Earth System
Dynamics
Discussions

Interactive comment on "A bias-corrected CMIP5 dataset for Africa using CDF-t method. A contribution to agricultural impact studies" *by* Adjoua Moise Famien et al.

Anonymous Referee #2

General comments

The authors present a dataset of CMIP5 GCM output of daily accumulated precipitation, daily mean, minimum and maximum near-surface air temperature, daily mean surface downwelling shortwave and daily mean wind speed bias-corrected using the CDF-t method and the global gridded observational dataset WFDEI over Western Africa. They compare the raw and the bias-corrected data to WFDEI and 2 other observational datasets, namely WFD and EWEMBI. Unsurprisingly, the bias-corrected data feature much smaller biases relative to WFDEI than the raw data. Specifically, the authors look into biases of seasonal mean values and mean annual cycles of temper-

ature, precipitation and radiation, the 95th percentile of temperature and precipitation, and the number of days on which temperature or precipitation are above or below certain threshold values. The analysis is confined to the five ISIMIP GCMs as this enables a comparison to the same CMIP5 GCM output bias-corrected using the ISIMIP method and the dataset WFD as it was done in ISIMIP. Lastly, a crop model is driven by the different climate input datasets and maize yields are evaluated in terms of both how well they reproduce historical observations and how they are projected to change under one RCP scenario.

The writing style of the manuscript is sloppy and the English is not good. They manuscript contains many figures, which on average have a rather low information content. The discussion of these figures in the text is mostly not very insightful. Comparisons of WFD, WFDEI and EWEMBI data are in many cases unnecessary since it is clear from the definition of these datasets that they are very similar or even identical in many cases. A table that summarizes the differences and commonalities of the datasets for the considered variables should be included in order to make this more transparent to the reader.

The manuscript does not add anything to the existing knowledge on bias correction. New findings are only present in parts of the comparison of the different datasets as well as in the sensitivity analysis of maize yields simulated using these different datasets as input. However, since the latter analysis is done for one crop model in combination with one GCM only, general conclusions cannot be drawn from it. A comparison of its results to corresponding ISIMIP or AgMIP studies might help put the results of this study into context.

In addition to that, the manuscript suffers from a few methodological flaws, see my specific comments below. Overall, I think that a major revision of the manuscript is needed before it can be published in ESD.

Specific comments

P2 L6: What are "robust biases"? And what is a "bias" in the context of this paper? I think you should write one or two sentences about that.

P2 L8: "statistical bias-corrections are necessary [...]" – Why? Please explain why you cannot just use the original GCM output as input to your impact models/what would happen if you did so.

P2 L17: "on 4 out of the 5 same CMIP5 GCMs" is not quite correct. Only 3 out of the 4 GCMs chosen in ISIMIP2b were also used in the first phase of ISIMIP.

P2 L22: Which kind of biases have been identified over Central, East and South Africa? Climate models have always been and will always be biased, so merely saying that a model is biased is an empty statement.

P3 L24: Why do you not present results for northern winter and autumn?

P5 L1: How did you interpolate the other variables?

P5 L13: Which WFDEI version do you use? The one with precipitation corrected using CRU (WFDEI-CRU) or GPCC (WFDEI-GPCC) estimates? Please specify.

P5 L22: You should inform the reader here that over land, EWEMBI is identical to WFDEI-GPCC for precipitation, daily mean, minimum and maximum near-surface air temperature and 10 m wind speed. Only for surface downwelling shortwave radiation there is a difference between EWEMBI and WFDEI-GPCC data over land.

P7 L4: Which SRB data exactly do you use for this comparison. Please describe that here or in Sect. 2.2.

Sections 3.1 and 3.2: This is an insufficient description of the CDF-t method. Merely referring the reader to Michelangeli et al. (2009), Déqué (2007) and Vrac et al., (2012, 2016) for all the details is not enough. Please be more specific about how you use $F_{obs,fut}$ to do the quantile mapping in the future period (I assume you map x to $F_{obs,fut}^{-1}(F_{mod,fut}(x))$) and please describe how you estimate $F_{obs,cal}$, $F_{mod,cal}$ and

$F_{mod,fut}$. Are these CDFs estimated parametrically or non-parametrically/empirically and how exactly do you do the estimation? Also I do not understand how you account for seasonality: Do you apply CDF-t month by month or using moving windows? Moreover, you state that "CDF-t preserves any long-term trend in the GCMs data" but do not give any reference that would corroborate that statement. Thinking about Eq. (1), I came to the conclusion that CDF-t does neither preserve trends in moments nor in quantiles. Please discuss this aspect in more detail since the users of your data product will want to know if and how you have modified the trends present in the original CMIP5 GCM data. Lastly, you state that "GCMs data have been interpolated to WFDEI grid before being bias-corrected." Which method do you use for that interpolation?

P8 L3: Not "every GCM has to be calibrated" but the bias-correction method has to be calibrated individually for every GCM.

Sect. 3.3 and Figure S2: I do not understand why you did what you did here. Let's take calibration period 1997–2013 as an example. Did you use 1997–2013 GCM and observational data to calibrate the CDF-t method and then apply the thus calibrated method to the same 1997–2013 GCM data? If that is what you did then I do not understand the purpose of these tests since in that case of course the remaining biases will be small and you cannot draw any conclusion in terms of overfitting. Therefore, also the concluding statement of the section would be nonsense. What you need to do to test for overfitting is a cross-validation.

P9 L2f: What you describe here is not what is shown in Fig. 2.

Figure 3: "Taylor diagrams relative to the mean of surface temperature" does not make sense. Please rephrase. I assume that all panels in the upper row refer to the Sahel box and all panels in the lower row to the Guinean box, correct? Please specify this in the caption. Also, please add a (separate) figure with a map showing the definition of the SAHEL, GUICOAST and any other region used in your study. Lastly, your Taylor diagrams suggest that 1979–2001 JAS mean EWEMBI and WFDEI tas have different

spatial standard deviations despite EWEMBI and WFDEI tas being identical by definition. Can you explain this? Have you maybe used a WFDEI data product version that is different from the one used for the production of EWEMBI? The same question applies to Figures 5, 8, 9, 12, 14, 16.

P18 L1: "daily values" of which variable?

P18 L9f: "CDF-t method is also a bit better than ISIMIP one when one refers to EWEMBI reference data" – of course, because the ISIMIP data were bias-corrected using WFD, your data were bias-corrected using WFDEI data, and EWEMBI is mostly identical to WFDEI (see my other comment above). There is no real point you are making here or elsewhere, where you have made the same statement. Basically, you could leave out the right panels in all Taylor diagram figures related to temperature and precipitation since there is no qualitative difference to the respective middle panel.

P20 L6f: Strange sentence. Please rephrase.

Figure 18: Maps of difference to WFDEI would be good here.

Figure 19: Maps of difference between, say 1970–2000 and 2070–2100 mean values would be good here. For which RCP is this? Do you get qualitatively similar results for the other four ISIMIP GCMs? I mean, why do you do all this work and then just show results for one GCM...?

Table 2 and its description in text: Why this selection of sensitivity experiments? Why, for instance, do you do the WFDEIWFDrsds experiment but no WFDEIWFDpr experiment even though the WFDEIpr and WFDEIrsds experiments suggest a higher sensitity of yields to pr than to rsds biases?

P25 L2: "It indicates also that WFD data and related bias-corrected simulations should not be used anymore" – I think your analysis as it currently stands is not sufficient to come to this conclusion because (i) you are comparing potential simulated yields to actual observed yields and (ii) you did not say anything about the quality of the

GDHY data. Nevertheless I think that you are right in terms of rsds as Weedon et al. (2014, doi:10.1002/2014WR015638) report substantial improvements of rsds in WFDEI relative to WFD.

P25 L11ff: "Interannual variability of simulated yields is proportional to the mean with a very weak variability for ISIMIP yield and higher variability for CDF-t and raw simulations. All projections show a clear decrease of maize yields by a factor of 2 over all of West Africa along the 21 st century." – I think you should also show that visually by plotting yields relative to mean 1950–1980 levels, for example.

Technical corrections

P2 L10: I think it would be better to write "in combination with" in place of "based on".

P27 L7ff: To be more precise, your list of variables should read "daily accumulated precipitation, daily mean, minimum and maximum near-surface air temperature, daily mean surface downwelling shortwave and daily mean wind speed" here and elsewhere in the manuscript.