

Point-by-point response to Reviewer #2

Florian Ehmele on behalf of all co-authors

October 31, 2019

Dear reviewer No. 2,

Thank you very much for your work and the useful and valuable comments that will help to improve the scientific quality of our manuscript. Especially your suggestion on how to implement the comments to the paper are very useful. Below you will find your comments given in gray and our responses to the individual points in black. Please also consider our comments to Reviewer 1 as there is some coincidence of the comments and the corresponding answers.

This paper analyzes long-term trends of heavy precipitation in multiple dynamically downscaled simulations for the historical period and the near future over Europe. The different sets of simulations are validated against gridded observations and tested whether they can be combined to a large ensemble for the detection of trends in the historic time period. This paper is relevant in terms of assessing the possibility of combining various simulations from the same RCM with varying driving data. As well as, the detection of trends within the historic time period.

General comments:

1) Please be more clear about what you are showing in the figures. In most cases it wasn't clear to me if you are showing the ensemble mean or a metric with pooled data from the entire ensemble.

Thank you for this feedback. We will change the figure caption to be more precise and accurate to become more clear. Therefore we will also include the related minor comments you wrote below.

2) Several sections need more clarification on what was analyzed and for what spatial extend and aggregation.

Going through your major and minor comments below and include them into the new version of the manuscript, we think this will clarify a lot of points within text. Please see our detailed comments to the specific points below.

Major comments:

. . . Model evaluation:

1) I have concerns with the comparison of E-OBS, CCLM and HYRAS over the sub-region AL. It is not clear to me whether the comparison was only performed for the HYRAS grid cells, which cover a substantially smaller area than E-OBS and CCLM, or whether E-OBS and CCLM represent the entire AL domain compared to a much smaller area in HYRAS. On P8-L204f you state this concern yourself '[. . .]' which might be a reason for the vanished differences between E-OBS and HYRAS and the resulting specious deviations to the RCM'. Did you compare the three datasets for the HYRAS grid cells only? If not please do so.

This is a crucial comment, thank you for that. Checking our data we found out that the analysis for the AL region indeed was performed for the entire region for CCLM and EOBS but HYRAS only for available grid cells. We will fix this in the new manuscript

version and also double check our results for the ME region to be done on HYRAS grid cells only.

2) Further, you state that '[. . .] by taking into account all grid points and all time steps within the investigation are (P6-L151)', does this mean that for both ME and AL you have included ocean grid cells in the spatial average of the RCM data? For both domains gridded observational datasets are only available over land. Please clarify this, and in any case ocean grid cells were included remove them from the comparison.

In every case ocean grid cells have been set to a missing value in every dataset and therefore they are not in the results. We will add a sentence on that in the method section for clarification.

2) In P8-L212f you state that '[. . .] HYRAS was aggregated to the E-OBS/RCM grid [. . .]'. However, you first mentioned this here for the Q-Q plots, so can I assume that the IPC's in Figure 2 are also based on aggregated HYRAS data? Please clarify this and if the aggregation of HYRAS applies to all related analysis then please move this detail to the methods section.

The HYRAS data have first been aggregated to the EOBS/RCM grid for all type of analysis in this study. We will clarify this by moving the corresponding statement more to the front of the manuscript into the method section as you have requested.

3) Further, if I understand correctly the evaluation is based on the TP1b time period (1950-2017), however the HYRAS data is only available for the period 1951-2006. Please comment on why the analysis wasn't based on the shorter HYRAS time period. I would recommend doing the analysis for 1951-2006.

In this case you are right with the different time periods. We assume that there will be only small changes when reducing TP1b to the HYRAS period 1951-2006 but nevertheless we will fix this for all analysis in terms of consistency.

4) The evaluation on such a highly aggregated level poses a risk of error compensation. It might be better to do the evaluation for each grid cell first (e.g. calculating the RMSE) and afterwards averaging the error metric.

We will restructure the evaluation part and will introduce additional skill measures in the evaluation, including grid-cell based measures.

5) Based on the concerns above, I don't really agree with your first point in the conclusion 'Extreme precipitation is well represented in the LAERTES-EU [P20-L352]'.

You and also RC1 have the same concerns about this first conclusion. Thinking about this a second time we came along that this statement might be too general. It was meant that heavy precipitation is consistent in all parts of LAERTES and that our results fit in the range of previous studies (e.g. Früh et al., 2010) and also in the range of observations knowing that the used observational datasets have uncertainties as well. We will rewrite this to be more precise.

. . . Added-Value:

6) Regarding your conclusions on the IPCs showing '[. . .] a clear added value of RCM data compared to coarser global models'. From that one figure I don't really see the added value, since you haven't compared the driving GCM with RCM simulation. You have compared the

IPC's to the 20CR reanalysis dataset. Because of the spatial averaging over such a large area, it might be that the trends in the GCM and RCM might be very close to each other.

To substantiate this statement we will include the IPCs of the MPI-ESM model in Fig. 2. At least we will include the IPC of data block 1 which used the LR version of MPI-ESM and if there is no loss in visibility we will also include the IPC of data block 3 which used the HR version as global forcing.

... CC-scaling

7) Your conclusion on the trends following the CC-scaling in your conclusion [P21-L379ff] are flawed. If you make a statement like this, please perform the temperature scaling with the LAERTES-EU temperature data and not by relating the precipitation change to a temperature approximation from another study. Please see Kröner et al (2017) and Pfahl et al (2017) for other effects than thermodynamics. Kröner et al (2017), Climate Dynamics, <https://doi.org/10.1007/s00382-016-3276-3> Pfahl et al (2017), Nature Climate Change, <https://doi.org/10.1038/nclimate3287>

We agree with the reviewer that a relationship should be established using LAERTES temperature data. We will do some brief analysis with the block 1 & 3 temperature data and put them into the context of the already cited studies on 20th century temperature changes. Then the argumentation should be more consistent and reasonable.

Minor comments:

P2-L25f: see also Zhang et al (2017) for a discussion on CC scaling Zhang et al (2017), Nature Geosciences, DOI: 10.1038/NGEO2911

P2-L27-34: Please add some more recent literature on this topic, e.g. Fischer and Knutti (2016), Nature, DOI: 10.1038/NCLIMATE3110 Alexander (2016), Weather and Climate Extremes, <http://dx.doi.org/10.1016/j.wace.2015.10.007> Barbero et al (2017), GRL, doi:10.1002/2016GL071917

P2-L45f: Please add a view more recent studies on trends in European floods. E.g.: Blöschl et al (2017), Science, DOI: 10.1126/science.aan2506; Blöschl et al (2019), Nature, <https://doi.org/10.1038/s41586-019-1495-6>

P2-L49f: Connection of Heavy Precipitation over central Europe and cyclones, see also Hoffstätter et al (2017), Int. Journal of Climatology, <https://doi.org/10.1002/joc.5386>

P2-L55f: Also see van der Wiel et al (2019) and Martel et al (2019) for the added value of large ensembles for flood risk or return periods of heavy precipitation van der Wiel et al (2019), GRL, <https://doi.org/10.1029/2019GL081967> Martel et al (2019), Journal of Climate, <https://doi.org/10.1175/JCLI-D-18-0764.1>

Thank you very much for this recent references. We will go through them and decide which one and where to include them in the new version of the manuscript.

P4-L114f: Please elaborate more on what you mean by '[. . .] more or less independent simulations'

We agree that this formulation is inept and we will remove it. What was meant is that the temporal evolution of the day-to-day weather in hindcasts is independent after a few weeks. The statement did not refer to the model system. The ensemble does not cover the full range of uncertainty, namely the model uncertainty. But, in the context of the paper we regard this as an advantage, since the dataset is homogeneous over time due to the consistent model setup.

P5-L139: What do you mean by un-initialized? Please clarify this for the reader, that by initialized you mean initialized by observational(-like) salinity and other variables, whereas the un-initialized data originate from a normal CMIP5 historical simulation. I had to go to Marotzke et al (2016) to understand what was meant by this.

We agree that this has been explained insufficiently and will state it more clearly in the revised manuscript.

P6-L157f: Are the 99th and 99.9th percentile based on all days or wet-days only? If you want to look at heavy precipitation it might be better to look at wet days only. Like this the values would not be affected by the dry-day adjustment as much. Further, it is not clear to me if you have first spatially aggregated and then calculated the percentiles, or the other way around. Please comment on whether you think that this will make a difference to your results. This could maybe also solve your concerns on P15-L282f '[. . .] an overestimation of precipitation [...], could be a result of missing data for the applied dry-day correction.'

The percentiles were estimated using all days including dry ones. As we focus on heavy precipitation we agree with you that using wet days only would be more appropriate. Nevertheless, the uncertainty in the first half of the century will remain. The dry-day correction adjusts the number of days without precipitation solely and not the values themselves. This means that the dry-day correction effects the percentiles anyway in some case. But, in order to get a more thorough analysis of heavy precipitation we will change to a wet days only calculation. Regarding your second point, we first did a spatial aggregation of precipitation to receive the areal precipitation and then calculate the percentile of this spatial mean values which are of deep interest in this study. The other way round we would get a spatial mean value of the percentile which is more relevant for a 2D analyses and related spatial variability giving local effects. We will include some sentences on that in the manuscript.

P7-L189: Could you briefly comment on why you chose the old 1961-1990 period as your reference climatology.

A couple of studies (e.g. Cahill et al., 2015 or Folland et al., 2018) showed that the climate change signal at least for global mean temperature is significantly increased since the early 1980s, which is to a lower degree applicable for Europe, too (Folland et al., 2001). Therefore, using the time period 1981-2010 would possibly include a strong changing signal to the analysis. Using 1961-1990 reduces the influence of these effects as this period shows more stable conditions to a certain degree. Doing so, there is more room for the interpretation of the future projection instead of comparing them to the directly preceding time period.

References:

Cahill et al. (2015), DOI: 10.1088/1748-9326/10/8/084002

Folland et al. (2018), DOI: 10.1126/sciadv.aao5297

Folland et al. (2001), DOI: 10.1029/2001GL012877

P9-L218f: Your conclusion to Table 2 stating that there is a higher correlation when driven by MPI-ESM-HR versus lower resolution MPI-ESM-LR is technically correct, however the differences are so marginal that I find it difficult to attribute the differences to resolution of driving data. Especially, when not only the resolution is different in the HR and LR simulations,

but also the initialization. Maybe add a short sentence [‘However, differences are only marginal.’].

Thanks for that comment. Yes, the differences are marginal but the differences between the LR and HR blocks are larger than those within the LR blocks or within the HR blocks. We will include a statement such as the suggested one to the manuscript.

Chapter 4.3: This is a nice analysis that shows the benefit of large ensembles, however since you are not looking at return values afterwards it could be nice to highlight another strength of large ensemble namely isolating the forced response from internal variability. Since you are looking at trends and variability this could be a better fit. But this is just a suggestion to improve the flow of the paper. Like a said it is a nice analysis as is.

We decided to use return values in this case as it easy to estimate and on a statistical perspective the amount of data has a significant influence on the estimates. Although you like the presented analyses, we would like to change Fig. 5 a little bit. For this particular figure the simulations were put together starting with block 1 simulation 1 and ending up with the last simulation in block 4. Doing so the shape of the curves strongly depend on the length of the single simulation runs. Therefore, we want to change the figure using the mean values of 100 random combinations of the simulation runs. The values of the signal-to-noise ration will not change that much and the given statements remain valid.

Figure 8a: Shouldn't there be also some more positive anomalies in the climTP period? Did I miss something? Because if you base the annual anomalies on this period, shouldn't you be having positive and negative anomalies within this period?

We are sorry, but unfortunately there was a wrong figure included at this point. Of course there should be and there are positive and negative anomalies in the climTP period. We replace the current plot with the correct one.

Figure 9: Nice plots!

Thank you!

P22-L396f: ‘[. . .] can be used as input for hydrological modeling’. In general, yes and especially when looking at higher return levels of floods. However, as mentioned a few lines above this ensemble is restricted by temporal homogeneity, which can play a very important role in hydrology.

In general, we agree with that and it definitely makes sense when investigating historical events, trends, flood frequency, and so on. In a particular application and as mentioned in the following sentence LAERTES serves as stochastic weather generator which leads to quasi stochastic hydrological simulations covering the internal climate variability and also a wider range of values occurring. For such statistical applications, LAERTES can be used to get robust hydrological statistics, too. For this specific case, it is necessary to do a bias correction to avoid too high discharges as a consequence of an overestimation of precipitation. This application as well as the bias correction will be part of a consecutive study (Kautz et al., planned submission in late 2019).

Technical Corrections:

Table 1: Projections for the period 2020-28 are missing

The projections are included in block 4 as the given year stand for the initialization years of the decadal simulations meaning an initialization in 2018 includes data for 2019-2028. We will change Table 1 accordingly so that it is clear that 'period' means the covered years.

Figure 2, 3: Please add the years of the period ([. . .] TP1b (1950-2017)). I had to go back and look for the TP1b definition. But I would anyway suggest changing the period to 1951-2006 (see major comments).

Both comments will be implemented in the revised version.

P3-L69: Typo ('Regionla', 'Regional')

P5-L128: grammar (replace 'it' with 'the')

P20-L342: grammar ('estimate' instead of 'estimated')

Thanks! We will fix that.

Figure 4: Please clarify what the RCM spread is. I assume it to be Min-May, right?

That's true. The RCM spread means the range between the minimum and maximum occurred value of the displayed variable. We will include a short clarification into the text.