This manuscript provides a comprehensive analysis and evaluation of ensembles of RCMs with parameterized convection and CPMs with explicitly resolved convection for two regions that are likely to see changes in rainfall convective rainfall extremes. The topic is well researched with a comprehensive background section, the analysis is reasonable, and the results are convincing. Further, the authors provide an insightful discussion on their results and the paper is almost entirely devoid of grammar issues. I think the paper is valuable and an interesting contribution to the literature and, accordingly, I recommend the paper be published pending the resolution of a handful of minor comments.

Thank you very much for your review, and helpful comments. Below you will find a response to them.

I have two key questions that the authors should address so as to better contextualize their results and conclusions.

1. What is the specific resolution (or range of resolutions) of the CPMs used in the study? The RCM resolution is listed at 12km, but I don't recall seeing the CPM resolution described in either the manuscript or the supplementary material. This is important as numerous studies show that for convection, diurnal convection in particular, 2km resolution is generally superior at reproducing the timing and intensity of diurnal convection (e.g., https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2014RG000475). If the models used in this study are coarser resolution than 2km, I think that's fine since the main consensus within the community is at 4km cutoff. That said, some context here would really improve the manuscript. On a similar note, any comments the authors can provide on the LSMs used in the models described in study would be helpful since at sub 4km resolutions, simulated convection becomes increasingly tied to the LSM. (e.g., are there any known biases in the LSMs that might impact the results or explain some of the shortcomings in the CPMs?).

   The resolution of the CPMs model is 2.2 km (COSMO) and 2.5 km (UKMO, HCLIM and AROME). This was indeed missing. We will also make a reference to the paper. We do agree that the soil scheme has a large influence on rainfall, for instance because land surface heterogeneity may influence the triggering on rain systems. But at the same time LSMs are so complex that it is hard to pinpoints at problems. It may be the formulation of the LSM, but also the underlying data base of soil and vegetation properties can strongly affect results. We therefore think this is outside the scope of the paper. We will add a few underlines the importance of the LSM in the discussion.

2. What is the generalized modality of the convection that makes up the extreme rainfall events (95 percentile, e.g.) in the regions described in the study? Diurnal air-mass "pop-corn" ordinary cells? Organized MCSs? Terrain-initiated convection? Convection forced by or embedded within synoptic scale systems (e.g., fronts)? Mix of all of the above? The authors provide an excellent discussion on the complexities of convection and extreme rainfall and how changes in absolute and relative humidity in the surface can influence or be related to cloud-scale dynamics, however without information on what convective modes are being simulated, this discussion is without context. For example, the authors discuss convective plume size vs. dry air entrainment as a physical reason for the observed behavior of extreme rain vs. dew point depression. However, my understanding is that this process is mostly only relevant for continental convection

associated with very dry mid-level air where wider convective plumes associated with larger surface DPDs can protect the convective core from dry-air entrainment in the mid-levels. If, for instance, the convection most associated with extreme rain in the NL or SFR was more characteristic of air-mass convection with very moist conditions throughout the atmospheric column, then this process likely isn't a great explanation for the DPD/extreme rain relationships observed. To be clear, I am not arguing that this process *isn't* relevant in this instance, I'm just arguing that without added context on storm modality, the authors' speculation on physical processes carries less weight. A brief discussion of predominant storm modality, and perhaps an example figure showing a snapshot of simulated rain-rate compared to radar image for an extreme rain event would be extremely helpful to the reader and would improve the manuscript. Such a discussion would also relate back to CPM resolution, since the differences between 4km and 2km resolutions would be quite different for a synoptically forced MCS vs air-mass thunderstorms.

These are good questions, and we thank the reviewer for asking them. Definitely, the statistics we get are based on many different events. We tried to rule out large orographic effect by focusing on stations with altitude below 400 m. For The Netherlands orography definitely does not play a large role as most of the country is very flat. In a previous paper we looked at the large-scale conditions associated with the events (Lenderink G, Barbero R, Loriaux JM, Fowler HJ (2017). https://doi.org/10.1175/JCLI-D-16-0808.1). Relative humidity associated with extreme events is shown in Figure 6 of that paper. We also found that the extreme statistics are dominated by larger scale events, where rainfall is embedded into large scale disturbances, with substantial large-scale lifting (e.g. Figs 6 and 8 of the paper). We also found that rainfall itself is not more intense when it occurs in larger clusters – the hourly rainfall distribution is exactly the same for small and large clusters – but since rainfall occurs in much larger areas they dominate the extreme statistics (Fig 3cd). This also appears to collaborate with the finding from Large-Eddy Simulation that large-scale lifting mostly affects the size of the systems, but that instability mostly affects intensity (Loriaux JM, Lenderink G, Siebesma AP (2017) https://doi.org/10.1175/JCLI-D-16-0381.1). We will add a brief description of this in the paper.

Finally, we note scaling is rather robust, and that sub-selecting on large-scale circulation types only moderately effects the results (see Figure below). Thus, it appears that many of the sensitivities we find are not so dependent on the type of convection and associated large-scale circulation patterns. But, it would also be interesting to investigate this further.

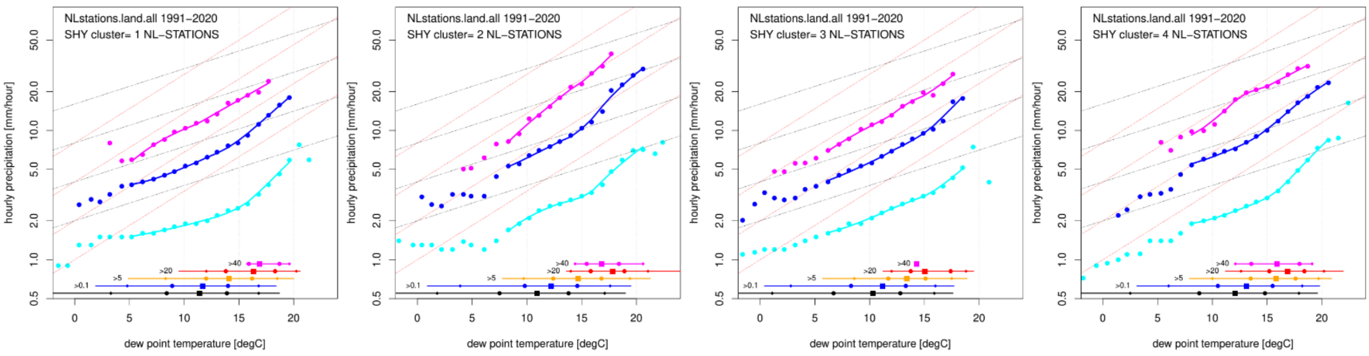

Figure 1. Scaling of hourly rainfall extremes, sub-selected on circulation type. From left to right, westerly flow, easterly flow, northly flow and southerly flow patterns over the Netherlands.

Minor comments:

1. Why not look directly at CAPE or CIN? When discussing /speculating on some of the convective processes and how they relate to changes in absolute and relative humidity? Were these data simply not available?
   No, this data is not available. Some models may have it, but there are definitely no reliable soundings for all stations, so we have to rely on reanalysis data. In earlier work we looked at CAPE and found that is quite sensitive to the quality of the surface observations and the time lag with the precipitation (Loriaux JM, Lenderink G, Siebesma AP (2016) https://doi.org/10.1002/2015JD024274; e.g. Figure 7). So, we think this is a much less robust measure.

2. Why use DPD instead of RH directly?
   This could have been done, but we prefer to keep everything in a "temperature" space. E.g. from the DPD it is easy to derive an approximate value for the cloud base given an undiluted surface parcel (LCL ~DPD*100 m).

3. I'm unclear as to the use of the 5x5 pixel sampling, and how it was used in the data comparisons, the mean was computed, but I don't know if it was actually used?
   This point was also noted by the other reviewer. In addition to the "sample" based value, we also did the analysis on the mean, to investigate whether dependencies on spatial average could affect our results, and explain part of the differences between the CPMs and the RCMs. But, it turned out that differences are rather marginal in most plots. We will show some results in the Supplement.

4. Line 246/247: dew-point temperature, TD, and TDD should be modified: TD and TDD. (no need to repeat dew-point temperature).
   OK

5. It is my understanding that the analysis performed comes from simulations that the authors themselves did not perform, but this point is not entirely clear as it relates to the

CPMs. I suggest the authors clarify this point.
Indeed, the groups that performed the CPM simulations are co-authors of the paper, but the RCMs were obtained within Copernicus Climate Change Service. We will check that in data statement.