Review by Benjamin Poschlod, Referee #1

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:

We thank Benjamin Poschlod for his useful and constructive comments and the time, he spent on the manuscript. A point-by-point reply can be found below. For a better overview, the reviewers' comments are shown in black, our replies in blue and proposed changes in the manuscript bold and blue.

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.

The reviewer remains unconvinced regarding the suitability of the DWD core ensemble (CE) for our analysis. To check the plausibility of our results, we compared them with the convection-permitting climate model (CPM) provided by the DWD, as recommended by the reviewer. Specifically, we compared rainfall extreme values with return periods of T=2 yrs and T=10 yrs for a duration of D=1 h between the CE and the CPM for the station subset A-E in the long-term future (2071-2100) (Fig. 1). A comparison is only for long-term future possible, as the near-term future covers a different period than in our study and for C20 no data was available via the link provided by the reviewer. In addition, the CPM data had to be aggregated to match the spatial resolution of the DWD core ensemble. Our analysis demonstrates that the extreme rainfall values obtained from the CPM fall within the range of those from the CE, thereby validating the plausibility of our results.

However, we did not include this comparison in our study in order to limit the results on the analysis of the DWD core ensemble only. The main focus of our study are rainfall extreme values with a subhourly temporal resolution. The CPM data has an hourly temporal resolution and therefore offers only small value in checking the plausibility of hourly rainfall extreme values. However, in our study we focus only on relative changes in rainfall extreme values between the C20 and the future. We have refrained from presenting absolute values in our study, as climate scenario analyses priorities relative changes above absolute changes and values. Relative changes for the future in CPM 20 could not be analysed as the C20 period was not available.

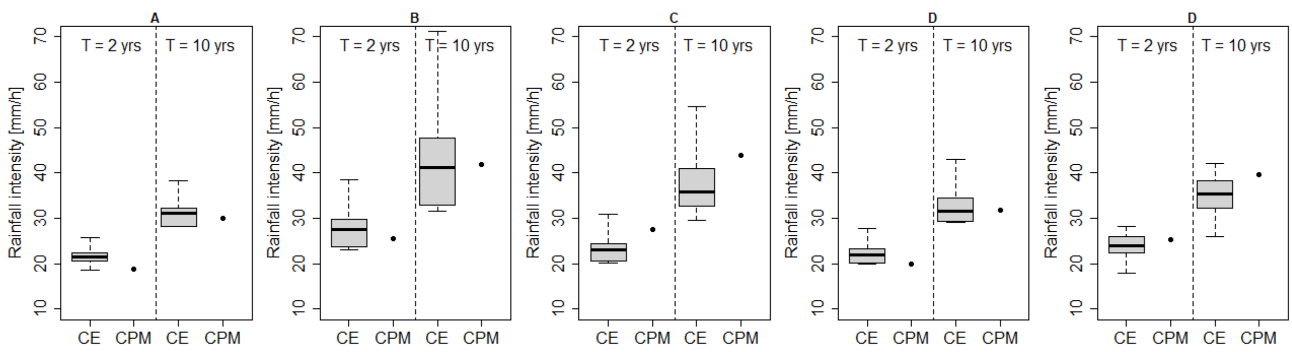

Figure 1: Comparison of rainfall extreme values with a return period of T=2 yrs and T=10 yrs for a duration of D=1 h for the disaggregated DWD core-ensemble (CE) and the convection-permitting climate model (CPM) proposed by the reviewer for the station subset A-E in the long-term future (2071-2100).

> 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?

As pointed out in our last reply, the bias-adjustment of the DWD core ensemble is outside of the scope of our study. We used bias-adjusted and spatial downscaled climate scenario data to analyse realistic climate scenario data sets in our study and as Fig. 1 shows, we can generate plausible rainfall extreme values with the bias-adjusted climate data.

Nevertheless, we provide in the following some details of the bias-correction. The method used for quantile mapping was the Quantile-delta change mapping (QDCM) developed by Ollson et al. (2009). QDCM involves the use of simulation data in the projection period to conserve projected changes in the quantiles of the climate projection through quantile-by-quantile adjustment.

In QDCM, both observed and modelled values (for historical and future periods) are initially sorted in ascending order and assigned to percentiles. The comparison of modelled and observed percentiles provides a specific bias (or adjustment value) for each percentile. Subsequently, historical or projected model data is then treated percentile by percentile with the corresponding adjustment value. Given that the highest temperature values, and consequently the highest percentiles, typically occur toward the end of the 21st century due to climate change, it's important to note that values at this juncture would frequently receive an adjustment value primarily tailored to high percentiles in the historical period. Conversely, a significant portion of values at the onset of the projection period (from 2006) would be adjusted with a value corresponding to a medium or low percentile in the historical period. To mitigate this issue, a "moving window approach" is employed, wherein climate projection data was sorted or adjusted for 35-year time slices, and the middle 15 years were retained.

The rainfall was adjusted univariately using QDCM. It should be noted that the QDCM method was only applied to values up to the 99.9th percentile, as rainfall amounts above the 99.9th percentile are not adequately represented in the reference and projection data. For values above the 99.9th percentile, the adjustment value was extrapolated linearly. Reviewer #1 points out that the univariate QDCM is a violation of the 'physics' behind the RCM results. We agree, but at the scale applied for the QDCM we have the transition from physical coherent simulations (so the regional climate model) to statistical methods anyway (the multiplicative cascade model). So, the application of the CDCM is not an 'additional' violation and is taken into account by the authors as a necessity.

The possible extension of the bias-adjustment description would be:

“For the bias-adjustment of the rainfall an univariately approach using the quantile-delta change mapping method was chosen by Hänsel et al. (2020). “

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.

The reviewer is right. A disaggregation from 1h to 5 min induces smaller methodological uncertainties compared to the disaggregation from 1d to 5min. Nevertheless, we have chosen this approach to analyse multiple GCM-RCM combinations to provide a range of possible future extreme values rather than choosing a single GCM-RCM combination.

Also, it's important to mention that the disaggregation from 1 d to subhourly time series is an established method with acceptable uncertainties and has been carried out in several studies (e.g. Bürger et al. (2019), Müller and Haberlandt (2018), Müller-Thomy (2020), Derx et al. (2023), Pidoto et al (2022)). However, in the “Summary and Conclusion” section we pointed out that the disaggregation is a tool and can be applied on other climate data with finer temporal resolutions.

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.

It remains unclear for the authors how the ‘adjustment of extremes is not straightforward due to limited sample size’. From our understanding the bias-correction was done with the best available data. The assessment, bias adjustment, spatial downscaling and selection of climate projections for the DWD core ensemble were conducted by climatologists and is beyond the scope of our study (e.g. Hänsel et al., 2020). However, it is noteworthy that the DWD core ensemble represents the state-of-technic climate scenario data for Germany and is recommended by the DWD for climate scenario analysis. This information has been included in our study to highlight the credibility and relevance of the dataset used. Again, if reviewer #1 can provide another climate model data set with different bias-corrected GCM-RCM combinations, we will add it in the outlook.

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.

We agree with the reviewer. The assumption we made is quite important for our study and should be verified beforehand. To do this, we used the data from the station Bochum, which is used in the supplement of our study.

The station Bochum (N51.5026 °, E7.2289 °) has a time series length of 45 years, consisting of the periods 1940-1959, 1979-1993 and 2008-2017. This time series is not part of the data set used in the accompanied manuscript, but was selected due to its time series length. Based on the annual temperature the time series was split into two time series, one with colder years and the other one with warmer years. The threshold to classify in cold and warm years was an annual mean temperature of 10.5 °C. The two resulting times series have a length of 21 years for $\text{condition}_{\text{cold}}$ and 24 years for $\text{condition}_{\text{warm}}$ each. The mean temperature difference between the two time series is 1.2 K. This temperature difference is comparable to the temperature increase (approximately 1.2 K) from C20 (1971-200) to the NTF (2021-2050) that is projected by the climate scenario data based on RCP 4.5 across all locations in our study.

To compare the scaling behaviour of both periods, we calculated the first three moments (Fig. 2). All three moments show an almost identical scaling behaviour in both time series. For coarser temporal resolutions, the scaling difference increases. However, the difference is relatively small (<5%). This analysis serves as evidence supporting our assumption.

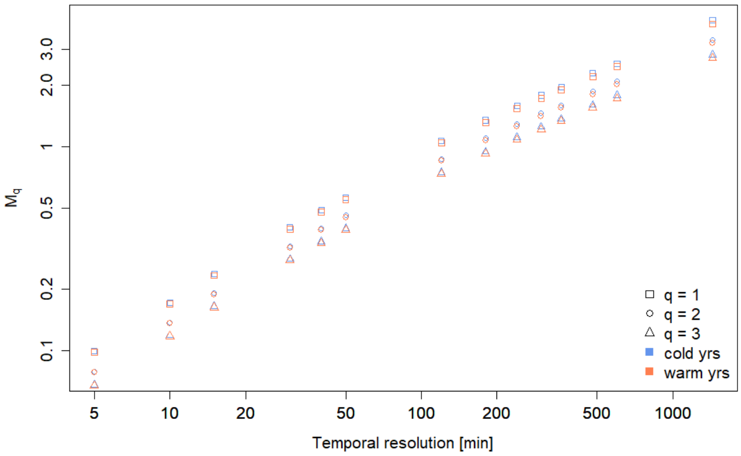

Figure 2: Probability-weighted moments of rainfall time series for the years with $\text{condition}_{\text{cold}}$ and $\text{condition}_{\text{warm}}$ for station Bochum (based on Müller and Haberlandt, 2018).

In accordance with the reviewer's recommendation, we further investigated the scaling behaviour using the convection-permitting climate model (CPM). The CPM provides data with an hourly temporal resolution, and through the link provided by the reviewer, only the near-term future (NTF) and long-term future (LTF) data were accessible. Consequently, we were able to assess the scaling behaviour between NTF and LTF only at hourly and daily resolutions (Fig. 3). Notably, the scaling behaviour observed in both time periods is highly similar. While there are slight variations in scaling behaviour for coarser temporal resolutions, the differences remain negligible. Thus, our assumption remains valid, supported by this additional analysis.

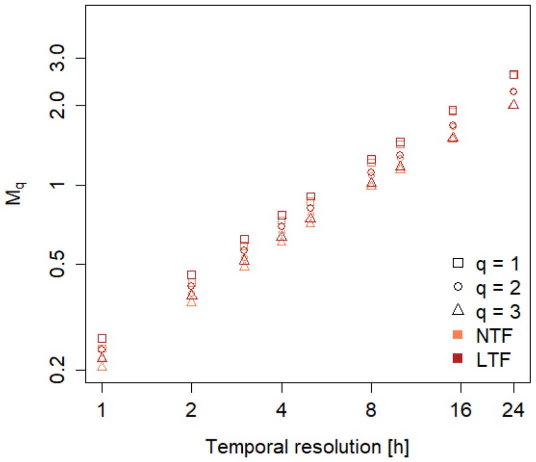

Figure 3: Probability-weighted moments of rainfall time series for the near-term future (NTF) 2031-2060 and long-term future (LTF) 2071-2100 for station Mühldorf with climate scenario data from convection-permitting climate model.

We propose to add the following text in the manuscript and the scaling behaviour analysis for station Bochum to the supplementary material section of the paper:

"The key assumptions for the application of cascade models for the disaggregation of future climate model data is the stationary scaling behaviour of rainfall, which was proven in the supplementary material with additional data."

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 3 rd 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.

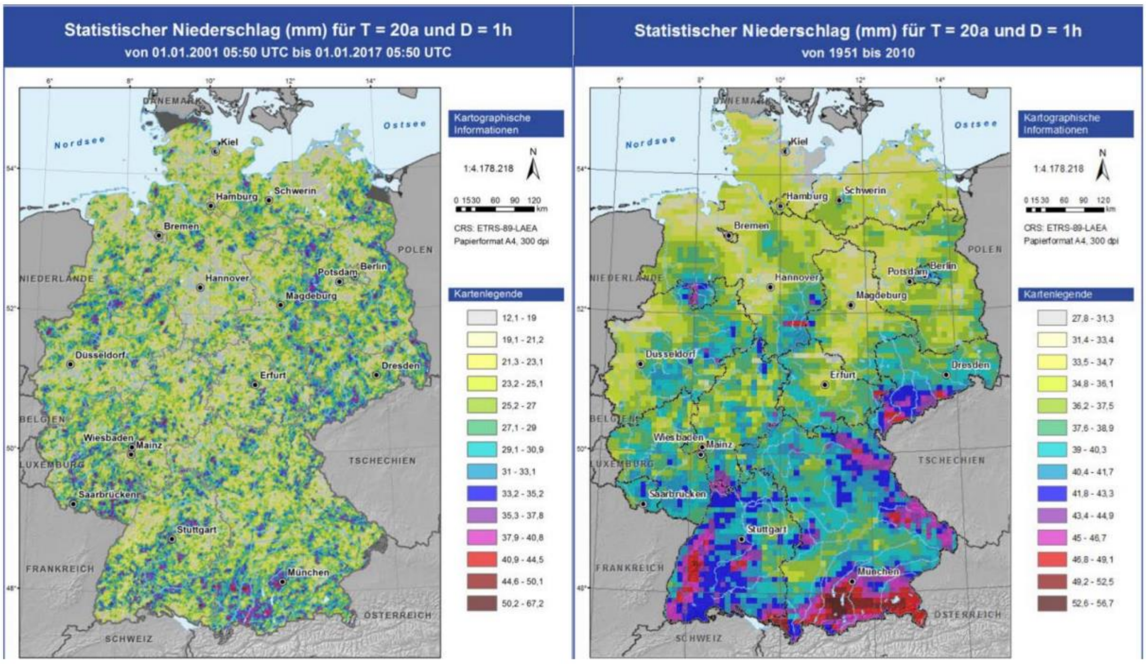

Figure R1: 1-hourly 20-year return levels based on radar (left) and station data (right). Taken from Winterrath et al., 2017:

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.

The rainfall extreme values were analysed according DWA-A 531. The empirical return periods T were estimated with the formula given in our study:

$$T = \frac{L+0.2}{k-0.4} \cdot \frac{M}{L}$$

, where L is the number of rainfall events that is considered to be 2.4 time the length of the analysed time series number in years (M) and k the running index of the sample sorted by size. The rainfall intensities assigned to return periods $d_R(T)$ were identified with the exponential distribution function:

$$d_R(T) = u + w \cdot \ln(T)$$

, where u and w are parameters determined by linear regression plotting dR(T) against ln(T). u corresponds to the y-axis value (dR) at ln(T)=0 and w is the slope of the linear regression.

A rainfall extreme intensity with a return period of T=10yrs is between the third and fourth most extreme rainfall intensities.

Regarding the second part of the reviewers comment, we are uncertain about the reviewers intentions. We have validated our extreme values from the disaggregated time series with observed values as presented in Fig. 8 of our study. We believe that a comparison of rainfall extreme values from climate scenario data with KOSTRA data is not plausible, since KOSTRA is a point dataset while the climate scenario data is a raster dataset. This difference in spatial resolution makes a comparison of extreme values impractical. Additionally, creating a map of rainfall extreme values for Germany is not feasible, as our analysis was focused on 45 specific locations within Germany. However, we have illustrated the changes in rainfall extreme values at these locations on a map of Germany in Fig. 11 of our study.

To clarify our approach regarding the calculation of return periods and rainfall extreme values we propose the following change to L312:

"First, the return periods of rainfall extremes values of the disaggregated time series are analysed. Therefore, empirical return periods (T) are estimated according to the German guideline DWA-531(DWA-531, 2012):

$$T = \frac{L+0.2}{k-0.4} \cdot \frac{M}{L} \quad (4)$$

, where L is the number of rainfall events that is considered to be 2.4 times the length of the analysed time series number in years (M) and k the running index of the sample sorted by size. The rainfall intensities assigned to return periods $d_R(T)$ were identified with the exponential distribution function:

$$d_R(T) = u + w \cdot \ln(T) \quad (5)$$

, where u and w are parameters determined by linear regression plotting $d_R(T)$ against ln(T) using (4). The return period allows a validation of the most extreme rainfall events."

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.

We would propose to change the elevation legend in Fig.1 to:

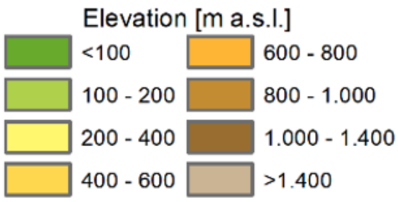

Figure 4: Revised elevation legend in Fig. 1 of the manuscript.

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

How is the RCM-inherent drizzle handled?

As explained above the bias-adjustment of the DWD core ensemble is not in the scope of our study. In our study we have shortly described and mentioned the methods of bias-adjustment. As proposed by the reviewer, we have added the exact quantile mapping method that was in the bias-adjustment of the rainfall data. In addition, we have referred to literature where the bias-adjustment of the DWD core ensemble is explained.

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

Please see our comment and proposed change in the major comment 3).

References:

Bürger, G., Pfister, A., and Bronstert, A.: Temperature-driven rise in extreme sub-hourly rainfall, Journal of Climate, 32, 7597–7609, 2019.

Derx, J., Müller-Thomy, H., Kılıç, H. S., Cervero-Arago, S., Linke, R., Lindner, G., Walochnik, J., Sommer, R., Komma, J., Farnleitner, A. H., and Blaschke, A. P.: A probabilistic-deterministic approach for assessing climate change effects on infection risks downstream of sewage emissions from CSOs, Water Research, 120746, https://doi.org/10.1016/j.watres.2023.120746, 2023.

Hänsel, S., Brendel, C., Fleischer, C., Ganske, A., Haller, M., Helms, M., Jensen, C., Jochumsen, K., Möller, J., Krähenmann, S., Nilson, E., Rauthe, M., Rasquin, C., Rudolph, E., Schade, N., Stanley, K., Wachler, B., Deutschländer, T., Tinz, B., Walter, A., Winkel, N., Krahe, P., and Höpp, S.: Vereinbarungen des Themenfeldes 1 im BMVI-Expertennetzwerk zur Analyse von klimawandelbedingten Änderungen in Atmosphäre und Hydrosphäre, https:// https://doi.bafg.de/BfG/2020/ExpNHS2020.2020.01.pdf, 2020.

Müller, H. and Haberlandt, U.: Temporal rainfall disaggregation using a multiplicative cascade model for spatial application in urban hydrology, Journal of Hydrology, 556, 847–864, https://doi.org/10.1016/j.jhydrol.2016.01.031, 2018.

Müller-Thomy, H.: Temporal rainfall disaggregation using a micro-canonical cascade model: possibilities to improve the autocorrelation, Hydrol. Earth Syst. Sci., 24, 169–188, https://doi.org/10.5194/hess-24-169-2020, 2020.

Olsson, J., Berggren, K., Olofsson, M., and Viklander, M.: Applying climate model precipitation scenarios for urban hydrological assessment: A case study in Kalmar City, Sweden, Atmospheric Research, 92, 364–375, https://doi.org/10.1016/j.atmosres.2009.01.015, 2009.

Pidoto, R., Bezak, N., Müller-Thomy, H., Shehu, B., Callau-Beyer, A. C., Zabret, K., and Haberlandt, U.: Comparison of rainfall generators with regionalisation for the estimation of rainfall erosivity at ungauged sites, Earth Surf. Dynam., 10, 851–863, https://doi.org/10.5194/esurf-10-851-2022, 2022.