The authors have partly addressed my comments. Some of the changes or clarifications have led to additional enquiries or issues, which are listed in the following:

Major comments:

1) I am still not convinced that the chosen climate simulations ("DWD core ensemble") are suitable for this analysis.

   a. Convective processes are parametrized in the 0.11° simulations. However, the study tries to assess changes in 5-min rainfall and rainfall extremes, whereby the extremes in Germany are mostly governed by convective processes at this time scale. The authors argue not to use available convection-permitting model simulations, as these simulations are not available as ensemble. However, they could include such analysis to compare the results and evaluate potential deviations of convection-permitting versus parametrized setups.

   b. The authors claim that the bias-adjustment is a major advantage of the data set. However, they do not discuss or reflect on the associated issues of such a bias adjustment for their study. Which kind of quantile mapping is applied? How are extreme values handled therein? Are trends preserved? Please clarify that temperature and precipitation are adjusted independently. This is from my perspective a major issue for the "physics inspired" temperature-dependent disaggregation, and needs to be discussed as well. Adjusting precipitation and temperature independently breaks the climate-model inherent physics. I'd suggest to analyse the dependence structure of daily precipitation and temperature for the observations and the climate models during the reference period (C20). This analysis should also explicitly address rainfall extremes.

   The bias adjustment largely governs the analysed output at the daily resolution, and its implications and limitations need to be well understood by the authors and carefully presented to the reader.

   The authors claim that they are mostly interested in the climate change induced changes – why is bias adjustment needed then?

   c. The DWD core ensemble provides daily resolution only. The aim of the whole study is the analysis of 5-min and 1-h rainfall. From my perspective, the disaggregation from daily to 5-min resolution induces larger methodological uncertainty than a disaggregation from hourly climate model output to 5-min resolution.

   d. For extreme rainfall, the authors have not shown the suitability of the DWD core ensemble. They assume that the bias adjustment has led to a proper representation of extremes, however the adjustment of extremes is not straightforward due to the limited sample size.

2) In L520, the authors argue: "The key assumptions for the application of cascade models for the disaggregation of future climate model data is that the scaling behaviour of rainfall remains stationary, which is not questioned to the authors knowledge."
   If this assumption is key, the authors should not assume it on the basis that they are not

aware of any other studies that call it into question. They should substantiate the assumption themselves. For example, they could investigate the stationarity of the respective scaling based on convection-permitting simulations (e.g. https://esgf.dwd.de/projects/dwd-cps/cps-scen-v2022-01). Or alternatively, they should refer to other peer-reviewed studies, which show that this scaling is stationary under strong climate change.

3) L312: As far as I understand the procedure, you are using the plotting position formula. Assuming 30-year periods, you include 2.4*30 events = 72 events. You analyse up to 10-year return levels. Is that the intensity of the 3rd most intense event? Or in between the 3rd and 2nd most extreme? How do you handle extreme value statistical uncertainties?
Following up on this calculation: L475: Can you show return level return period plots for the disaggregated 5min and 1h extremes? Can you provide a map to present the spatial pattern of the return levels? Does it follow the topography (as in KOSTRA) or is it more chaotic as in RADOLAN (see Fig R1). Can you provide an evaluation of the disaggregated 2-year and 10-year return levels of 5min and 1h for the reference period (C20) compared to official rainfall guidelines return levels from KOSTRA-DWD? (https://www.dwd.de/DE/leistungen/kostra_dwd_rasterwerte/kostra_dwd_rasterwerte.html). A successful evaluation would make a strong argument that your disaggregation procedure works for rainfall extremes for the reference period.

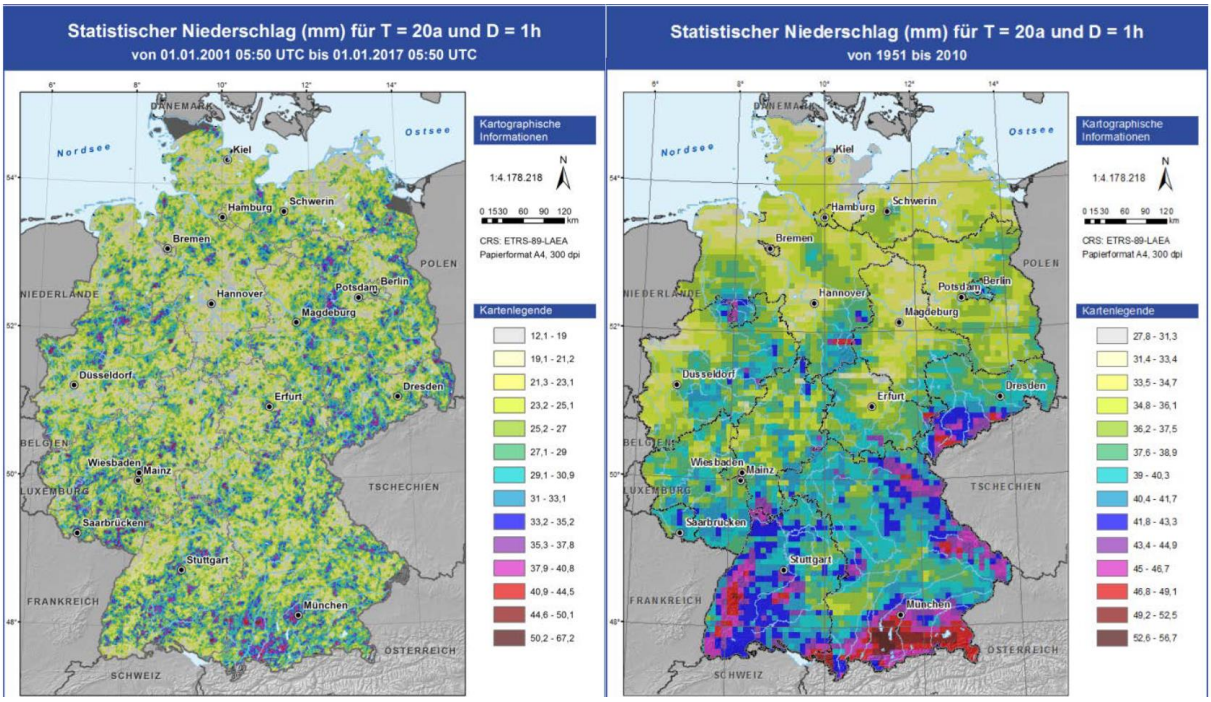

Figure R1: 1-hourly 20-year return levels based on radar (left) and station data (right). Taken from Winterrath et al., 2017:

Winterrath T, Brendel C, Hafer M, Junghänel T, Klameth A, Walawender E, Weigl E and Becker A 2017 Erstellung einer radargestützten Niederschlagsklimatologie (= Berichte des deutschen Wetterdienstes 251). Offenbach, Selbstverlag des Deutschen Wetterdienstes.

Minor comments:

The authors have addressed the majority of the minor comments of the previous review sufficiently.

L145: I'd still prefer a readable elevation legend in Fig. 1 instead of the squeezed micro-colorscale.

L175ff.: The bias adjustment is not explained sufficiently. See major comment 1b.

How is the RCM-inherent drizzle handled?

L312: Plotting positions / extreme value analysis: see major comment 3)