We thank the anonymous reviewer for providing this comprehensive review. We acknowledge the suggestions to improve the quality of the manuscript and we hope that the respective modifications of the manuscript will satisfy your concerns. Below, we provide our point-by-point answer to your comments. [Color code: Reviewer comment: blue, authors' answer: green, revised text: black]

"Major Comments Numerical models present important limitations impacting the results, including: an incomplete understanding of the climate system, an imperfect ability to transform our knowledge into accurate mathematical equations, the limited power of computers, the models' inability to reproduce important atmospheric phenomena, and inaccurate representations of the complex natural interconnections. In addition, it is commonly recognized that numerical models present recurrent spurious precipitation from the numerical processes. The authors present a very good discussion in section 5. Another issue arises from converting station data (point point) to grid data. Given all that has been stated, what is your degree of confidence in the validation of the model and your results?"

Our degree of confidence in the validation varies according to the density of the rain gauges and the length of the underlying observational rainfall data. This measurement density seems to be mainly driven by the expected impact of sub-daily rainfall. Northern countries with lower shares of convectional sub-daily precipitation (Netherlands, Denmark, Sweden, Finland) provide only one return level value for the whole country (Sweden: four values), as they pooled the available rainfall data in order to acquire a data base, which is sufficient for extreme value analysis. Hence, our confidence level in the validation of these regions is "moderate", as the observational data base provides an order of magnitude of the return levels, but no spatial variance. For hourly to 12-hourly durations, Norway provides only a study, where results have been clearly stated to be experimental, and therefore uncertain, which affects the confidence in the validation as well.

Additionally, the deviations of the observational data of the Netherlands to Germany & Belgium and Norway/Sweden to each other (see text L300-305) support the argumentation that the confidence in the validation of these countries is only "moderate". Though we still wanted to include all possible observational data (as stated in the text in section 5.1). Also, the findings that country-wise data show deviations at the border provide some added value, as this investigation shows, where the existing 10-year return levels show major uncertainties and should, therefore, be revised.

In the remaining countries, the rain gauge measurement density and the length of available observations are sufficient for the area-wise calculation of 10-year return levels in our opinion. We have no possibility to validate our homogenized observational return level product, though we can look at the borders of the different country-wise return level calculations. For the non-northern countries, we see no major deviations at the borders of the different data sets, and topographical features are well preserved (Fig. 2,3,4: Alpine and Pre-alpine areas of Germany, Austria, Switzerland, Slovenia, Italy and France fit together very well for all durations). For 24-hourly duration (Fig. 4) we see almost no major deviations at any border. The good fit between these data sets also increases our confidence level in the quality of our homogenized observational product and, therefore, our confidence in the validation.

In order to address your question also within the manuscript, we will briefly add our thoughts at the end of section 5.1 (L368ff.), as it fits there perfectly thematically:

[This paragraph also features some revisions due to the reviewer comment RC2 of the other reviewer]

"…Even though the combined observational data set is subject to different limitations and uncertainties, it is a necessary approach to evaluate the return levels of climate models not only locally or countrywide, but to perform a validation on (almost) continental scale. To our knowledge, such an assessment has not been carried out before. The confidence level in this validation varies by country depending on the underlying rainfall database and the procedure of the return level calculation, which has been described within section 3. The obvious deviations in our homogenized observational return level product at the country borders between Norway and Sweden, between Italy and France and Switzerland as well as between the Netherlands and Germany and Belgium (as described in section 4), clearly show that the validation in these regions is subject to major uncertainties for hourly to 12-hourly durations. On the other hand, the good fit and the preservation of topographic features at the borders of Germany, Denmark, Belgium, France, Austria, Switzerland, and Slovenia support the confidence level in the validation for these regions. For the 24-hourly duration we find no major deviations along the country borders, which increases the confidence in this return level duration."

"In the manuscript are presented elements that do not appear referenced in the text. For example, Table 3, Figure 5, and Figure 6."

Figures 5 and 6 are referenced and analyzed in lines 379ff and 389ff.
A reference to Table 1 (there is only one table) is indeed missing, thanks a lot for pointing at this. We will include a reference to Table 1 in section 3 in line 148:

"… We describe the data processing for each national data set in the following. Table 1 provides an overview of the applied observational data and how they were accessed."

“Minor Comments

L 100 and from this line forward. The acronym CRCM-LE appears. What is LE? Each word or phrase should have only one meaning, and should be used consistently throughout the documentation.”

LE = large ensemble. The acronym is introduced in line 72:

2.1 The Canadian Regional Climate Model Version 5 Large Ensemble (CRCM5-LE)

“L123 – size of the window?”

The window sizes refer to the chosen durations of the 10-year return levels: hourly, 3-hourly, 6-hourly, 12-hourly and 24-hourly as introduced in line 60. For better readability, we will modify the text in L122 ff.:

“Due to the hourly resolution of the CRCM5-LE data, the hourly maxima are constrained to the fixed window at the full hour (e.g. 6:00 to 7:00). For all other durations (3-hourly, 6-hourly, 12-hourly and 24-hourly, respectively) we allow hourly moving windows for the selection of maxima.”

“L275- L290 This paragraph is confusing for the reader. Please clarify what Figure 1 shows: if the medians of the sums (L283) if the sums (L279). We are directed to a similar figure- Berg (2009) - referring to summer precipitation. Please clarify whether in Figure 1 we are analyzing summer or another season. In the caption of Figure 1 include the clarifications made, to help the reader in interpreting the figure more easily.”

Thanks for this comment, here we were not precise enough. We will modify the sentence in the text (L279ff.) and in the caption of Figure 1 (L800):

“Figure 1 shows the rainfall intensity for hourly and 12-hourly precipitation return levels for the European domain based on the median of the 50-member CRCM5-LE. Though covering the whole year, Figure 1 can be compared to the 10-year return levels of nine RCM setups of the EURO-CORDEX ensemble, which were calculated for summer-time precipitation only (Berg et al., 2019).”

“Figure 1: 10-year return levels of hourly (left) and 12-hourly (right) precipitation over Europe based on the median of the 50-member CRCM5-LE.”

“L 291- L 300 See the comments in the previous paragraph.”

We will modify:

L291: “For the 12-hourly duration, these areas also show the highest median rainfall intensities, with the Norwegian west coast and the Atlantic coast of northern Portugal and Spain also exhibiting high values.”

L293: “The 12-hourly 10-year return levels based on the median of the CRCM5-LE are similar to all nine RCM-GCM combinations of Berg et al. (2019) in terms of spatial patterns and rainfall intensities.”

“L 300 From this line to the end of the paragraph. These text is confusing and needs clarification. First, it is necessary that the authors clearly identify which figures are under analysis. This block of text is close to imperceptible without the clear identification of the figures. Analyze the figures in the same order as they are presented (Figure 2, text; Figure 3, text, and so on).”

We have sorted this paragraph not by the duration of the return levels in order to avoid unnecessary repetitions. Therefore, we sorted this paragraph by the topic of the analysis:

L300-305: Describing the observational dataset and the country-wise deviations.

L306-313: Analysis of the areas, where observations are/are not in range of the CRCM5-LE.

L314-321: Description of the biases.

L322-326: Analysis of the spatial correlation.

If we would change this order and present the results from figure to figure, the text would contain a lot of duplications and the comparability between the different durations would decrease.

Still, we see your concerns regarding the clarification of figure assignment and analysis. Hence, we will add more references to the analyzed figures within the paragraph:

[This paragraph also features some revisions due to the reviewer comment RC2 of the other reviewer]

“The combined observational datasets (see Fig. 2, 3, 4) show quite smooth transitions between most of the different data sources and methods. The biggest deviation is found at the border of Norway and Sweden for hourly to 12-hourly durations (see Fig. 2 & 3), as the estimate of the rainfall return level for western Sweden by Olsson et al. (2018) is a lot higher than the estimate by Dyrrdal et al. (2015) for eastern Norway. This is due to the sparse sampling of observations and differing approaches to derive return levels (see Sect. 3.1). We also find slight deviations for the Netherlands, where the return levels by Beersma et al. (2018) are higher than the surrounding levels for northern Belgium and western Germany. For the shorter durations of hourly and 3-hourly return levels (see Fig. 2), deviations occur at the border between Italy and France as well as between Italy and Switzerland. This is due to the higher ARF applied in Italy (see Section 3.2). These deviations emphasize the need for homogeneous data sets of extreme precipitation.

As the 50 members of the CRCM5-LE also provide a range of equally probable estimations of return levels, we hatch areas, where the observations are not within the range of the regional climate model ensemble. The rainfall intensity of the observational data set is within the range of the climate model generated intensities in 60 % (77 %, 78 %, 83 %, 78 %) of the area for hourly (3-hourly, 6-hourly, 12-hourly, 24-hourly) durations (see Fig. 2, 3, 4). This fraction of areas is gradually increasing between hourly and 12-hourly durations, whereas it slightly decreases for the 24-hourly duration. For the 24-hourly return period, data for the Iberian Peninsula and Poland was added, whereby no data for these countries was available for the hourly to 12-hourly evaluation. Without these additional data sets, the fraction of areas, where 24-hourly observational return levels are within the CRCM5-LE return levels, would amount to 80 %. In addition, in the Netherlands, Switzerland and Norway, different data bases are used for the estimations of the return levels of hourly to 12-hourly durations and 24-hourly duration (see Section 3).

The hourly intensities are generally underestimated by the CRCM5-LE except for England and Wales, northern Italy, northern Austria and the northern part of Norway, resulting in an areal average bias of -16.3 % (see Fig. 2). There is also an area-wide underestimation in the Mediterranean as well as Scandinavia in all 50 members of the large ensemble, which is why the observations are not in the range of the CRCM5-LE for large parts of these areas (see Fig. 2). For durations of three to twelve hours, the biases over the whole area decrease to -1.0 %, -0.5 % and +0.1 % (see Fig. 2 & 3). The high intensities of southern France, southern Switzerland and parts of Italy are underestimated (see Fig. 2 & 3). Also in Sweden and Finland the observational data sets report higher rainfall intensities. For the 24-hourly aggregation, the bias amounts to +8.2 % (see Fig. 4). The CRCM5 overestimates 24-hourly rainfall intensities in western Norway and at the Atlantic coast of the northern Iberian Peninsula, which is why the observations are not in the range of the 50 CRCM-LE members (see Fig. 4).

We calculate the Spearman's rank correlation coefficient ρ as a measure to compare the spatial patterns. For the median of the return levels of the CRCM5-LE and the observational data the coefficient amounts to 0.83 (0.81, 0.76, 0.78, 0.83, respectively) of the area for hourly (3-hourly, 6-hourly, 12-hourly, 24-hourly, respectively) durations. These values confirm the visual impression of a high spatial pattern correlation when comparing both data sets (see Fig. 2, 3, 4)."

"L305. what is the figure under discussion? Figure 4? In relation to Figure 4, the authors explain well the deviation in Norway and the Netherlands but what about southern Europe?"

See modification of the text above. This paragraph refers to the country-wise deviations of the observational datasets. We find deviations in Norway/Sweden and the Netherlands/Germany/Belgium. The observational datasets in central and southern Europe do not show any noticeable deviations to each other at the borders of the countries.

[After the application of different areal reduction factors (ARF) for Italy due to the reviewer comment RC2, we find deviations at the border between Italy and France/Switzerland. We add the description of these differences within the text.]

"Table 1 ???? Figure 5/6 is presented, but the analysis is missing."

See answer above in the Major Comments section.

"Figure 2. This is not Europe; this is some regions of Europe."

We will specify the figure caption of Figures 2, 3 & 4:

"….precipitation over parts of Europe. …"