The author present a nice study, evaluating existing metrics and a new metric for the onset and end of the rainy season with respect to vegetation data. One major result is that threshold based metrics needed to be calibrated region-specifically as they have done with the NDVI. Furthermore, they test sensitivities of the metrics, and they apply them to climate projections, showing that despite projected increase in precipitation, the rainy season onset and end do not show significant trends.

The study is interesting and relevant, however, the study could profit from an improved structure and some clarifications in the section Material & Methods as stated below in the specific comments..

Line 30: "hereafter named metrics" -> this seems a bit unnecessary. Maybe, it's fine if you just start the next sentence with: Broadly, these rainy season metrics…

Line 37: It's called the Standardized Precipitation Index (SPI), maybe denote it like this. Also, it is a bit a confusing citation. You are talking about RSO and RSE metrics, but the SPI is used to separate the rain season into wet and dry periods? Consider removing this citation.

Line 46: it would be nice if you mention some examples of uncertainties here.

Line 61: How do you deal with irrigation when using a vegetation metric for independent validation?

Line 65: I think the sentence is nicer, if you remove "regarding such independent data"

Fig. 1: The topography colorscale here is not very intuitive. It could be useful to reduce the colorscale of topography to something simple, i.e. green to brown in order to give an intuitive idea of the topography

Line 145-162: I would suggest to first mention the three observational datasets (WRF, CHIRPS, AWS) and describe them shortly, and only then the scenarios. I.e. Move everything after "In addition, Potter et al. (2023)" to a later position in the paragraph.
It could be nice to mention the abbreviation that you use in the figures already here (WRF, CHIRPS, AWS). For example, its' quite obvious that AWS means automatic weather stations, but it's never mentioned, I think. For reproducibility, could you state which three stations you used?

With respect to CMIP 5 downscaling, maybe it's good to mention briefly, what type of emission scenarios RCP4.5 and RCP 8,5 are.

Line 196: and the our -> "the" not needed?

Line 196-205: this small section is quite relevant. Maybe you could consider making an own small chapter out of it in a new chapter, for example: 2.2 Rainy season metrics, 2.3 The new "bucket" metric, 2.4. Calibration of threshold-based metrics, 2.5 Sensitivity analysis 2.6 Future projections. Also, in the Figures, you compare the thresholds provided by the authors to your calibrated

thresholds. Maybe, you can mention that you will be comparing them as well in the section on the calibration?

Are all the necessary information given, on how you apply the Differential Evolution optimization for reproducibility of your study?

Line 232: what is an almost large number of rainy seasons? Maybe consider deleting the "almost"?

Line 247: individual models were excluded individually -> maybe one individual is enough?

Sect. 2.4: It could also be helpful to separate the sensitivity analyses and projections in to separate chapter, because you will be doing two separate things with the climate projections?
Do you state anywhere how you calculated the trends for the past and the future (maybe I just didn't see it)? That could go into such a short section?

Line 266: always add the unit after the RMSE, i.e. 8.8 and 14.4 days.

Fig 2 and 3: I find the figures nice, but very full. Could you consider either moving the calibration data or the RMSE (or both?) out of the Figure into a separate table?
Please mention more clearly that INIT WRF are the thresholds provided by the authors, maybe even mention this in the section very you describe the calibrations instead of just mentioning it in the caption?

Line 278ff: Here, mention that you talk about the INIT WRF shown in the figure, otherwiseit is not so clear. Also, specifically state the very high RMSE, when the parameters of the rainy season metrics are not tuned. I think this is a very relevant result of your study. For the agricultural perspective, a RMSE of 34 days (e..g Gurgiser) is very different from an RMSE of 12 days.
Also, could you state why you looked at the original threshold only using WRF, and not for example the AWS?

Fig 4: Please add inside the Figure a legend with the colors of the precipitation data sets. This can increase the readability a lot.

Lines 308 ff: You could also call Sect. 32. Sensitivity analysis of rain season metrics?

Fig. 5 This figure is quite small. Maybe you can enlarge the figures a bit by adding the rainy season metric title at the top of the figure (i.e horizontally). Like this you have more space in the horizontal dimension.

Conclusion: Could you add a sentence summarizing the results of the sensitivity tests?

Line 456: I don’t think you can say it is an “unprecedented” number of future projections since these have been published by Potter already?