Response to RC1 on nhess-2022-162

NOTE: Reviewer's comments are in black, our responses to the comments are given in blue below.

The main achievement of the analysis performed by Petrovic et al. consists of providing insights into the utility of improving spatial resolution and customizing the model setup of RCMs aimed at reproducing (and, possibly, projecting) drought characteristics. Through a simple yet straightforward experiment, they give clear answers that can drive further development of RCMs aimed at reproducing and projecting drought characteristics. I only have two main suggestions for the authors:

1) I wonder if SPEI behaviour for other aggregation scales would be the same. Would it be possible to give any information (even as supplementary material) also for SPEI-6 and SPEI-12?

> Yes, we agree with the reviewer that this information might be useful to the readers. Thus, we will add supplemental material about SPEI-6 and SPEI-12 and add some text to the respective sections. It will mainly be about the changes depending on the aggregation scale.

2) In the Introduction, the authors declare the objective of gaining insights into drought development for Germany etc., but I can't find specific indications about that in the paper. A detailed section could help; otherwise, I suggest avoiding emphasizing this objective in the introduction.

> Drought development was meant to refer to the meteorological drought development in Germany (and the near surroundings) in the considered time period 1980 – 2009 based on the E-OBS reference data. These results are mainly included in section *4.5 Drought Characteristics Analysis* and the numbers from the respective tables are given in the abstract and in the conclusions. We agree that another term than *development* would be more accurate, so we have decided to use the term *course*: "2. To gain insights into the meteorological drought course for Germany and the near surroundings between 1980-2009."

L84: ERA-Interim, capital I.

> Correct. We will change it.

L87: "concluded". In the line below: "saw".

> Correct. We will change both.

Table 1: are those listed the only models available? Are there others (e.g. HIRHAM)? Please specify in the text. If some models are neglected, please explain why.

> The models listed depict the complete subset for the two requirements: Containing the necessary variables (precipitation, Tmax and Tmin) and coverage of the time period 1980 – 2009. To make this clearer in the text, we will change it to: "At the time of selection, these were all available model runs that cover the study period 1980 – 2009 and contain the relevant meteorological variables needed for the analysis."

L222: to make the paper more self-consistent, I suggest providing more information about how the SPAEF metric works. On the other hand, less room can be devoted to the Mann-Kendall test, which is older and more widely known.

> We can see the point and will provide more information about the SPAEF metric. We will also work on shortening the section with the Mann-Kendall trend test.

Section 4.1: Taylor diagrams don't provide information about possible bias. I suggest adding this piece of information.

> Indeed, there are no bias information included. Therefore, we will add a table with bias values from the spatially and temporally averaged monthly values of the three important SPEI variables. Text for description will be added in the results section and also in the conclusions.

L296: from 50, I guess.

> Thanks for the hint. It should be "from 12.5 to 5 km". We will change the sentence (also after a comment of reviewer 2) to "From our results, we obtain that, if existent, the benefits of a resolution increase from 12.5 to 5 km are less distinct.".

L314: I suggest removing Table 4 and introducing the mean correlation coefficient as an inset in the respective map (or near the title).

> We agree. Table 4 will be removed and the mean correlation coefficient values will be added underneath the title of the respective models.

Section 4.3: why not use WRF@15 km here, at least in Table 5? It would further highlight the benefits of model settings.

> We agree with the reviewer. We will add the map of the WRF@15 km run to Figure 3 and the relevant scores to Table 5. Figure 3 will then contain both WRF runs next to the E-OBS reference. Additional text will be added in this section and in the conclusions section.

Table 5: are the Mean SPEI values and the other statistics averaged over the German territory or the whole domain?

> They were averaged over the whole domain.

L347: … of this RCM …

> Correct. *Of* will be included: "The only exception is ALADIN with the maximum value of 0.55. REMO holds the lowest SPAEF value (-1.89), which completes the overall bad performance of this RCM in this regard."

L355: in the authors' opinion, why do we see these results with the Tiedtke scheme? However, from Table 2, I observe that also RegCM uses Tiedtke.

> We have further analyzed the relevant time steps (months) of the spatially averaged time series from the three meteorological variables to check if the deviations from the reference are also the highest in this case. It turned out that the RCMs, which were run with the Tiedkte convection scheme, do not show extraordinary bias values here. From this we conclude that the deterioration in SPEI-3 performance for the 2003 event is rather not related to the Tiedtke scheme. The fact that RegCM was also run with Tiedtke corroborates this conclusion. Consequently, we will remove the section in question.

Figs. 4 and 5: I would merge them. In general, I suggest always considering as two different configurations WRF@5km and WRF@15km (see my comment to Section 4.3).

> We agree that merging the two figures would make sense, but see the difficulty in the resulting figure dimensions since this would result in nine panels. As a compromise we would like to keep the Figs. 4 and 5 the way they are and add the WRF@15 km scores in the Tables 7 – 9 in the drought characteristics section. Then the WRF@15 km information will be included in every single section. This implies also some adjustments in the results and conclusions section.

Section 4.5.1: E-OBS drought frequency looks too high for some areas (up to 22/24 times over 30 years). What is the reference period on which the index is calculated? 1980-2009?

> We can see that the drought frequency values appear relatively high for some areas. It must be kept in mind that already events with an SPEI-3 value equal to or below -1 are considered as droughts here. This does not necessarily imply drought events to be severe or extreme. Due to the definition of the SPEI, this can also imply a just drier than normal period (dry anomaly), which is then considered as a drought event. This can also happen in autumn and winter months. For clarification, some description for the proper classification will be added to the text.
> Yes, the reference period for the index calculation is 1980 – 2009.

L414: so, it is mean absolute error (MAE).

> Correct, it is the mean absolute error. This information will be added to the text: "For the drought characteristics we used the mean absolute error (MAE) as a measure for the domain mean bias (third column in Table 6) since values with opposite signs can balance each other out, thus making the information less meaningful."

Sections 4.4 and 4.5: discussion, specifically in terms of comparison to existing literature, is mostly missing.

> We can see the point here. Searching for literature dealing with similar topics (trend detection and drought characteristics) for the same or similar region was not very fruitful. However, some discussion can be added to section 4.4 regarding model capabilities to reproduce observed trends using the Mann-Kendall trend test: "Nasrollahi et al. (2015) applied the Mann-Kendall trend test to the outputs of 41 CMIP5 models to evaluate their ability to replicate observed drought trends on the global scale between 1901 – 2005. They used the SPI-3 as drought index (and SPI-6 in the supporting material). Their results revealed that about 75 % of the models reproduce the global drying trend, but most models fail at reproducing regional wetting and drying trends (at most about 40 % with agreement). In most locations, less than 10 % of the models showed agreement with the observations. Greater agreement was found in higher latitudes. Um et al. (2017) also performed the Mann-Kendall trend test on grid cell based SPEI-12 time series from outputs of four (HadGEM3-RA, MM5, RegCM4 and RSM) RCMs from CORDEX East Asia and of their ensemble mean for the time period 1980 – 2005 over East Asia. They found distinct differences among the single model outputs regarding their capability to reproduce observed drying and wetting trends. While HadGEM3-RA and MM5 generally captured the proper trends, RegCM4 and RSM were only partially successful. This is why the ensemble mean showed relatively poor performance compared to the two former RCMs. These results highlight the spread in the model's capability in reproducing observed trends of wetting and drying, which is found in this study as well.".
>
> Regarding the drought characteristics, matching literature is sparse. Moreover, different methods, time and aggregation scales, definitions, reference data sets etc. make comparisons with existing literature not really meaningful in most cases.

Conclusions: maybe, the main achievements of the paper can be highlighted with bullet points (e.g., increased resolution and setup are not useful for drought characteristics; they are helpful for correlations; they are useful for trends, etc.).

We thank for the suggestion and comprehend that this would make it very clear. Nevertheless, we would like to keep the format in continuous text. Also, because it appears more common in NHESS.