physical climate drivers”.

The methods are not fully described, and this limits the interpretation of the results in places.

The method and data used for Maximum Covariance Analysis (section 3.2) is not described, thus limiting the interpretation of these results. This part on changes in the Walker Circulation is very interesting and a very important part of the paper, but without a clear explanation of how these figures are created it is hard to understand the interpretation. I also don't think it shows the contribution of each individual region, which is important if these regions are to be described as “physical climate drivers”.

The set of years for hindcasts/ reforecasts etc is not given. This is an important detail and should be explained. Are the results robust to different training and validation periods? This is an important consideration, given the recent trends in the long rains, and is not presented.

More detail needs to be given for datasets (e.g. period, resolution). More than one rainfall dataset should be considered. Is GPCC shown to be a good dataset over the region? More justification for the choice of this dataset should be included, and the results should definitely be tested using at least one other rainfall dataset.

The method for defining the regions in Figure 1 is only defined in the caption of Figure 1, not in the methods. It is later noted that much of SSRZ does not have rainfall during

MAM – if this is the case then why is it included in the region. The regions should be designed such that they have rainfall during the season of interest. It is also not clear why this appears so early in the paper – if this is given to locate Eastern Africa, then more context (e.g. location within Africa) should be given.

It is said that correlation analysis is used to identify the regions, yet there seem to be some regions selected with almost no correlation shown in the region, and no reasoning (e.g. comparison with the literature) is given for why these extra regions are included. As an example, ASTI is shown as an SST region in Figure 3, but this is a land region, and ATL3 has almost no correlation in either December or January. The Azores basin shows very low SLP correlation in both months. There needs to be better description of the selection of the regions. Some gradients and dipoles are calculated but seems to be little justification given for why these gradients were selected. This should also be done using the background literature – there is limited mention or reference to the Western V for example. Literature should be used to support the choice of regions.

The EOF analysis should be shown – is this done using seasonal means? How much of the variance is explained? This is important to include, as all the correlations are done using this analysis. It also says in line 104 that “basin-mean correlation coefficients with EA MAM rainfall PC1 are given in Table 1”, but these do not seem to be included. These should be shown.

There seems to be errors and inconsistencies. For example, Figure 3 shows SST correlations, but shows correlations over land regions – there shouldn't be SST correlations over land regions. There are also some odd straight lines in the correlation maps – are these physical? It is mentioned that changes between Dec and Jan are considered, but then not clear how these are included.

The discussion of the Beeswarm plot seems to be incorrect – it states that “higher values of most predictors are associated with enhanced rainfall” and “the exception is southern Pacific 2-m temperature, which shows a negative contribution to EQZ rainfall”. Yet figure 6a shows high feature values for SPT2 linked with a positive SHAP value, indicating a positive contribution to MAM rainfall? This either needs to be corrected, or more clearly explained.

The figures are difficult to read – a lot of the text is very small and in bold font, which makes it hard to read. It is important to be able to see which region is which, but this is hard to do with the current font. Why are there white patches in Figure 11?

Linked to the large number of regions, there are a lot of abbreviations included in the text – this makes it hard to read and follow. It maybe good to include a full name (e.g. West of Australia) at first use.

OND_ANOM_PY seems to be first mentioned in line 450 – it is not clear how this is defined or included. This should be made clearer.

Minor comments:

- Line 44: Could cite Hiron and Turner, 2018.
- Line 50: Could cite Kisenber et al., 2026 <https://doi.org/10.1175/JHM-D-25-0080.1>
- Line 160: “random search and cross-validation for each country” – do you mean region rather than country?
- Line 176: How was the masking done? This has not been described.
- Figure 4 caption (and elsewhere) - I am very unsure about the use of the phrase “thermal regions”. I think this should be reconsidered throughout.
- Line 206: “nearly the same board regions” – this comes back to my question of independence – if we are looking at the same regions then should some of the regions be combined?
- Skill metrics are not included in the methods – these need to be described. As previously mentioned, it should be clear which years are included in this.
- Figure 12: what is OV?