# Point-by-point response to Reviewer #1

Florian Ehmele on behalf of all co-authors
October 31, 2019

**Dear reviewer No. 1,**
**Thank you very much for your work and the useful and valuable comments that will help to improve the scientific quality of our manuscript. 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 2 as there is some coincidence of the comments and the corresponding answers.**

*This manuscript addresses the issue of heavy precipitation in RCM simulations. This is a very timely issue with importance for many sciences, which rely on RCM simulations. My fundamental concern with this study is the first conclusion ("Extreme precipitation is well represented in LAERTES-EU."). The same is expressed in the authors' short summary ("The simulations show a good agreement with observations for both statistical distributions and time series of heavy precipitation."). I am sorry, but I just can not see enough support for this crucial statement in the manuscript.*
**You and also RC2 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 what was meant to be stated here. However, we do think that, for instance, the IPCs do support the statement in terms of the statistical distribution of precipitation values. The time series of LAERTES (ensemble mean) is within the range of both analyzed observational data sets, and the ensemble spread covers the observed variability. Please note that it was never intended that LAERTES shows a one-by-one agreement with historical events.**

*1) The authors state that E-OBS underestimates precip by almost a third. To me, this means that these data are not useful to evaluate the performance of extreme value simulations. As E-OBS is only available for land surfaces, I also find it surprising that the ME box includes parts of the North Sea.*
**E-OBS has some limitations, like a certain underestimation especially for extremes, but these have been mentioned by several previous studies like Haylock et al (2008) or Hofstra et al. (2009). As these and other studies already used and analyzed E-OBS, we follow their conclusions and did not perform a further analysis on data quality. Keeping the limitations in mind E-OBS can be useful for evaluation. Unfortunately, there is no other high-resolution daily precipitation data set available that covers entire Europe for a quit long time period. As the focus of this study is on intensive areal precipitation we think it does not make sense to use single ground based observations that potentially are available for longer time scales, or in terms of the focus on long-term evolution other products like satellite data with a very limited time frame are not helpful and also have limitations. We will add a comment on this situation to the revised manuscript.**

**The prudence region are defined as regular lat/lon-boxes and therefore cover ocean areas as well. But, 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) The evaluation using IPCs is good, but doing this on a highly aggregated level seems to limit the opportunity to really test the simulated precip. Here I would like to see more creative tests such as IPCs for smaller areas and/or IPCs for certain seasons. As the analysis is done now, there is a risk for error compensation.*

**For the IPCs no aggregation was done. We take all grid point values within the investigation area (e.g. ME) and at all timesteps and group them into a histogram giving the probability of occurrence (=IPC). On purpose, we only use all year data and no seasonal differences as the paper would have become to long doing a seasonal analysis for every part of it. A seasonal as well as a more detailed 2D analysis is planned for a following study. This study was meant to be an introduction to LAERTES and some long term investigations on the upper part of the precipitation distribution.**
**For a more appropriate evaluation of LAERTES we will follow Reviewer 2 and include some further analysis using other quantities like RMSE, skill scores, or related estimates.**

*3) I am sorry, but I do not see how the Q-Q plots help to evaluate the performance for heavy precip. If anything, the total precip is evaluated. But even then, comparing cumulative values introduces a spurious correlation, and on top of that, R2 is no suitable measure as a value of one does not ensure a 'perfect' model. May be I miss something here, but I find this analysis not convincing.*

**We will restructure the evaluation part and will introduce additional skill measures in the evaluation (see also answers to RC2).**

*4) The 99% of precip (=around 10 mm) is not really 'heavy precipitation.'*

**Technically, this is correct. However, the focus of this study is intensive areal precipitation, which is related to widespread flooding along the great major river networks. For a single grid cell, 10 mm is no big deal but 10 mm on average over a large area is quite a lot rain. Maybe the term 'heavy precipitation' somehow is irritating at this point. We will include a clarification on that at the beginning of the revised manuscript. Furthermore, please see also the report of Reviewer 2. We will change the percentile calculation to wet days only, as currently dry days are included.**

*Minor comments:*
*P4L115: "more or less independent simulations". This needs to be clarified. In some respect, these simulations might be independent, but as the same RCM is used, the simulations obviously are dependent!*

**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.**

*P6L149: does this mean there was a bias correction? Were extreme precipitation simulations affected by this at all? I assume not but would like to get some clarification.*

**The dry-day adjustment only corrects the number of days without precipitation (R<0.1mm/day) in the model as RCMs tend to produce too much days with very small rainfall amounts (drizzle effect). The absolute values are not affected. A bias correction meaning an adjustment of the absolute precipitation values by e.g. a quantile mapping was not performed at this stage. In a consecutive study (Kautz et al, planned submission in late 2019), a specific application of LAERTES for hydrological issues will be presented for which such a bias correction is mandatory. For any other application a reduction of the drizzle effect has to be done anyway.**

*P7L197ff: I can see the argument that GCMs underestimate heavy precip, but the same argument should, although to a smaller degree, apply to RCMs. So, what is the physical reason that RCMs 'tend to overestimate precipitation intensities'?*

**Two effects are of relevance at this point and which act together. The limited time period of observations results in unknown distributions especially at the heavy tail. In a dataset of 65 years, extreme events with return periods of 100 years or more are not represented in a statistically robust way. The RCM have a physical background when calculating precipitation amounts which makes it possible to reach higher than observed values. Furthermore, the huge number of simulations allows for a more robust estimate of the high-intensity tail of the distribution, whereas the observations display only a few single events in this range.**